

Principales medidas en epidemiología

Alejandra Moreno-Altamirano, C.D., M. en C.,⁽¹⁾ Sergio López-Moreno, M.C.,⁽²⁾
 Alexánder Corcho-Berdugo, M.C.⁽²⁾

Concepto de medición, variables y escalas

Una vez que se ha identificado un problema científico y se ha aventurado una explicación hipotética, es necesario someterla a prueba. Para contrastar la hipótesis se requiere descomponerla en un conjunto suficientemente pequeño de variables susceptibles de ser evaluadas empíricamente. Si los procedimientos empíricos no refutan la hipótesis planteada ésta se acepta como probablemente verdadera. En pocas palabras, este es el camino que el científico sigue más frecuentemente al realizar su trabajo. Dado que en la mayoría de los casos es necesario medir las variables durante la contrastación empírica de la hipótesis, la medición resulta un procedimiento indispensable en la práctica científica.

En epidemiología, el proceso de investigación es similar al utilizado en el resto de las ciencias. Cuando se investiga la salud de la población también se proponen una o varias explicaciones hipotéticas que posteriormente son sometidas a contrastación empírica. En este proceso, los conceptos de *medición* y de *variable* resultan fundamentales.

Concepto de variable

La función de las variables consiste en proporcionar información asequible para descomponer la hipótesis planteada en sus elementos más simples. Las variables pueden definirse como aquellos atributos o características de los eventos, de las personas o de los

grupos de estudio que cambian de una situación a otra o de un tiempo a otro y que, por lo tanto, pueden tomar diversos valores. Para su estudio es necesario medirlos en el objeto investigado, y es en el marco del problema y de las hipótesis planteadas donde adquieren el carácter de variables.

De acuerdo con la relación que guardan unas con otras, las variables se clasifican en independientes (o variables explicativas) y dependientes (o variables respuesta). Cuando se supone que una variable produce un cambio en otra, se considera a la primera como independiente (o causa) y a la segunda como dependiente (o efecto). En los estudios epidemiológicos la enfermedad o evento es por lo general la variable dependiente y los factores que determinan su aparición, magnitud y distribución son las variables independientes, o exposición. No obstante, el concepto de dependencia e independencia es contextual, es decir, obedece al modelo teórico planteado. Una vez que se han identificado las variables el investigador debe definir las de manera operativa, especificando el método y la escala con las cuales llevará a cabo su medición.

El uso de variables permite a la epidemiología la elaboración de modelos descriptivos, explicativos y predictivos sobre la dinámica de la salud poblacional. En los modelos más sencillos (por ejemplo, en los modelos en los que se considera una sola exposición y un solo daño o evento) las variables generalmente se expresan en tablas simples de dos categorías mutuamente excluyentes (llamadas dicotómicas), representadas por la ausencia y la presencia de la exposición y la

(1) Departamento de Salud Pública, Facultad de Medicina, Universidad Nacional Autónoma de México.

(2) Dirección de Políticas y Planeación, Centro de Investigación en Sistemas de Salud, Instituto Nacional de Salud Pública, México.

ausencia y la presencia del evento. Al combinar ambas categorías se forma una tabla con dos filas y dos columnas, conocida como tabla tetracórica o tabla de 2 por 2. Cuando, en cambio, existen más de dos categorías de exposición, o varias formas de clasificar el evento, esta relación se expresa en tablas de varias columnas y varias celdas. En este texto se analizará la elaboración de medidas epidemiológicas basadas en categorías dicotómicas y el uso de tablas de 2 X 2.

Concepto de medición

La medición consiste en asignar un número o una calificación a alguna propiedad específica de un individuo, una población o un evento usando ciertas reglas. No obstante, la medición es un proceso de abstracción. En términos estrictos no se mide al individuo sino cierta característica suya, abstrayéndola de otras propiedades. Uno no mide al niño sino que obtiene información sobre su estatura o su peso. Además, lo que se hace es comparar el atributo medido en otros individuos (o en el mismo individuo en otro momento), con el fin de evaluar sus cambios en el tiempo o cuando se presenta en condiciones distintas de las originales.

Para medir es necesario seguir un proceso que consiste, en breves palabras, en el paso de una entidad teórica a una escala conceptual y, posteriormente, a una escala operativa.

En general, los pasos que se siguen durante la medición son los siguientes: a) se delimita la parte del evento que se medirá, b) se selecciona la escala con la que se medirá, c) se compara el atributo medido con la escala y, d) finalmente, se emite un juicio de valor acerca de los resultados de la comparación. Para medir el crecimiento de un menor, por ejemplo, primero se selecciona la variable a medir (la edad, el peso, la talla); luego se seleccionan las escalas de medición (meses cumplidos, centímetros, gramos); inmediatamente después se comparan los atributos con las escalas seleccionadas (un mes de edad, 60 cm de talla, 4 500 gramos de peso) y, por último, se emite un juicio de valor, que resume la comparación entre las magnitudes encontradas y los criterios de salud aceptados como válidos en ese momento. Como resultado, el infante se califica como bien nutrido, desnutrido o sobrenutrido.

Como se puede notar, la medición es un proceso instrumental sólo en apariencia, ya que la selección de la parte que se medirá, de la escala de medición y de los criterios de salud que se usarán como elementos de juicio deben ser resultado de un proceso de decisión teórica. En otras palabras, sólo puede medirse

lo que antes se ha concebido teóricamente. La medición, sin embargo, nos permite alcanzar un alto grado de objetividad al usar los instrumentos, escalas y criterios aceptados como válidos por la mayor parte de la comunidad científica.

Principales escalas de medición

Las escalas se clasifican en cualitativas (nominal y ordinal) y cuantitativas (de intervalo y de razón). Un requisito indispensable en todas las escalas es que las categorías deben ser exhaustivas y mutuamente excluyentes. En otras palabras, debe existir una categoría para cada caso que se presente y cada caso debe poder colocarse en una sola categoría.

Escala nominal

La medición de carácter nominal consiste simplemente en clasificar las observaciones en categorías diferentes con base en la presencia o ausencia de cierta cualidad. De acuerdo con el número de categorías resultantes, las variables se clasifican en dicotómicas (dos categorías) o politómicas (más de dos categorías). En las escalas nominales no es posible establecer un orden de grado como mejor o peor, superior o inferior, o más o menos. La asignación de códigos numéricos a las categorías se hace con el único fin de diferenciar unas de otras y no tienen interpretación en lo que se refiere al orden o magnitud del atributo. Como ejemplos de este tipo de medición en la investigación epidemiológica se pueden mencionar el sexo (masculino "0", femenino "1"), el estado civil (soltero, casado, viudo, divorciado), la exposición o no a un factor X, y el lugar de nacimiento, entre otras.

Escala ordinal

En contraste con las escalas nominales, en este tipo de medición las observaciones se clasifican y ordenan por categorías según el grado en que los objetos o eventos poseen una determinada característica. Por ejemplo, se puede clasificar a las personas con respecto al grado de una enfermedad en leve, moderado o severo. Si se llega a utilizar números en este tipo de escalas su única significación consiste en indicar la posición de las distintas categorías de la serie y no la magnitud de la diferencia entre las categorías. Para la variable antes mencionada, por ejemplo, sabemos que existe una diferencia de grado entre leve y severo, pero no es posible establecer con exactitud la magnitud de la diferencia en las enfermedades de una u otra personas.

Escala de intervalo

Esta es una escala de tipo cuantitativo en la que, además de ordenar las observaciones por categorías del atributo, se puede medir la magnitud de la distancia relativa entre las categorías. Esta escala, sin embargo, no proporciona información sobre la magnitud absoluta del atributo medido. Por ejemplo, se puede obtener una escala de intervalo para la altura de las personas de un grupo si, en lugar de medirlas directamente, se mide la altura de cada persona con respecto a la altura promedio. En este caso, el valor cero es arbitrario y los valores asignados a la altura no expresan su magnitud absoluta. Esta es la característica distintiva de las escalas de intervalo en comparación con las de razón.

El ejemplo más conocido de las escalas de intervalo es la escala de Celsius para medir la temperatura, en la que por convención el grado cero corresponde al punto de congelación del agua y donde, por lo tanto, la razón entre dos objetos con temperaturas de 10 y 20 grados no indica que uno de ellos sea realmente dos veces más caliente (o más frío) que el otro. En ciencias de la salud, un buen ejemplo de este tipo de escalas es la utilizada para medir el coeficiente intelectual.

Escalas de razón

Esta escala tiene la cualidad de que el cero sí indica la ausencia del atributo y, por lo tanto, la razón entre dos números de la escala es igual a la relación real existente entre las características de los objetos medidos. En otras palabras, cuando decimos que un objeto pesa 8 kg estamos también diciendo que pesa el doble que otro cuyo peso es de 4 kg, y que un avión que viaja a 600 km por hora tardará en llegar a su destino la mitad del tiempo que tardaría si viajara a 300 km por hora. Muchas características biofísicas y químicas que pueden ser medidas en las unidades convencionalmente aceptadas (metros, gramos, micras, mol/kg, mg/dl, etc.) son ejemplos de mediciones que corresponden a este tipo de escala. En materia de investigación social y de salud, el ingreso económico y la concentración de plomo en sangre son buenos ejemplos de este tipo de escalas.

Cálculo de proporciones, tasas y razones

Un rasgo característico de la contrastación en los estudios epidemiológicos es que las relaciones causales postuladas entre las variables se traducen en términos probabilísticos. Es decir, se trata de establecer si la mayor o menor probabilidad de que un evento

ocurra se debe precisamente a los factores que se sospecha intervienen en su génesis y no al azar. Para cumplir con este objetivo, la investigación epidemiológica se basa en la construcción de tres tipos de medidas: a) de frecuencia; b) de asociación o efecto, y c) de impacto potencial. La construcción de estas medidas se realiza por medio de operaciones aritméticas simples y de los instrumentos matemáticos conocidos como razones, proporciones y tasas. Antes de abordar las medidas utilizadas en los estudios epidemiológicos repasaremos brevemente estos tres conceptos.

Proporciones

Las proporciones son medidas que expresan la frecuencia con la que ocurre un evento en relación con la población total en la cual éste puede ocurrir. Esta medida se calcula dividiendo el número de eventos ocurridos entre la población en la que ocurrieron. Como cada elemento de la población puede contribuir únicamente con un evento es lógico que al ser el numerador (el volumen de eventos) una parte del denominador (población en la que se presentaron los eventos) aquel nunca pueda ser más grande que éste. Esta es la razón por la que el resultado nunca pueda ser mayor que la unidad y oscile siempre entre cero y uno.

Por ejemplo, si en un año se presentan tres muertes en una población compuesta por 100 personas, la proporción anual de muertes en esa población será:

$$P = \frac{3 \text{ muertes}}{100 \text{ personas}} = 0.03$$

A menudo las proporciones se expresan en forma de porcentaje, y en tal caso los resultados oscilan entre cero y 100. En el ejemplo anterior, la proporción anual de muertes en la población sería de 3 por 100, o de 3%. Nótese, asimismo, que el denominador no incluye el tiempo. Las proporciones expresan únicamente la relación que existe entre el número de veces en las que se presenta un evento y el número total de ocasiones en las que se pudo presentar.

Tasas

Las tasas expresan la dinámica de un suceso en una población a lo largo del tiempo. Se pueden definir como la magnitud del cambio de una variable (enfermedad o muerte) por unidad de cambio de otra (usualmente el tiempo) en relación con el tamaño de la población que se encuentra en riesgo de experimentar el suceso.

En las tasas, el numerador expresa el número de eventos acaecidos durante un periodo en un número determinado de sujetos observados.

A diferencia de una proporción el denominador de una tasa no expresa el número de sujetos en observación sino el tiempo durante el cual tales sujetos estuvieron en riesgo de sufrir el evento. La unidad de medida empleada se conoce como tiempo-persona de seguimiento. Por ejemplo, la observación de 100 individuos libres del evento durante un año corresponde a 100 años-persona de seguimiento; de manera similar, 10 sujetos observados durante diez años corresponden a 100 años-persona.

Dado que el periodo entre el inicio de la observación y el momento en que aparece un evento puede variar de un individuo a otro, el denominador de la tasa se estima a partir de la suma de los periodos de todos los individuos. Las unidades de tiempo pueden ser horas, días, meses o años, dependiendo de la naturaleza del evento que se estudia.

El cálculo de tasas se realiza dividiendo el total de eventos ocurridos en un periodo dado en una población entre el tiempo-persona total (es decir, la suma de los periodos individuales libres de la enfermedad) en el que los sujetos estuvieron en riesgo de presentar el evento. Las tasas se expresan multiplicando el resultado obtenido por una potencia de 10, con el fin de permitir rápidamente su comparación con otras tasas.

$$\text{Tasa} = \frac{\text{número de eventos ocurridos en una población en un periodo } t}{\text{sumatoria de los periodos durante los cuales los sujetos de la población libres del evento estuvieron expuestos al riesgo de presentarlo en el mismo periodo}} \times \text{una potencia de 10}$$

Razones

Las razones pueden definirse como magnitudes que expresan la relación aritmética existente entre dos eventos en una misma población, o un solo evento en dos poblaciones. En el primer caso, un ejemplo es la razón de residencia hombre: mujer en una misma población. Si en una localidad residen 5 000 hombres y 4 000 mujeres se dice que, en ese lugar, la razón de residencia hombre:mujer es de 1:0.8 (se lee 1 a 0.8), lo que significa que por cada hombre residen ahí 0.8 mujeres. Esta cantidad se obtiene como sigue:

$$\text{Razón hombre: mujer} = \frac{4\,000}{5\,000} = 0.8$$

En este caso, también se podría decir que la razón hombre:mujer es de 10:8, pues esta expresión aritmética es igual a la primera (1:0.8).

En el segundo ejemplo se encuentran casos como la razón de tasas de mortalidad por causa específica (por ejemplo, por diarreas) en dos comunidades. En este caso, la razón expresaría la relación cuantitativa que existe entre la tasa de mortalidad secundaria a diarreas registrada en la primera ciudad y la tasa de mortalidad secundaria a diarreas registrada en la segunda. La razón obtenida expresa la magnitud relativa con la que se presenta este evento en cada población. Si la tasa de mortalidad por diarreas en la primera ciudad es de 50 por 1 000 y en la segunda de 25 por 1 000 la razón de tasas entre ambas ciudades sería:

$$\text{RTM} = \frac{\text{tasa de mortalidad en la ciudad B}}{\text{tasa de mortalidad en la ciudad A}} = \frac{50 \times 1\,000}{25 \times 1\,000} = 2.0$$

Donde RTM es la razón de tasas de mortalidad (en este caso, por diarreas) entre las ciudades A y B. El resultado se expresa como una razón de 1:2, lo que significa que por cada caso en la ciudad A hay 2 en la ciudad B.

Medidas de frecuencia

El paso inicial de toda investigación epidemiológica es medir la frecuencia de los eventos de salud con el fin de hacer comparaciones entre distintas poblaciones o en la misma población a través del tiempo. No obstante, dado que el número absoluto de eventos depende en gran medida del tamaño de la población en la que se investiga, estas comparaciones no se pueden realizar utilizando cifras de frecuencia absoluta (o número absoluto de eventos).

Por ejemplo, si en dos diferentes poblaciones se presentan 100 y 200 casos de cáncer cervicouterino, respectivamente, se podría pensar que en el segundo grupo la magnitud del problema es del doble que en el primero. Sin embargo, esta interpretación sería incorrecta si el segundo grupo tuviera el doble de tamaño que el primero, ya que la diferencia en el número de casos podría deberse simplemente al mayor tamaño de la segunda población y no a la presencia de un factor de riesgo extraordinario. Aunque la frecuencia absoluta cambie la magnitud del problema puede ser la misma.

En consecuencia, para comparar adecuadamente la frecuencia de los eventos de salud es necesario construir una medida que sea independiente del tamaño

de la población en la que se realiza la medición. Este tipo de medidas, denominadas medidas de frecuencia relativa, se obtiene, en general, relacionando el número de casos (numerador) con el número total de individuos que componen la población (denominador). El cálculo correcto de estas medidas requiere que se especifique claramente qué constituye el numerador y el denominador. Es evidente, por ejemplo, que los varones no deben ser incluidos en el denominador durante el cálculo de la frecuencia relativa de carcinoma del cérvix.

La parte de la población que es susceptible a una enfermedad se denomina población en riesgo. Así, por ejemplo, los accidentes laborales sólo afectan a las personas que trabajan, por lo que la población en riesgo es la población trabajadora. Si, en cambio, queremos investigar el efecto de un contaminante generado por una fábrica podríamos ampliar el denominador a toda la población expuesta al mismo, sea o no trabajadora.

Las medidas de frecuencia más usadas en epidemiología se refieren a la medición de la mortalidad o la morbilidad en una población. La mortalidad es útil para estudiar enfermedades que provocan la muerte, especialmente cuando su letalidad es importante. Empero, cuando la letalidad es baja y, en consecuencia, la frecuencia con la que se presenta una enfermedad no puede analizarse adecuadamente con los datos de mortalidad, la morbilidad se convierte en la medida epidemiológica de mayor importancia.

En ocasiones, la morbilidad también puede servir para explicar las tendencias de la mortalidad, ya que los cambios en la mortalidad pueden ser secundarios a cambios ocurridos antes en la morbilidad o, por el contrario, las tendencias en la mortalidad pueden explicar los cambios en los patrones de morbilidad cuando, por ejemplo, la disminución en la mortalidad infantil explica los aumentos aparentes en el volumen de enfermedades en otras edades. Por ambas razones, el análisis de las condiciones de salud de las poblaciones se basa siempre en los cambios observados en las medidas de mortalidad y morbilidad.

Las principales fuentes de información de morbilidad son los datos hospitalarios y los registros de enfermedad. Sin embargo, debido a las limitaciones de estos registros, los estudios epidemiológicos se basan en información obtenida mediante métodos de detección especialmente diseñados para ello. A continuación se presenta un resumen de los elementos más importantes de las medidas de mortalidad y morbilidad.

Medidas de mortalidad

El concepto de mortalidad expresa la magnitud con la que se presenta la muerte en una población en un

momento determinado. A diferencia de los conceptos de muerte y defunción que reflejan la pérdida de la vida biológica individual, la mortalidad es una categoría de naturaleza estrictamente poblacional. En consecuencia, la mortalidad expresa la dinámica de las muertes acaecidas en las poblaciones a través del tiempo y el espacio, y sólo permite comparaciones en este nivel de análisis. La mortalidad puede estimarse para todos o algunos grupos de edad, para uno o ambos sexos y para una, varias o todas las enfermedades. La mortalidad se clasifica de la siguiente manera: a) general y b) específica.

Mortalidad general

La mortalidad general es el volumen de muertes ocurridas por todas las causas de enfermedad, en todos los grupos de edad y para ambos sexos. La mortalidad general, que comúnmente se expresa en forma de tasa, puede ser cruda o ajustada, de acuerdo con el tratamiento estadístico que reciba.

La mortalidad cruda expresa la relación que existe entre el volumen de muertes ocurridas en un periodo dado y el tamaño de la población en la que éstas se presentaron; la mortalidad ajustada (o estandarizada) expresa esta relación pero considera las posibles diferencias en la estructura por edad, sexo, etcétera, de las poblaciones analizadas, lo que permite hacer comparaciones entre éstas. En este caso, las tasas se reportan como tasas ajustadas o estandarizadas. La tasa cruda de mortalidad se calcula de acuerdo con la siguiente fórmula:

$$\text{Tasa mortalidad general} = \frac{\text{número de muertes en el periodo } t}{\text{población total promedio en el mismo periodo}} \times 10n$$

Mortalidad específica

Cuando existen razones para suponer que la mortalidad puede variar entre los distintos subgrupos de la población ésta se divide para su estudio. Cada una de las medidas obtenidas de esta manera adopta su nombre según la fracción poblacional que se reporte. Por ejemplo, si las tasas de mortalidad se calculan para los diferentes grupos de edad, serán denominadas tasas de mortalidad por edad. De la misma manera pueden calcularse la mortalidad por sexo, por causa específica, etcétera.

En algunos casos pueden calcularse combinaciones de varias fracciones poblacionales, y cuando es así, se especifican los grupos considerados (por ejemplo, mortalidad femenina en edad reproductiva). Las ta-

tas de mortalidad específica por edad y sexo se calculan de la siguiente forma:

$$TME = \frac{\text{total de muertes en un grupo de edad y sexo específicos de la población durante un periodo dado}}{\text{población total estimada del mismo grupo de edad y sexo en el mismo periodo}} \times 10n$$

Donde TME es la tasa de mortalidad específica para esa edad y sexo.

Tasa de letalidad. La letalidad es una medida de la gravedad de una enfermedad considerada desde el punto de vista poblacional, y se define como la proporción de casos de una enfermedad que resultan mortales con respecto al total de casos en un periodo especificado. La medida indica la importancia de la enfermedad en términos de su capacidad para producir la muerte y se calcula de la manera siguiente:

$$\text{Letalidad (\%)} = \frac{\text{número de muertes por una enfermedad en un periodo determinado}}{\text{número de casos diagnosticados de la misma enfermedad en el mismo periodo}} \times 100$$

La letalidad, en sentido estricto, es una proporción ya que expresa el número de defunciones entre el número de casos del cual las defunciones forman parte. No obstante, generalmente se expresa como tasa de letalidad y se reporta como el porcentaje de muertes de una causa específica con respecto al total de enfermos de esa causa.

Medidas de morbilidad

La enfermedad puede medirse en términos de prevalencia o de incidencia. La prevalencia se refiere al número de individuos que, en relación con la población total, padecen una enfermedad determinada en un momento específico. Debido a que un individuo sólo puede encontrarse sano o enfermo con respecto a cualquier enfermedad, la prevalencia representa la probabilidad de que un individuo sea un caso de dicha enfermedad en un momento específico.

La incidencia, por su parte, expresa el volumen de casos nuevos que aparecen en un periodo determinado, así como la velocidad con la que lo hacen; es decir, expresa la probabilidad y la velocidad con la que los individuos de una población determinada desarrollarán una enfermedad durante cierto periodo.

Prevalencia

La prevalencia es una proporción que indica la frecuencia de un evento. En general, se define como la proporción de la población que padece la enfermedad en estudio en un momento dado, y se denomina únicamente como prevalencia (p). Como todas las proporciones, no tiene dimensiones y nunca puede tomar valores menores de 0 o mayores de 1. A menudo, se expresa como casos por 1 000 o por 100 habitantes.

En la construcción de esta medida no siempre se conoce en forma precisa la población expuesta al riesgo y, por lo general, se utiliza sólo una aproximación de la población total del área estudiada. Si los datos se han recogido en un momento o punto temporal dado, p es llamada prevalencia puntual.

Prevalencia puntual. La prevalencia puntual es la probabilidad de un individuo de una población de ser un caso en el momento t , y se calcula de la siguiente manera:

$$p = \frac{\text{número total de casos existentes al momento } t}{\text{total de la población en el momento } t} \times 10n$$

La prevalencia de una enfermedad aumenta como consecuencia de una mayor duración de la enfermedad, la prolongación de la vida de los pacientes sin que éstos se curen, el aumento de casos nuevos, la inmigración de casos (o de susceptibles), la emigración de sanos y la mejoría de las posibilidades diagnósticas. La prevalencia de una enfermedad, por su parte, disminuye cuando es menor la duración de la enfermedad, existe una elevada tasa de letalidad, disminuyen los casos nuevos, hay inmigración de personas sanas, emigración de casos y aumento de la tasa de curación. En resumen, la prevalencia de una enfermedad depende de la incidencia y de la duración de la enfermedad.

Dado que la prevalencia depende de tantos factores no relacionados directamente con la causa de la enfermedad, los estudios de prevalencia no proporcionan pruebas claras de causalidad aunque a veces puedan sugerirla. Sin embargo, son útiles para valorar la necesidad de asistencia sanitaria, planificar los servicios de salud o estimar las necesidades asistenciales.

Anteriormente era común el cálculo de la llamada prevalencia de periodo (o lápsica), que buscaba identificar el número total de personas que presentaban la enfermedad o atributo a lo largo de un periodo determinado. No obstante, debido a las confusiones

que origina, esta medida es cada vez menos empleada, y en materia de investigación es mejor no utilizarla.

Incidencia

En los estudios epidemiológicos en los que el propósito es la investigación causal o la evaluación de medidas preventivas, el interés está dirigido a la medición del flujo que se establece entre la salud y la enfermedad, es decir, a la aparición de casos nuevos. Como ya se mencionó anteriormente, la medida epidemiológica que mejor expresa este cambio de estado es la incidencia, la cual indica la frecuencia con que ocurren nuevos eventos. A diferencia de los estudios de prevalencia, los estudios de incidencia inician con poblaciones de susceptibles libres del evento en las cuales se observa la presentación de casos nuevos a lo largo de un periodo de seguimiento. De esta manera, los resultados no sólo indican el volumen final de casos nuevos aparecidos durante el seguimiento sino que permiten establecer relaciones de causa-efecto entre determinadas características de la población y enfermedades específicas. La incidencia de una enfermedad puede medirse de dos formas: mediante la tasa de incidencia (basada en el tiempo-persona) y mediante la incidencia acumulada (basada en el número de personas en riesgo). La tasa de incidencia (también denominada densidad de incidencia) expresa la ocurrencia de la enfermedad entre la población en relación con unidades de tiempo-persona, por lo que mide la velocidad de ocurrencia de la enfermedad. La incidencia acumulada, en cambio, expresa únicamente el volumen de casos nuevos ocurridos en una población durante un periodo, y mide la probabilidad de que un individuo desarrolle el evento en estudio. La incidencia acumulada, por esta razón, también es denominada riesgo.

Tasa de incidencia o densidad de incidencia. La tasa de incidencia (TI) es la principal medida de frecuencia de enfermedad y se define como “el potencial instantáneo de cambio en el estado de salud por unidad de tiempo, durante un periodo específico, en relación con el tamaño de la población susceptible en el mismo periodo”. Para que una persona se considere expuesta al riesgo en el periodo de observación debe iniciar éste sin tener la enfermedad (el evento en estudio).

El cálculo del denominador de la TI se realiza sumando los tiempos libres de enfermedad de cada uno de los individuos que conforman el grupo y que permanecen en el estudio durante el periodo. Este número se mide generalmente en años, pero pueden ser meses,

semanas o días, y se conoce como tiempo en riesgo o tiempo-persona.

El número de individuos que pasan del estado sano al estado enfermo durante cualquier periodo depende de tres factores: a) del tamaño de la población, b) de la amplitud del periodo de tiempo, y c) del poder patógeno de la enfermedad sobre la población. La tasa de incidencia mide este poder, y se obtiene dividiendo el número observado de casos entre el tiempo total en el que la población ha estado en riesgo, equivalente a la sumatoria de los periodos individuales en riesgo. Al sumar periodos de observación que pueden variar de uno a otro individuo y considerar sólo el tiempo total en riesgo la TI corrige el efecto de entrada y salida de individuos al grupo durante el periodo de seguimiento.

A menudo no es posible calcular exactamente la duración del tiempo-persona para los individuos que ya no están en riesgo, debido a que desarrollaron la enfermedad. No obstante, para este grupo el valor total del tiempo-persona en riesgo puede estimarse de manera aproximada –y generalmente satisfactoria– multiplicando el tamaño medio de la población por la duración del periodo de observación.

La TI no es una proporción –como la prevalencia y la incidencia acumulada– dado que el denominador expresa unidades de tiempo y, en consecuencia, mide casos por unidad de tiempo. Esto hace que la magnitud de la TI no pueda ser inferior a cero ni tenga límite superior. La fórmula general para el cálculo de la TI es la siguiente:

$$\text{Tasa de incidencia} = \frac{\text{número de casos nuevos}}{\text{suma de todos los periodos libres de la enfermedad durante el periodo definido en el estudio (tiempo-persona)}}$$

Incidencia acumulada. La incidencia acumulada (IA) se puede definir como la probabilidad de desarrollar el evento, es decir, la proporción de individuos de una población que, en teoría, desarrollarían una enfermedad si todos sus miembros fuesen susceptibles a ella y ninguno falleciese a causa de otras enfermedades. También se ha definido simplemente como la probabilidad, o riesgo medio de los miembros de una población, de contraer una enfermedad en un periodo específico.

Las cifras obtenidas mediante el cálculo de la IA son relativamente fáciles de interpretar y proporcionan una medida sumamente útil para comparar los diferentes riesgos de distintas poblaciones. Para calcular la IA en el numerador se coloca el número de personas

que desarrollan la enfermedad durante el periodo de estudio (llamados casos nuevos) y en el denominador el número de individuos libres de la enfermedad al comienzo del periodo y que, por tanto, estaban en riesgo de padecerla. La incidencia acumulada es una proporción y, por lo tanto, sus valores sólo pueden variar entre 0 y 1. A diferencia de la tasa de incidencia la IA es adimensional. Su fórmula es la siguiente:

$$IA = \frac{\text{número de personas que contraen la enfermedad en un periodo determinado}}{\text{número de personas libres de la enfermedad en la población expuesta al riesgo en el inicio del estudio}}$$

Como la duración del periodo de observación influye directamente sobre la IA su amplitud debe considerarse siempre que se interprete esta medida. Cuando los miembros de una población tienen diferentes periodos bajo riesgo –debido a que se incorporan o abandonan el grupo a lo largo del periodo de seguimiento– la IA no puede calcularse directamente.

Medidas de asociación o de efecto

Las medidas de asociación son indicadores epidemiológicos que evalúan la fuerza con la que una determinada enfermedad o evento de salud (que se presume como efecto) se asocia con un determinado factor (que se presume como su causa).

Epidemiológicamente, las medidas de asociación son comparaciones de incidencias: la incidencia de la enfermedad en las personas que se expusieron al factor estudiado (o incidencia entre los expuestos) contra la incidencia de la enfermedad en las personas que no se expusieron al factor estudiado (o incidencia entre los no expuestos). Estadísticamente, lo que estos indicadores miden es la magnitud de la diferencia observada. Debido a que las medidas de asociación establecen la fuerza con la que la exposición se asocia a la enfermedad, bajo ciertas circunstancias estas medidas permiten realizar inferencias causales, especialmente cuando se pueden evaluar mediante una función estadística. En este documento se abordará el cálculo de medidas de asociación para variables dicotómicas.

Las medidas de asociación más sólidas se calculan utilizando la incidencia, ya que esta medida de frecuencia nos permite establecer, sin ninguna duda, que el efecto (el evento o enfermedad) es posterior a la causa (la exposición). En estos casos, se dice, existe una correcta relación temporal entre la causa y el efecto. Empero, en los estudios en los que no existe suficiente información para calcular la incidencia (como las

encuestas transversales y la mayoría de los estudios de casos y controles) no es posible calcular la incidencia. En estos casos puede estimarse la asociación entre el evento y la exposición al comparar las prevalencias a partir de la razón de prevalencias (RP) o de productos cruzados (RPC).

En general, hay dos tipos de medidas de asociación: las de diferencia (o de efecto absoluto) y las de razón (o de efecto relativo).

Medidas de diferencia

Como indica su nombre, estas medidas expresan la diferencia existente en una misma medida de frecuencia (idealmente la incidencia) entre dos poblaciones.

En general, las medidas de diferencia indican la contribución de un determinado factor en la producción de enfermedad entre los que están expuestos a él. Su uso se basa en la suposición de que tal factor es responsable de la aparición de la enfermedad y en la presunción de que, de no existir, los riesgos en ambos grupos serían iguales. Por este motivo, se dice que las medidas de diferencia indican el riesgo de enfermar que podría evitarse si se eliminara la exposición. Como sinónimo se emplea el término riesgo atribuible. Estas medidas se calculan de la siguiente manera:

$$\text{Diferencia} = E_i - E_o \times 100$$

donde,

E_i es la frecuencia de enfermar o morir de un grupo expuesto, y

E_o es la frecuencia de enfermar o morir en el grupo no expuesto.

El resultado se interpreta de la siguiente forma:

Valor =0 indica no-asociación (valor nulo).

Valores <0 indica asociación negativa y puede tomar valores negativos hasta infinito.

Valores >0 indica asociación positiva y puede tomar valores positivos hasta infinito.

Debe señalarse que el término riesgo atribuible carece de justificación cuando no existe una relación causa-efecto entre la exposición y la enfermedad. No obstante, como la diferencia de incidencias –ya sea diferencia de tasas de incidencia (DTI) o diferencia de riesgos (DR)– puede llegar a indicar diferencias verdaderamente atribuibles a la exposición, estas medidas se siguen usando para estimar la magnitud de proble-

mas de salud pública, aunque ya casi nunca se usan en investigación.

La diferencia de prevalencia (DP), usada en estudios transversales, puede ser en algunas condiciones un estimador aceptable de la diferencia de incidencia, pero sus resultados sólo indican asociación y no causalidad.

Medidas de razón

Estas medidas también cuantifican las discrepancias en la ocurrencia de enfermedad en grupos que difieren en la presencia o no de cierta característica. Como se señaló antes, una razón puede calcularse tanto para dos eventos en una misma población como para un solo evento en dos poblaciones. Las razones que con mayor frecuencia se calculan son del segundo tipo, y se obtienen con la siguiente fórmula:

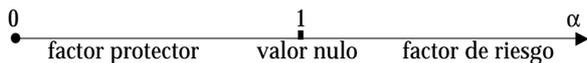
$$\text{Razón} = \frac{\text{medida de frecuencia en un grupo expuesto (E}_i\text{)}}{\text{medida de frecuencia de un grupo no expuesto (E}_o\text{)}}$$

La razón representa cuántas veces más (o menos) ocurrirá el evento en el grupo expuesto al factor, comparado con el grupo no expuesto. El resultado se interpreta de la siguiente forma:

Valor = 1 indica ausencia de asociación, no-asociación o valor nulo.

Valores <1 indica asociación negativa, factor protector.

Valores >1 indica asociación positiva, factor de riesgo.



La interpretación de estas medidas se basa en el hecho de que si se dividen dos cantidades entre sí y el resultado es 1, estas cantidades son necesariamente iguales, y tener o no la característica estudiada es lo mismo, pues ello no afecta la frecuencia de enfermedad. Cuando, en cambio, la razón es mayor de 1, el factor se encuentra asociado positivamente con el riesgo de enfermar y la probabilidad de contraer el padecimiento será mayor entre los expuestos. Si el resultado es menor de 1, el factor protege a los sujetos expuestos contra esa enfermedad.

Conforme el resultado se aleja más de la unidad, la asociación entre el factor y la enfermedad es más fuerte. Un valor de 4 indica que el riesgo de enfermar entre los expuestos es cuatro veces mayor que entre los no expuestos. Asimismo, un valor de 0.25 indicaría que el riesgo de enfermar entre los expuestos es cuatro veces menor que entre los no expuestos.

La incidencia y la mortalidad son las medidas de frecuencia más empleadas en la construcción de las medidas de razón. Con la densidad de incidencia se obtiene la razón de densidad de incidencia (RDI), y con la incidencia acumulada se obtiene la razón de incidencia acumulada (RIA) también llamado riesgo relativo (RR). Ambas medidas –que se obtienen en estudios de cohorte– permiten asumir inferencia etiológica, ya que siempre implican la posibilidad de establecer adecuadamente una relación de temporalidad causal.

Razón de densidad de incidencia

Esta medida es útil para identificar la velocidad con la que se pasa del estado sano al de enfermo según se esté expuesto o no a determinado factor.

Razón de incidencia acumulada o riesgo relativo

Compara el riesgo de enfermar del grupo de expuestos (IA_i) con el riesgo de enfermar del grupo de no expuestos (IA_o). Es útil si lo que se desea es conocer la probabilidad de padecer la enfermedad en función de la exposición, y es la medida que mejor refleja su asociación.

$$RR = \frac{IA_i}{IA_o} = \frac{a/n_i}{c/n_o}$$

donde,

IA_i es la incidencia acumulada o riesgo de enfermar entre los expuestos, y

IA_o es la incidencia acumulada o riesgo de enfermar entre los no expuestos (para observar gráficamente la ubicación de las celdas a, c, n_i y n_o , véase la tabla de 2 X 2).

Razón de prevalencias

La razón de prevalencias (RP) se utiliza en los estudios transversales y se calcula de forma similar a la estimación del RR en los estudios de cohorte. Si la duración del evento que se estudia es igual para expuestos y no expuestos, la RP puede ser buen estimador de la velocidad con la que se pasa del estado sano al de enfermo, pero, en general, esta medida subestima la RDI.

Razón de productos cruzados

La razón de productos cruzados (RPC u OR) se estima en los estudios de casos y controles –donde los su-

jetos son elegidos según la presencia o ausencia de enfermedad, desconociéndose el volumen de la población de donde provienen— por lo que no es posible calcular la incidencia de la enfermedad. La RPC también se conoce con los términos en inglés *odds ratio* (OR) y *relative odds*, y en español como *razón de momios* (RM), *razón de ventaja* y *razón de disparidad*. La RM es un buen estimador de la RDI, sobre todo cuando los controles son representativos de la población de la que han sido seleccionados los casos. La RM también puede ser un buen estimador del RR. Esta medida se calcula obteniendo el cociente de los productos cruzados de una tabla tetracórica:

$$RPC = \frac{a/c}{b/d} = \frac{ad}{bc}$$

donde,

	casos	controles	
Exposición	presente a	b	Total de expuestos (n _e)
	ausente c	d	Total de no expuestos (n _o)
	Total (m _e)	Total (m _o)	Total de sujetos (n)

Al igual que en las medidas anteriores, esta fórmula expresa el caso más sencillo, cuando la exposición y la enfermedad se reportan simplemente como presentes o ausentes.

El resultado se interpreta de la misma forma que en el resto de las medidas de razón. Cuando la OR tiene un valor de 1 (o nulo), el comportamiento del factor es indiferente; si el valor es superior a 1, el factor puede considerarse como de riesgo, y si es inferior a 1 es valorado como factor protector.

Medidas de impacto potencial

La razón de densidad de incidencia, el riesgo relativo y la razón de momios describen la asociación entre la exposición y el evento en términos de la magnitud de la fuerza de la asociación entre estos, información que es muy importante cuando evaluamos la existencia de asociaciones causales. Sin embargo, estas medidas no se pueden traducir fácilmente en el contexto de la salud de la población. ¿Qué tan importante es una exposición? ¿Qué proporción de las enfermedades se pueden atribuir a esta variable? Para poder estimar el efecto de cierta exposición en la población en estudio o en la población blanco se requiere esti-

mar otro tipo de medidas, conocidas como medidas de impacto.

Las principales medidas de impacto potencial son el riesgo atribuible (o fracción etiológica), que se estima cuando el factor de exposición produce un incremento en el riesgo (RR>1), y la fracción prevenible, relacionada con factores que producen una disminución en el riesgo (RR<1).

Riesgo atribuible

Anteriormente era muy frecuente el uso del término fracción etiológica para referirse a este indicador; sin embargo, actualmente se recomienda utilizarlo únicamente para referirse a relaciones causales bien demostradas. El término que se usa con mayor frecuencia y que es más conservador es el riesgo atribuible proporcional. Para esta última medida se han derivado dos dimensiones, el *Riesgo Atribuible Proporcional en el grupo Expuesto* (RAP_{Exp}) y el *Riesgo Atribuible Proporcional en la Población blanco* (RAPP). Ambas medidas son proporciones, por lo que toman valores entre cero y uno e indican la importancia relativa de la exposición al factor en estudio con relación al total de eventos. El RAP_{Exp} tiene interpretación en el ámbito de la población en estudio, mientras que el RAPP expresa la importancia en el ámbito poblacional, o población blanco.

El RAP_{Exp} estima la proporción de eventos en el grupo expuesto que se pueden atribuir a la presencia del factor de exposición. En otras palabras, refleja el efecto que se podría esperar en el grupo expuesto de la población en estudio si se eliminara el factor de riesgo en cuestión. El RAP_{Exp} se puede calcular utilizando la siguiente fórmula:

$$RAP_{exp} = \frac{DI_E - DI_{NE}}{DI_E} = \frac{RDI - 1}{RDI}$$

donde

- DI_E = Densidad de incidencia en expuestos,
- DI_{NE} = Densidad de incidencia en no expuestos, y
- RDI = Razón de densidad de incidencia

El RAP_{Exp} se puede estimar también en estudios donde la medida de frecuencia es la incidencia acumulada, utilizando el riesgo relativo. Además, dado que la razón de momios es un buen estimador de la RDI, el RAP_{Exp} también se puede estimar en los estudios de casos y controles, utilizando la siguiente fórmula:

$$RAP_{Exp} = \frac{RM-1}{RM}$$

Para ilustrar su interpretación y cálculo supongamos que se desea estimar el RAP_{Exp} de los resultados derivados de un estudio de casos y controles sobre tabaquismo y cáncer pulmonar. En el mencionado estudio se documenta una asociación entre el riesgo de cáncer de pulmón y el tabaquismo (RM) de 12.5. El RAP_{Exp} se podría estimar dividiendo 12.5 menos 1 entre 12.5, lo que daría un RAP_{Exp} de 0.92 (o 92%), lo que indicaría que el 92 % de los casos de cáncer pulmonar en el grupo expuesto al tabaco podrían atribuirse a esta exposición. Esto significa que el RAP_{Exp} indica el porcentaje de casos en el grupo expuesto que se podría prevenir si se eliminara la exposición, asumiendo que la exposición es la única causa del evento y que el resto de las causas de cáncer de pulmón se distribuyen de igual manera entre los fumadores (grupo expuesto) y los no fumadores (grupo no expuesto), como se indica en la figura 1. Para el ejemplo anterior indicaría que se podrían prevenir cerca del 92% de los casos de cáncer de pulmón que ocurren en el grupo de fumadores.

El RAPP se puede considerar como una proyección del RAP_{Exp} hacia la población total. En este caso, los resultados obtenidos en el grupo de expuestos se extrapolan hacia la población blanco estimando el impacto de la exposición a nivel poblacional. Siguiendo el ejemplo anterior, la estimación del RAPP nos indicaría cuántos casos de cáncer de pulmón en la población total son atribuibles al tabaco o se podrían evitar supo-

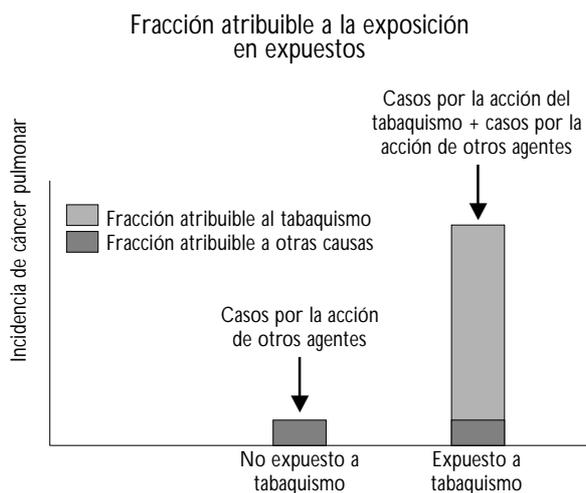


FIGURA 1. REPRESENTACIÓN HIPOTÉTICA DE UN ESTUDIO DE COHORTE PARA EVALUAR EL EFECTO DEL TABAQUISMO SOBRE EL RIESGO DE DESARROLLAR CÁNCER DE PULMÓN

niendo que se eliminara el tabaquismo en la población general. EL RAPP se estima ponderando el RAP_{Exp} de acuerdo con la proporción de sujetos expuestos en la población blanco. El RAPP se puede estimar utilizando la siguiente fórmula:

$$RAPP = \frac{Pe (RDI-1)}{Pe (RDI-1)+1}$$

Al igual que en el caso anterior, el RAPP se puede estimar para estudios de cohorte, donde se estima la incidencia acumulada, o en estudios de casos y controles, donde se estima la razón de momios. En este último caso, se puede utilizar la prevalencia de exposición en los controles para estimar la prevalencia en la población blanco o población de referencia. En el estudio antes mencionado sobre tabaquismo y cáncer pulmonar se observó una prevalencia del 28.5 de tabaquismo en el grupo control. Dado que la serie de controles se puede considerar como representativa de la población base, en este estudio se podría estimar directamente el RAPP, lo que daría una fracción de 0.76. Esta última cifra indicaría que, en la población blanco, el 76% de los casos de cáncer pulmonar pueden ser atribuidos al tabaquismo, asumiendo que el tabaquismo es su única causa.

Mediante el cálculo del RAP_{Exp} y del RAPP es posible identificar diversos escenarios:

- Con un RR alto y una prevalencia de expuestos alta, la reducción del riesgo de enfermedad puede considerarse como de alto impacto.
- Cuando el RR es bajo y la prevalencia de expuestos es alta, la supresión del factor de riesgo posee un impacto moderado, pero notable entre los expuestos.
- Cuando el RR es alto pero la prevalencia de expuestos es baja, la eliminación del factor de riesgo tiene un impacto relativamente bajo tanto entre la población blanco como entre los expuestos, y
- Cuando el RR es bajo y la prevalencia de expuestos también es baja, la eliminación del factor de riesgo no es una prioridad en salud pública, ya que su impacto en la población blanco y en los expuestos sería irrelevante.

Fracción prevenible

Esta medida se aplica cuando a partir de las medidas de asociación se obtienen factores protectores o negativos ($RR < 1$). También existen dos modalidades: fracción prevenible poblacional y fracción prevenible entre expuestos.

La fracción prevenible poblacional es la proporción de todos los casos nuevos que potencialmente podrían haber ocurrido entre la población general en un determinado periodo en ausencia de una exposición protectora específica; o bien, es la proporción de casos potenciales que serían realmente prevenibles o evitados si existiera la exposición entre la población.

Finalmente, la fracción prevenible para los expuestos es la proporción de casos nuevos entre los expuestos que potencialmente podría haber ocurrido en un determinado periodo en ausencia de una exposición particular. Es decir, es la proporción de casos expuestos potenciales que realmente se evitarían si la población se expusiera al factor protector.

Bibliografía

1. MacMahon B, Pugh TF. Epidemiology. Principles and methods. Boston: Little Brown & Co., 1970.
2. Kleimbaum DG, Kupper LL, Morgenstern H. Epidemiologic research. New York (NY): Van Nostrand Reinhold Co., 1982.
3. Rothman JK. Modern epidemiology. Boston: Little Brown & Co., 1986.
4. Ahlbom A, Norell S. Fundamentos de epidemiología. Madrid: Siglo XXI Editores 1987.
5. Gordis L. Epidemiology. Philadelphia: W.B. Saunders Co. 1995.
6. Jenicek M. Epidemiología: la lógica de la medicina moderna, Barcelona: Masson, 1996.
7. Martínez NF, Antó JM, Castellanos PL, Gili M, Marset P, Navarro V. Salud pública. Madrid: McGraw-Hill-Interamericana, 1998.

Diseño de estudios epidemiológicos

Mauricio Hernández-Avila, Ph.D.,⁽¹⁾ Francisco Garrido-Latorre, M. en C.,⁽²⁾
Sergio López-Moreno, M.C.⁽²⁾

Los principales objetivos de la investigación epidemiológica son, por un lado, describir la distribución de las enfermedades y eventos de salud en poblaciones humanas y, por otro, contribuir al descubrimiento y caracterización de las leyes que gobiernan o influyen en estas condiciones. La epidemiología no representa un dominio del conocimiento claramente delimitado como el que tienen otras ciencias médicas como, por ejemplo, la bioquímica o la fisiología. La epidemiología se emplea en las distintas ramas de la medicina como una herramienta para el estudio de diferentes enfermedades o eventos relacionados con la salud, especialmente cuando se busca evaluar la repercusión de éstos en el ámbito de la población. Así, es posible encontrar aplicaciones de la epidemiología tanto para definir los mecanismos de transmisión de una enfermedad infecciosa como para evaluar la respuesta médica organizada para contender con la misma o para evaluar el impacto, en el ámbito poblacional, del desarrollo de resistencia a los diferentes tratamientos. El principal objetivo de la epidemiología es desarrollar conocimiento de aplicación a nivel poblacional (cuadro I), y por esta razón es considerada como una de las ciencias básicas de la salud pública.

La información necesaria para cumplir con los objetivos de la investigación epidemiológica, ya sea de tipo descriptivo o analítico, se deriva de la experimentación con seres humanos o, más frecuentemente, de la observación directa de grupos poblacionales. A pesar de que para la epidemiología es de interés princi-

pal derivar conocimiento de aplicación poblacional, raramente estudia a la población en su conjunto. Por ello, tanto para la experimentación con voluntarios como para la observación de grupos poblacionales es necesario desarrollar estrategias muestrales y de medición que permitan, en primera instancia, estudiar subgrupos de la población y, en un segundo término, hacer extrapolaciones del conocimiento generado hacia el total de la población. La validez de la información derivada de los estudios epidemiológicos depende de manera importante de lo adecuado y apropiado de los métodos utilizados. El reconocimiento de la importancia que tienen los aspectos metodológicos como un

Cuadro I PRINCIPALES USOS DE LA EPIDEMIOLOGÍA EN SALUD PÚBLICA

- Identificación de la historia natural de las enfermedades
- Descripción de la distribución, frecuencia y tendencias de la enfermedad en las poblaciones
- Identificación de la etiología y los factores de riesgo para la aparición y desarrollo de enfermedades
- Identificación y explicación de los mecanismos de transmisión y diseminación de las enfermedades
- Identificación de la magnitud y tendencias de las necesidades de salud
- Identificación de la magnitud, vulnerabilidad y formas de control de los problemas de salud
- Evaluación de la eficacia y efectividad de las intervenciones terapéuticas
- Evaluación de la eficacia y efectividad de la tecnología médica
- Evaluación del diseño y ejecución de los programas y servicios de salud

(1) Centro de Investigación en Salud Poblacional, Instituto Nacional de Salud Pública (INSP), México.

(2) Centro de Investigación en Sistemas de Salud, INSP, México.

eje necesario para el desarrollo y avance del conocimiento epidemiológico ha propiciado que se asuma como un objetivo mismo de la epidemiología el desarrollo y el estudio de nuevos métodos de aplicación en el campo. Esto, sin duda, ha contribuido de manera importante a mejorar la calidad y la validez del conocimiento derivado de estudios epidemiológicos y a consolidar a la epidemiología como una ciencia básica necesaria para el avance de la salud pública y de la medicina.

En este trabajo se hace una revisión y una actualización de los principales diseños epidemiológicos utilizados en investigaciones de tipo epidemiológico para conformar los grupos poblacionales de estudio. Asimismo, se propone un esquema de clasificación de dichas estrategias, conforme a una escala ordinal en términos de la evidencia que aporta cada diseño para establecer relaciones causa efecto entre dos variables de interés.

En diferentes libros de texto y trabajos que abordan la aplicación y desarrollo de los métodos epidemiológicos se han propuesto diversos esquemas para agrupar y caracterizar a los distintos tipos de estudio, los cuales se han clasificado de acuerdo con: a) el tipo de asignación de la exposición o variable en estudio; b) el número de mediciones que se realiza en cada sujeto de estudio para verificar la ocurrencia del evento o cambios en la exposición; c) la temporalidad del inicio de la exposición o de la ocurrencia del evento; d) los criterios utilizados para la selección de la población a estudiar, y e) la unidad de análisis donde se mide el evento en estudio (cuadro II).¹⁻⁵

En términos de causalidad la asignación de la exposición* es el criterio más importante de clasificación y divide a los estudios epidemiológicos en tres tipos: a) experimentales, cuando el investigador controla la exposición y utiliza la aleatorización como método de asignación; b) pseudo-experimentales (o de intervención no aleatorizados), cuando el investigador controla la exposición pero no utiliza procedimientos de aleatorización en la asignación, y c) no-experimentales u observacionales, cuando la exposición ocurre sin la participación del investigador y de acuerdo con variables que están fuera de control del investigador.

De acuerdo con el número de mediciones que se realiza en cada sujeto de estudio para medir la ocurrencia del evento[‡] o cambios en la variable de exposición a lo largo del tiempo, los estudios se pueden dividir en: a) longitudinales, cuando se realizan al menos dos mediciones: la medición basal para determinar el estado inicial y una subsecuente para

* En este trabajo utilizaremos el término exposición para referirnos a la variable en estudio. Utilizaremos exposición como un término de significado amplio, que puede abarcar desde la exposición a una bacteria o una sustancia tóxica hasta la exposición a un suplemento nutricional, una vacuna, un programa de salud o un estilo de vida.

‡ Utilizaremos el término evento para referirnos a la variable respuesta o cambio que se espera detectar con relación a la exposición. Se utiliza también de manera amplia y puede referirse tanto a la ocurrencia de una enfermedad, como al cambio de estado o cambio promedio en una variable continua; por ejemplo, cambio en las concentraciones de colesterol sérico o seroconversión después de aplicada una vacuna.

Cuadro II
CLASIFICACIÓN DE LOS ESTUDIOS EPIDEMIOLÓGICOS

Tipo de estudio	Asignación de la exposición	Número de observaciones por individuo	Criterios de selección de la población en estudio	Temporalidad	Unidad de análisis
Ensayo aleatorizado	Aleatoria	Longitudinal	Ninguno	Prospectivo	Individuo
Pseudo-experimentales	Por conveniencia	Longitudinal	Ninguno	Prospectivo	Individuo
Cohorte	Fuera de control del investigador	Longitudinal	Exposición	Prospectivo o retrospectivo	Individuo
Casos y controles	Fuera de control del investigador	Longitudinal o transversal	Evento	Prospectivo o retrospectivo	Individuo
Estudio de encuesta	Fuera de control del investigador	Transversal	Ninguno	Retrospectivo	Individuo
Ecológico o de conglomerado	Fuera de control del investigador	Longitudinal o transversal	Ninguno	Retrospectivo	Grupo (o población)

determinar la ocurrencia del evento, y b) transversales, cuando se realiza una sola determinación en los sujetos de estudio y se evalúan de manera concurrente la exposición y el evento de interés.

En términos de causalidad existe una diferencia importante entre estos dos tipos de estudio, ya que en los longitudinales es posible verificar que la exposición antecede a la ocurrencia del evento, con lo que se cumple el principio temporal de causalidad –la causa antecede al efecto; en tanto que en los transversales resulta imposible verificar este tipo de relaciones cuando se estudian exposiciones que cambian con el tiempo. Los estudios transversales sí pueden proporcionar información valiosa cuando se estudian factores que no varían (como el sexo y la carga genética) o exposiciones únicas que no cambian con el tiempo (por ejemplo, el caso de la población expuesta a una bomba atómica).

El criterio de temporalidad en la ocurrencia del evento se utiliza para distinguir entre los estudios retrospectivos y prospectivos. El punto de referencia para esta clasificación es la ocurrencia del evento de interés (la variable respuesta). Si al inicio del estudio, el evento investigado ya ocurrió y el investigador planea reconstruir su ocurrencia en el pasado utilizando registros o entrevistando a los mismos sujetos de estudio, se considera que el estudio es retrospectivo. Si la ocurrencia del evento se registra durante el estudio, es decir, si los sujetos de estudio están libres del evento de interés al iniciar su participación en el estudio, el diseño se considera de tipo prospectivo. En general, podríamos afirmar que los estudios prospectivos tienen mayor puntaje en la escala de causalidad, dado que en este tipo de estudios se pueden diseñar instrumentos para la medición y registro del evento que aseguren la calidad de las mediciones. En cambio, en los estudios retrospectivos la calidad de medición y registro del evento dependen con frecuencia de instrumentos que no fueron diseñados de manera expresa para observar el evento en cuestión ni para responder a los objetivos de la investigación. Los estudios que incluyen eventos que ocurrieron antes de iniciar la investigación y eventos evaluados de manera prospectiva son referidos en algunos textos como mixtos o ambispectivos.³

La selección de los participantes en el estudio se puede llevar a cabo de acuerdo con la exposición, el evento o sin considerar ninguna de estas características de los sujetos elegibles para el estudio. La selección con base en estos atributos se utiliza con frecuencia para distinguir entre los diferentes estudios epidemiológicos de tipo observacional. Cuando los sujetos son seleccionados con base en la exposición, es decir, se

elige un grupo expuesto y uno no expuesto, en los que posteriormente se determinará la ocurrencia del evento, se considera que se trata de un estudio de cohorte. En contraste, cuando se selecciona a los participantes con base en el evento de estudio, es decir, se elige de manera independiente un grupo de sujetos que tienen el evento de interés (casos) y un grupo de sujetos que no lo tienen (controles) y en estos grupos se determina la exposición, entonces hablamos de un estudio de casos y controles. Finalmente, cuando la selección es indistinta de la ocurrencia de la exposición o del evento, es decir, los sujetos de estudio son seleccionados sin considerar información sobre la exposición o el evento, y la ocurrencia de éstos se determina una vez conformada la población en estudio, entonces los estudios se denominan de encuesta, aunque algunos textos también los clasifican como de tipo transversal.² La característica principal de esta última estrategia de muestreo es que la evaluación de la exposición y de la ocurrencia del evento se hacen de manera simultánea. En términos de causalidad se puede decir, de manera general, que los estudios de cohorte tienen mayor peso que los estudios de casos y controles y que los transversales. Sin embargo, es importante mencionar que los estudios de casos y controles, con ciertas características, pueden ser tan informativos (en términos de causalidad) como un estudio de cohorte.

Por último, la unidad de análisis se ha utilizado para clasificar a los estudios en ecológicos (de conglomerados) e individuales. A diferencia de los estudios individuales, en los que la unidad de análisis es el individuo y se cuenta con al menos una medición de cada uno de ellos, en los estudios ecológicos la unidad de análisis es un grupo (por ejemplo, un país o una región) y se cuenta con el promedio de eventos o de exposición para el grupo, desconociéndose a nivel individual la condición de evento o exposición para cada individuo de la población. Este tipo de estudios conlleva problemas importantes en su interpretación ya que, dado que los datos se encuentran agrupados, no es posible corregir por diferencias en otras variables (posibles variables confusoras*) que pudieran explicar los resultados observados. Por esta razón, los resultados de este tipo de estudios tienen el peso más bajo en la escala de causalidad.

A continuación describiremos brevemente las características de las principales estrategias epidemiológicas utilizadas para estudiar grupos poblacionales.

* Una variable confusora está asociada con la exposición y el evento en estudio, por lo que una diferencia en su distribución en los grupos que se comparan puede distorsionar los resultados.

Ensayos epidemiológicos aleatorizados

Los ensayos epidemiológicos aleatorizados son estudios experimentales que, cuando se llevan a cabo de manera adecuada, proporcionan el máximo grado de evidencia para confirmar la relación causa efecto entre la exposición y el evento en estudio. Se distinguen de los estudios observacionales (no experimentales) porque el investigador tiene control sobre la asignación de la exposición y porque ésta se lleva a cabo mediante un proceso aleatorio. Además, dado que se trata de estudios longitudinales y prospectivos, y en los que la unidad de análisis es el individuo, es posible prevenir la introducción de sesgos y lograr altos índices de validez. En este tipo de estudios es posible minimizar la ocurrencia de sesgos mediante tres procedimientos, los cuales tienen como objetivo garantizar la comparabilidad: a) de intervenciones (o de la exposición), b) de los grupos en estudio, y c) de los procedimientos para recopilar la información obtenida de la población en estudio.⁵

La comparabilidad de intervenciones es indicativa de la "pureza" del contraste entre los grupos experimentales, y se refiere a que la única diferencia entre las exposiciones que se comparan es la parte o sustancia activa de la exposición en estudio. El concepto de *comparabilidad* de intervenciones se puede entender como una extensión en epidemiología del concepto de *efecto placebo* observado de los ensayos clínicos. Este concepto se deriva del hecho de que el efecto de un medicamento es el resultado de la suma de dos componentes: uno, causado por la sustancia activa del medicamento y, otro, producido por la acción de recibir atención o por el componente psicológico asociado con la idea de recibir un medicamento. La idea de identificar una intervención exactamente igual a la que se pretende probar, pero sin la sustancia activa, es precisamente eliminar de la comparación el efecto atribuido al placebo y de esta manera estimar únicamente la diferencia atribuible a la sustancia activa o intervención en cuestión.

La comparabilidad de poblaciones se logra cuando los grupos experimentales que reciben las diferentes intervenciones son similares en todas y cada una de las características que pudieran tener relación con el evento en estudio o con la manera en que actúa la exposición. En términos epidemiológicos este concepto indica la ausencia de factores de confusión o modificación de efecto. En una situación ideal, la comparabilidad de poblaciones se podría lograr observando a los mismos sujetos en estudio en condiciones experimentales diferentes. Sin embargo, las condiciones necesarias para que esto ocurra raramente se

consiguen en el contexto epidemiológico. Como alternativa para lograr la comparabilidad de poblaciones se ha utilizado la aleatorización. Mediante este proceso se deja al azar la distribución de los sujetos en los diferentes grupos experimentales y se espera que en promedio los grupos tengan características comparables. Es importante mencionar que la comparabilidad que se obtiene con la asignación aleatoria a los grupos experimentales depende del tamaño muestral, y que la simple aleatorización no garantiza completamente que las variables se distribuirán homogéneamente en los distintos grupos de intervención. Dado que las variables se distribuyen al azar será siempre necesario verificar el resultado de la aleatorización, debido a que existe la posibilidad de no funcionar adecuadamente. Esto se puede llevar a cabo comparando la distribución de las variables medibles en los diferentes grupos experimentales. La distribución homogénea de estas variables entre los grupos experimentales sería indicativa de éxito en la aleatorización.

Por último, la comparabilidad de la información se logra cuando se utilizan exactamente los mismos métodos de seguimiento y de medición en todos los participantes en el estudio. Una manera de lograr esto es cegando o enmascarando a los evaluadores y a los participantes en el estudio respecto a la condición de exposición. Si el o los evaluadores desconocen la pertenencia de los sujetos que evalúan, con relación a los grupos experimentales, es probable que la medición no se vea afectada por esta información. La "ignorancia", en este caso, hace que los grupos sean tratados en igual forma. Si se pusiera especial interés en determinar la ocurrencia del evento en el grupo experimental, los resultados podrían ser erróneos dada la diferencia en los procedimientos utilizados. Este concepto abarca tanto la calidad de la información como la proporción de los sujetos en estudio que se pueden perder durante el seguimiento en cada grupo experimental. La pérdida diferencial de participantes puede ocasionar errores importantes. Algunos textos refieren este último problema dentro de los sesgos de selección.²

Los pasos para la realización de este tipo de estudios (figura 1) incluyen la definición de la población blanco, que es aquella a la cual se pretenden extrapolar los resultados del estudio. Al aplicar los criterios de inclusión en el estudio se define la población elegible y de esta población se seleccionan los participantes en el estudio; esto último se puede llevar a cabo ya sea mediante el reclutamiento de voluntarios o mediante la selección de una muestra representativa de la población blanco. Es importante mencionar que siempre que se trabaja con poblaciones humanas se tendrá un grupo

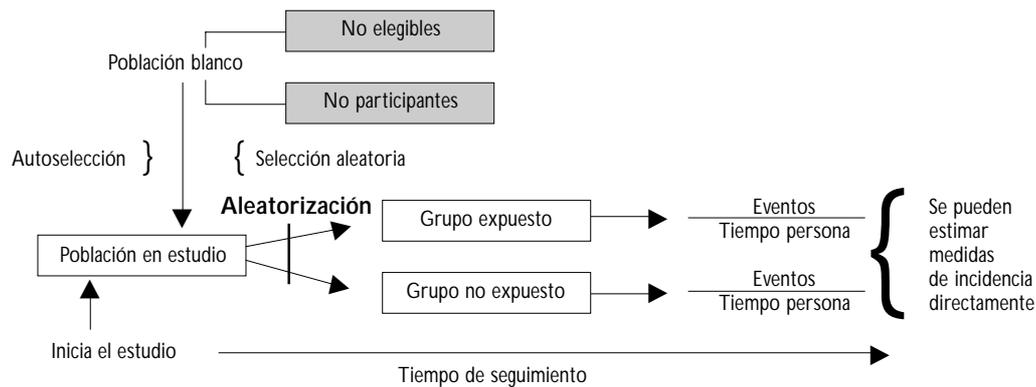


FIGURA 1. ENSAYO CLÍNICO ALEATORIZADO

autoseleccionado de la población que corresponde a aquellos sujetos que otorgan el consentimiento informado para participar en el proceso experimental. La autoselección o la selección de un grupo representativo de la población blanco tiene un alto grado de importancia en términos de la validez externa* de los resultados, ya que éstos serán aplicables a la población blanco en la medida de que la población en estudio represente adecuadamente a esta población. Una vez identificados los participantes en el estudio y que éstos han dado su consentimiento para participar en el proceso experimental, los sujetos se asignan de manera aleatoria a los grupos experimentales (figura 1). Posteriormente, los sujetos se siguen en el tiempo con el fin de documentar la ocurrencia del evento en estudio y posibles cambios en otras covariables de interés.

Algunos textos⁴ mencionan la existencia de un grupo intermedio entre los estudios experimentales y no experimentales; éstos son llamados quasi-experimentales. En estos diseños el investigador controla la asignación de la exposición, sin embargo, esta asignación no se hace de manera aleatoria. Este tipo de estudios de intervención incluye los diseños tipo *antes* y *después* o de *comparación concurrente*.

En los estudios observacionales (no experimentales) la asignación de la exposición ocurre sin la participación del investigador. En este tipo de diseños es común que la exposición ya haya ocurrido al iniciar el estudio, y que ésta se haya dado por algún factor independiente y fuera del procedimiento experimental.

* Validez externa: se refiere a la generalización de los resultados observados.

Estudios de cohorte

Entre los estudios observacionales, este tipo de diseño representa lo más cercano al diseño experimental y también tiene un alto valor en la escala de causalidad, ya que es posible verificar la relación causa efecto correctamente en el tiempo. Sin embargo, dado que se trata de estudios observacionales tienen la importante limitación de que la asignación de la exposición no es controlada por el investigador ni asignada de manera aleatoria, por lo que no es posible controlar completamente las posibles diferencias entre los grupos expuesto y no expuesto en relación con otros factores asociados con la ocurrencia del evento.

La selección de los participantes con base en la exposición caracteriza a los estudios de cohorte (figura 2). En este tipo de diseño epidemiológico la población en estudio se define a partir de la exposición y debe estar conformada por individuos en riesgo de desarrollar el evento en estudio. Los sujetos de estudio se seleccionan de la población que tiene la exposición de interés y de grupos poblacionales comparables, pero que no tienen la exposición. Una vez conformada la población en estudio ésta se sigue en el tiempo y se registra en ella la ocurrencia del evento de interés o variable respuesta.

El diseño de cohorte es especialmente eficiente para estudiar exposiciones raras o poco frecuentes; por ejemplo, las exposiciones ocupacionales que se presentan en poblaciones muy reducidas de trabajadores. En general, cuando se requiere evaluar los riesgos asociados con algún tipo particular de ocupación, se selecciona a grupos ocupacionales y se establece un grupo de comparación (no expuesto) tomado de la población general o, incluso, ubicado en la misma in-

dustria o en otra similar, pero no en contacto con la exposición en estudio.

Los estudios de cohorte también se utilizan regularmente para estudiar exposiciones que se presentan con una alta frecuencia en la población general. Para este tipo de exposiciones es común seleccionar aleatoriamente grupos representativos de la población que posteriormente se clasifican de acuerdo con la exposición; la cohorte (población en estudio) queda conformada con los participantes que no tienen el evento en estudio y que están en riesgo de desarrollar el evento, posteriormente este grupo se sigue en el tiempo con el fin de registrar la ocurrencia del evento. El procedimiento antes descrito se refiere a un estudio prospectivo (figura 2); sin embargo, los estudios de cohorte también pueden ser retrospectivos (figura 3). En este tipo de estudios, se inicia con la definición de los grupos expuesto y no expuesto en algún punto en el pasado y posteriormente se reconstruye la experiencia de la cohorte en el tiempo, identificando a los sujetos en el tiempo actual (cuando se realiza el estudio) y evaluando si a la fecha referida ya han desarrollado el evento de interés.

En su concepción más simple (figuras 2 y 3), un estudio de cohorte consiste en seleccionar un grupo expuesto y otro no expuesto de la población elegible, observarlos durante un tiempo determinado y compararlos en términos de la ocurrencia del evento de interés. La validez de la comparación dependerá de que no existan diferencias (aparte de la exposición) entre los grupos expuesto y no expuesto. Cualquier diferencia con relación a una tercera variable entre el grupo expuesto y no expuesto que esté relacionada con la ocu-

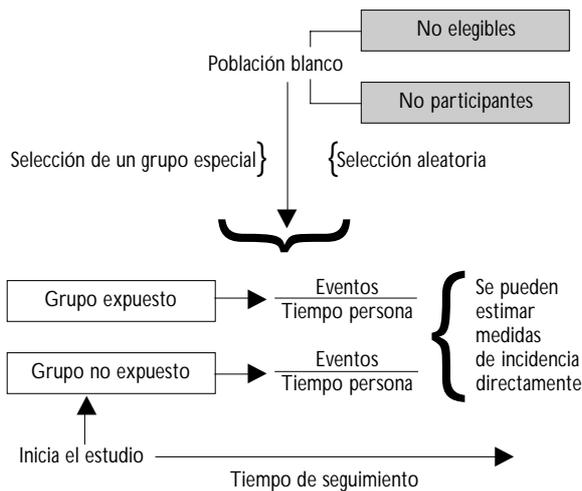


FIGURA 2. ESTUDIOS DE COHORTE PROSPECTIVA

rrencia del evento podría distorsionar los resultados sobre la asociación real entre la exposición y el evento.

Los estudios de cohorte son difíciles de realizar y, además, son costosos. Se considera que este tipo de diseño es poco eficiente para el estudio de enfermedades raras, ya que para registrar un número adecuado de eventos se requeriría un número muy grande de participantes y de tiempos prolongados de seguimiento (cuadro III).

Estudios de casos y controles

Durante los últimos años se ha desarrollado una estrategia de selección sobre la base del evento en estudio,

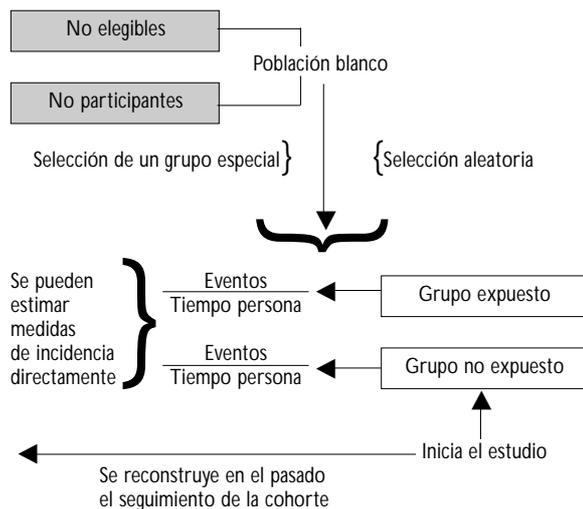


FIGURA 3. ESTUDIOS DE COHORTE RETROSPECTIVA

Cuadro III

VENTAJAS Y DESVENTAJAS DE LOS ESTUDIOS DE COHORTE

Ventajas

- Más cercanos a un experimento
- La relación temporal causa efecto es verificable
- Se pueden estimar medidas de incidencia
- Eficientes para evaluar exposiciones poco frecuentes
- Se pueden estudiar varios eventos
- Se pueden fijar criterios de calidad en la medición del evento
- Bajo riesgo de sesgo de selección (en especial en estudios prospectivos)

Desventajas

- Cuando se trata de eventos poco frecuentes la complejidad y el costo pueden aumentar considerablemente, ya que requiere estudiar y seguir un número grande de participantes
- Son estudios difíciles de realizar

este tipo de muestreo se ha denominado en la literatura epidemiológica como estudios de casos y controles,¹⁻⁴ o casos y referentes.⁵ La característica principal de este diseño epidemiológico es que el criterio de selección de la población en estudio se basa en la presencia (casos) o ausencia (controles) del evento en estudio y en que es el investigador quien fija el número de eventos a estudiar, así como el número de sujetos sin evento (controles) que se incluirán como población de comparación o referencia. De esta manera la población en estudio queda compuesta por un grupo de sujetos con el evento en estudio (casos) y un grupo de sujetos sin el evento (controles), posteriormente estos grupos se comparan en términos de la exposición que tuvieron al factor causal en estudio. A diferencia de los estudios de cohorte en los que se iguala la proporción de sujetos expuestos y no expuestos en la población de estudio, en este diseño se tiende a igualar la población en estudio en términos de los sujetos con y sin el evento en estudio.

En general, este tipo de estudios se lleva a cabo utilizando sistemas de registro que permiten identificar fácilmente a los sujetos que desarrollaron el evento en estudio (casos). Los sistemas de registro tradicionalmente utilizados incluyen centros hospitalarios o registros con base poblacional, como son los de neoplasias o malformaciones congénitas. El común denominador en este tipo de estudios es la utilización de un sistema que permite concentrar información sobre la población que presenta el evento en tiempos relativamente cortos y, en general, sin la necesidad de invertir cuantiosos recursos económicos, como los que se podrían requerir para concentrar el mismo número de eventos en el contexto de un estudio de cohorte. Mediante un mecanismo de selección, independiente al utilizado para los casos, se selecciona como grupo de contraste una serie de sujetos que no han desarrollado el evento en cuestión. La comparación directa de sujetos con (casos) y sin (controles) el evento, con relación al antecedente de exposición se utiliza con frecuencia para establecer asociaciones entre la exposición y el evento en estudio. Sin embargo, a pesar de que frecuentemente se utiliza esta comparación de manera automática para establecer asociaciones causales entre la exposición y el evento, es muy importante recalcar que para que ésta pueda considerarse como válida, se requiere del cumplimiento de ciertas condiciones sobre el origen de los casos y los controles. Entre otras cosas se requiere, por ejemplo, que los casos y controles tengan su origen en la misma base poblacional, que los controles representen de manera adecuada la población de donde provienen los casos y que cumplan con la condición de que si desarrollaran el evento

en estudio, hubieran sido reclutados en la población de estudio como casos. Es claro que a menos de que el estudio se desarrolle al interior de una cohorte bien definida no será posible verificar el cumplimiento de las condiciones anteriormente mencionadas; sin embargo, el cumplimiento teórico de las mismas se podrá analizar en la medida que pensemos en los estudios de casos y controles como una alternativa para estudiar una cohorte imaginaria, la cual es posible delinear en tiempo y espacio, pero que es estudiada por medio de la selección de una muestra representativa de los eventos (casos) y de los individuos que no desarrollan el evento de interés (controles). El supuesto del origen común de los casos y controles se cumple si ambos se originan de la misma cohorte y representan tanto a los eventos como a la población en riesgo que no desarrolló el evento.

Los estudios de casos y controles son frecuentemente realizados de manera retrospectiva (figura 4), por lo que no tienen una relación causal perfecta, ya que el evento se evalúa antes que la causa y no siempre se puede inferir que la causa antecedió al evento. La naturaleza retrospectiva de los estudios de casos y controles hace que este tipo de estudios sea particularmente vulnerable a la introducción de errores en los procesos de selección o de recolección de la información. Por esta razón, este tipo de estudios se ha considerado tradicionalmente con un puntaje bajo en la escala de causalidad. Sin embargo, en ciertas ocasiones también es posible realizar este tipo de estudio de manera prospectiva y en ese contexto este tipo de estudios tiene mayor peso en la escala de causalidad. Las principales ventajas y desventajas de este tipo de estudios se describen en el cuadro IV.

Estudios transversales

Finalmente, la población en estudio puede ser seleccionada de manera aleatoria sin considerar la exposición o el evento como criterios de selección. Este tipo de estudio se ha denominado como de *encuesta* o *transversal* en los diferentes textos,¹⁻⁴ y se distingue porque se indaga sobre la presencia de la exposición y la ocurrencia del evento una vez conformada la población en estudio, y porque sólo se hace una medición en el tiempo en cada sujeto de estudio (figura 5). El número de eventos así como la proporción de sujetos con la exposición están determinados por la frecuencia con que ocurren éstos en la población elegible y, por lo tanto, quedan fuera del control del investigador. Esto último contrasta con los estudios de cohorte o de casos y controles en los que el investigador puede fijar con anterioridad, ya sea la proporción de expuestos

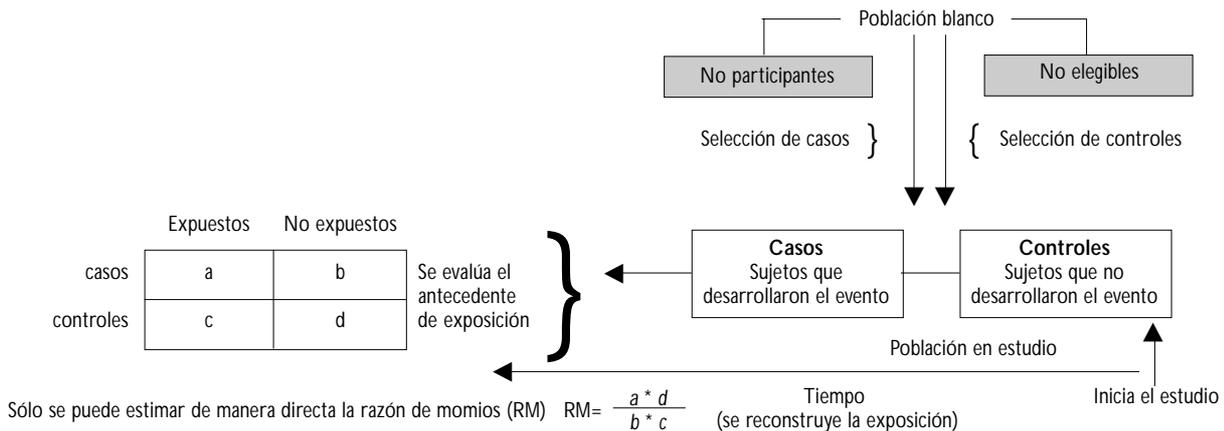


FIGURA 4. ESTUDIOS DE CASOS Y CONTROLES

(estudio de cohortes) o la prevalencia del evento en la población en estudio (estudio de casos y controles). Los estudios transversales se caracterizan porque sólo se hace una medición en el tiempo en cada sujeto de estudio.

Este tipo de estrategia comparte muchas de las limitaciones de los estudios de casos y controles (cuadro V), son retrospectivos y se basan en el estudio de casos prevalentes, los que en general representan a los sujetos con periodos de mayor sobrevivencia o duración de la enfermedad (evento). Cualquier factor que esté relacionado con la duración del evento y la exposición puede ser una fuente de error en este tipo de estudios. Por lo anterior, los estudios de encuesta tienen

una escala baja en términos de causalidad y deben ser interpretados con mucha cautela. Sin embargo, son estudios útiles para la planeación de los servicios de salud y para caracterizar el estado de salud de la población en un punto en el tiempo.

Estudios ecológicos o de conglomerado

En general, en los diferentes tipos de estudios epidemiológicos que hemos mencionado, la unidad de análisis es el sujeto que compone la población en estudio, y es en éstos en los que se mide la exposición y se registra la ocurrencia del evento en el estudio. Sin embargo, en ocasiones la unidad de análisis puede no ser el individuo, sino un conjunto o conglomerado de individuos miembros de la población en estudio. Los conglomerados pueden estar constituidos por grupos poblacionales, comunidades, regiones, o países. La característica principal de este tipo de estudios es que se cuenta con información sobre la exposición o el evento para el conglomerado en su totalidad, desconociéndose la información a nivel individual para cada uno de los miembros del conglomerado. En este tipo de estudios es común asignar la misma exposición (exposición promedio) a todo el conglomerado, ignorando o no considerando la variación individual de cada miembro del conglomerado. Lo mismo sucede con la medición del evento; dado que sólo se cuenta con el número de eventos registrados para el conglomerado, no podemos discernir sobre los eventos que se presentaron en los sujetos expuestos de los que ocurrieron en los no expuestos, al interior de cada conglomerado, por lo que es frecuente atribuir la totalidad de eventos –sin una veri-

Cuadro IV
VENTAJAS Y DESVENTAJAS DE LOS ESTUDIOS DE CASOS Y CONTROLES

Ventajas

- Eficientes para el estudio de enfermedades raras
- Eficientes para estudiar enfermedades con periodos de latencia o inducción prolongados
- Se pueden estudiar varias exposiciones simultáneamente
- En comparación con los estudios de cohorte son menos costosos y se pueden realizar en menor tiempo

Desventajas

- No se pueden estimar de manera directa medidas de incidencia o prevalencia
- Susceptibles a sesgos de selección
- Se puede presentar causalidad reversa
- Problemas para definir población fuente de los casos
- Problemas para medir adecuadamente exposición

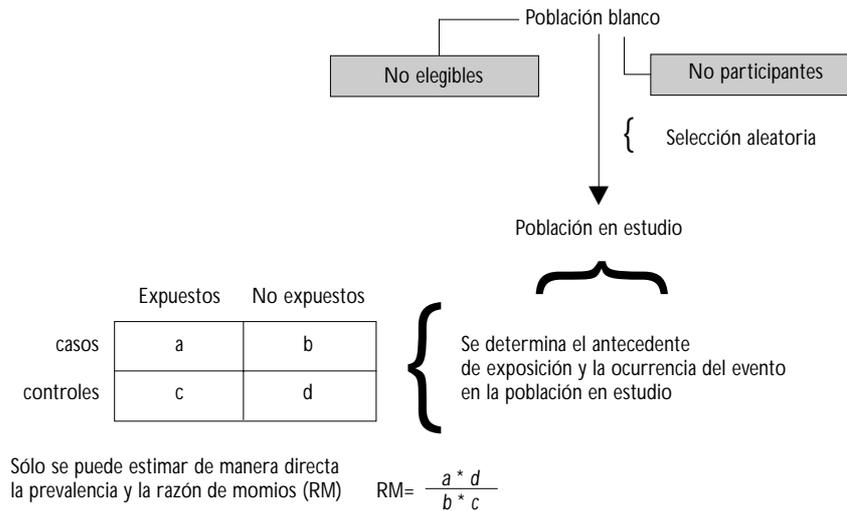


FIGURA 5. ESTUDIO TRANSVERSAL O DE ENCUESTA

ficación real- a la exposición promedio que se registró al interior del conglomerado.

Los estudios de conglomerados (ecológicos) (cuadro VI) permiten estudiar grandes grupos poblacionales en poco tiempo y con un costo relativamente muy bajo, ya que en general utilizan estadísticas existentes recolectadas con otros fines. Sin embargo, dado que tienen el puntaje más bajo en la escala de causalidad deben ser considerados únicamente para sugerir hipótesis, que tendrán necesariamente que ser verificadas mediante otros estudios más rigurosos. Los principales problemas de este tipo de estudios son que se ignora la variabilidad individual de los integrantes de los conglomerados y que no es posible corregir por diferencias en otras variables que pudieran estar también asociadas con la exposición y el evento en estudio.

Conclusiones

Hemos revisado brevemente las principales estrategias de muestreo o diseños de investigación utilizados en los estudios epidemiológicos. Sin duda el ensayo aleatorizado es la estrategia que se reconoce como más poderosa para establecer relaciones causa efecto. Sin embargo, frecuentemente no es posible utilizar este tipo de diseño en estudios epidemiológicos, en particular cuando evaluamos los efectos de la exposición a sustancias tóxicas, que pueden ser de uso común pero que su aplicación deliberada a un grupo experimental sería inaceptable, en estas circunstancias tendremos que

Cuadro V
VENTAJAS Y DESVENTAJAS DE LOS ESTUDIOS TRANSVERSALES

Ventajas

- Eficientes para estudiar la prevalencia de enfermedades en la población
- Se pueden estudiar varias exposiciones
- Son poco costosos y se pueden realizar en poco tiempo
- Se puede estimar la prevalencia del evento

Desventajas

- Problemas para definir y medir exposición
- Sesgos de selección
- Sesgos por casos prevalentes
- La relación causa efecto no siempre es verificable
- Sobrerrepresentación de enfermos con tiempos prolongados de sobrevida o con manifestaciones con mejor curso clínico
- Se puede presentar causalidad débil

basarnos en diseños de tipo observacional y realizar estudios en poblaciones que han estado expuestas por alguna circunstancia, ya sea de tipo ocupacional, ambiental o accidental.

Un ejemplo de la situación antes mencionada es la exposición al DDT. Pese a que se consideraría inaceptable la exposición intencional (experimental) de un grupo poblacional a este insecticida con el único fin de evaluar efectos tóxicos, un gran número de personas se expone frecuentemente a esta sustancia, ya sea por actividades ocupacionales o por vivir en regio-

Cuadro VI

VENTAJAS Y DESVENTAJAS DE LOS ESTUDIOS ECOLÓGICOS

Ventajas

- Se pueden estudiar grandes grupos poblacionales
- Relativamente fáciles de realizar
- Aumenta el poder estadístico
- Aumenta la variabilidad en exposición
- Se puede utilizar información de estadísticas vitales

Desventajas

- No se tiene información del individuo por lo que no se puede ajustar por diferencias a nivel individual (no se puede saber quién sí está expuesto o quién sí desarrolló el evento de interés)
- No se tiene información sobre factores de confusión y no se puede corregir por éstos

nes palúdicas donde el DDT se utiliza como un método de control del paludismo.

A lo largo de los últimos años se ha desarrollado un número importante de estudios epidemiológicos cuyo objetivo principal ha sido el de evaluar si la exposición al DDT en poblaciones humanas tiene efectos sobre la salud, en particular se han evaluado los efectos reproductivos del DDT y su posible asociación con un incremento en el riesgo de cáncer. Mencionaremos algunos de ellos con el fin de ilustrar la utilización de los diferentes diseños epidemiológicos descritos anteriormente en un problema de actualidad en el campo de la salud pública.

Cocco y colaboradores⁶ realizaron un estudio de cohorte retrospectivo con el fin de estudiar la asociación entre exposición laboral al DDT y riesgo de desarrollar cáncer. Para este estudio se definió como población expuesta a los sujetos que participaron en el programa de control del paludismo en Sardinia, Italia. Los investigadores obtuvieron información sobre el estado de salud y la causa de muerte de los trabajadores que participaron en las actividades de control. En este grupo se registraron 1 043 muertes, la distribución y frecuencia de causas de muerte observada se compararon con las que se registraron para la población italiana del mismo sexo y edad.

Hunter y colaboradores⁷ estudiaron la asociación entre los niveles séricos de DDE (indicador de exposición a DDT) y la incidencia de cáncer mamario. Este grupo de investigadores realizó un estudio de casos y controles prospectivo en una cohorte bien definida de enfermeras norteamericanas iniciado en 1986. Para el reporte antes mencionado se incluyeron en la población en estudio 240 casos de cáncer mamario que se registraron en la cohorte de enfermeras entre 1990 y 1992. Los niveles de DDE en suero se compararon con los

observados en un grupo de participantes de la misma cohorte y del mismo grupo de edad, pero que no desarrollaron cáncer mamario. Los niveles de DDE se determinaron en muestras de suero que habían sido recolectadas en años previos al diagnóstico de cáncer mamario.

Romieu y colaboradores⁸ realizaron un estudio de casos y controles para estudiar la asociación entre la exposición ambiental a DDT y el riesgo de desarrollar cáncer mamario en mujeres residentes de la ciudad de México. En este trabajo, los investigadores estudiaron una serie de 120 casos incidentes de cáncer mamario reclutados en unidades médicas del Instituto Mexicano del Seguro Social, el Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado y la Secretaría de Salud, y compararon de manera concurrente los niveles de exposición a DDT, con los observados en una serie de 123 controles que se reclutaron mediante un muestreo aleatorio de mujeres de la ciudad de México.

Wang y colaboradores,⁹ realizaron un estudio multinacional de conglomerados para estudiar la asociación entre exposición al DDT y cáncer. En este estudio los autores recuperaron muestras de cerumen de 3 800 sujetos entre 35 y 54 años de edad de diferentes poblaciones (35 países) y determinaron las concentraciones promedio de exposición a DDT en las muestras de cerumen. La concentración promedio de DDT fue utilizada para caracterizar la exposición promedio de cada país al DDT y posteriormente correlacionaron esta exposición promedio de DDT con la tasa de mortalidad por cáncer de los países incluidos en el estudio.

Finalmente, Ayotte y colaboradores* estudiaron la asociación entre los niveles plasmáticos de DDT y parámetros reproductivos determinados mediante una espermatobioscopia. Para fines de este estudio los investigadores reclutaron voluntarios de una de las zonas palúdicas de México y caracterizaron de manera transversal la asociación entre los niveles séricos de DDT y el número de espermatozoides, así como de otros parámetros funcionales. En este estudio también se evaluó de manera transversal la asociación del DDT en plasma con diferentes proteínas transportadoras de hormonas.

Los diferentes métodos epidemiológicos descritos en este trabajo representan los estudios más utilizados. La aplicación de cada método a diferentes problemas

* Ayotte P, Hernández-Avila M, Farías P, Villanueva CA, Giroux S, Dewailly E. Male reproductive function and endocrine profile in Mexican peasants exposed to *p,p'*-DDEm.

de salud requiere de creatividad y de conocimiento de las limitaciones de cada uno así como de los métodos de análisis que se han desarrollado de manera específica para cada aplicación. En los capítulos subsecuentes se presentará con mayor detalle cada uno de estos diseños.

Referencias

1. Walker AM. Observation and inference. An introduction to the methods of epidemiology. Chestnut Hill, MA: Epidemiology Resources Inc, 1991
2. Kelsey JL, Thompson WD, Evans AS. Methods in observational epidemiology. Nueva York: Oxford University Press, 1986.
3. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research. Principles and quantitative methods. Belmont, CA: Lifetime Learning Publications, 1982.
4. Rothman KJ, Greenland S. Modern epidemiology. 2a. edition. East Washington Square, PA: Lippincott-Raven Publishers, 1998.
5. Miettinen OS. Theoretical epidemiology. Principles of occurrence research medicine. Nueva York: A Wiley Medical Publication, 1985.
6. Cocco P, Blair A, Congia P, Saba G, Ecça AR, Palmas C. Long-term health effects of the occupational exposure to DDT. A preliminary report. Ann NY Acad Sci 1997;837:246-256.
7. Hunter DJ, Hankinson SE, Laden F, Colditz GA, Manson JE, Willett WC *et al*. Plasmaorganochlorine levels and the risk of breast cancer [see comments]. N Engl J Med 1997;337(18):1253-1258.
8. Romieu I, Hernández M, Lazcano E, Weber JP, Dewali E. Breast cancer, lactation history and serum organochlorines. Am J Epidemiol 2000. En prensa.
9. Wang XQ, Gao PY, Lin YZ, Chen CM. Studies on hexachlorocyclohexane and DDT contents in human cerumen and their relationships to cancer mortality. Biomed Environ Sci 1988;1(2):138-151.

Estudios ecológicos

Víctor Hugo Borja-Aburto, M.C., M.S.P., Ph.D.⁽¹⁾

Uno de los diseños de estudio más sencillos y frecuentemente empleados en la descripción de la situación de salud o en la investigación de nuevas exposiciones en poblaciones humanas son los estudios ecológicos. Sin embargo, por lo limitado de sus mediciones, pueden ser más susceptibles de sesgos que los estudios que se basan en observaciones individuales. En este trabajo revisaremos los conceptos generales de este tipo de diseños, las limitantes en la inferencia de sus resultados y presentaremos ejemplos de los mismos.

Los estudios ecológicos en epidemiología se distinguen de otros diseños en su unidad de observación, pues se caracterizan por estudiar grupos, más que individuos por separado. Frecuentemente se les denomina estudios exploratorios o generadores de hipótesis, dejando a los diseños experimentales y algunos diseños observacionales la característica de ser estudios etiológicos o probadores de hipótesis. Se les llama también diseños incompletos debido a que, por emplear promedios grupales, frecuentemente se desconoce la distribución conjunta de las características en estudio a nivel de cada individuo.¹ Comúnmente las unidades de observación son diferentes áreas geográficas o diferentes periodos de tiempo en una misma área, a partir de las cuales se comparan las tasas de enfermedad y algunas otras características del grupo.

Es frecuente que ante las primeras sospechas de efectos negativos a la salud por algún producto o condición ambiental se exploren estas asociaciones en el ámbito grupal. Probablemente la principal motivación para los estudios ecológicos es la fácil disponibilidad de los datos; comúnmente se emplean datos

registrados rutinariamente con propósitos administrativos o legales. Así, las instituciones gubernamentales tienen disponibles estadísticas de mortalidad y morbilidad, al igual que datos de los servicios de salud, mediciones ambientales, venta y consumo de productos de los cuales se sospecha algún efecto. El trabajo realizado por Yang y colaboradores² para estudiar la relación entre cloración del agua y bajo peso al nacer es un ejemplo de lo anterior. Los autores compararon la frecuencia de bajo peso al nacer de hijos nacidos de mujeres primíparas en 14 municipios de Taiwán, donde 90% de la población era abastecida con agua clorada, con otros 14 municipios semejantes en sus características sociodemográficas y de urbanización, donde menos de 5% de la población era abastecida con agua clorada.

Otra motivación para los estudios ecológicos es que la comparación entre diversas áreas permite la evaluación de múltiples niveles de exposición, lo cual puede ser imposible en una sola área geográfica cuando se tienen exposiciones casi homogéneas. Siguiendo el ejemplo anterior, sería difícil hacer una comparación de niveles de exposición relevantes a los productos de la cloración entre las mujeres de un mismo municipio cuando la población comparte la misma fuente de suministro de agua.

Los estudios ecológicos han sido empleados por sociólogos y por epidemiólogos en diversas áreas, que van desde las enfermedades cardiovasculares hasta los efectos de la contaminación ambiental. Una revisión de los estudios ecológicos utilizados para investigar la disminución en el consumo de sal en la dieta y la reducción en la presión arterial, así como una

(1) Centro Nacional de Salud Ambiental, Instituto Nacional de Salud Pública, México.

Solicitud de sobretiros: Dr. Víctor Hugo Borja-Aburto, Centro Nacional de Salud Ambiental, Instituto Nacional de Salud Pública.
 Rancho Guadalupe s/n, 52140 Metepec, Estado de México, México.
 Correo electrónico: vborja@insp.mx

comparación con los estudios a nivel individual y ensayos clínicos se puede consultar el trabajo de Law y colaboradores.³

Tipos de estudios ecológicos

Frecuentemente, se clasifica a los estudios ecológicos como exploratorios, de grupos múltiples, de series de tiempo y mixtos.^{4,5}

Estudios exploratorios. En los estudios exploratorios se comparan las tasas de enfermedad entre muchas regiones continuas durante un mismo periodo, o se compara la frecuencia de la enfermedad a través del tiempo en una misma región. En ninguno de los dos casos se hace una comparación formal con otras variables de los grupos, y el único propósito es buscar patrones espaciales o temporales que podrían sugerir hipótesis sobre las causas. Por ejemplo, Lee y colaboradores⁶ estudiaron las tasas de melanoma maligno de piel y mostraron un incremento cercano a 3% por año en las poblaciones de Inglaterra, Canadá y los Estados Unidos de América (EUA), de 1951 a 1971. El análisis por cohorte de nacimiento mostró que este aumento se debe a un incremento en el riesgo de cáncer de piel en las cohortes de nacimiento más recientes. Este trabajo no propone hipótesis sobre los mecanismos de daño, pero sugiere la necesidad de realizar estudios epidemiológicos más detallados. En otro ejemplo, para mostrar la importancia del estudio de factores genéticos y ambientales, Croen y colaboradores⁷ estudiaron las variaciones en la prevalencia de labio leporino y paladar hendido de acuerdo con la raza de los padres y el lugar de nacimiento de la madre en el registro de nacimientos de California, EUA.

Estudios de grupos múltiples. Este es el tipo de estudio ecológico más común. En un estudio analítico de comparación de grupos múltiples se evalúa la asociación entre los niveles de exposición promedio y la frecuencia de la enfermedad entre varios grupos; comúnmente grupos geopolíticos. La fuente de datos suele ser las estadísticas de morbilidad y mortalidad rutinarias. Hatch y Susser⁸ compararon las tasas de cáncer de niños en relación con la exposición de radiación gamma en 69 pequeñas áreas geográficas cercanas a una planta nuclear. No se contó con información en el ámbito individual sobre la exposición a radiación, sólo se obtuvo información sobre las características sociodemográficas generales de las áreas en estudio.

Estudios de series de tiempo. En un estudio de series de tiempo analítico se comparan las variaciones temporales de los niveles de exposición (uso o consumo

de productos o servicios, conductas, o concentraciones promedio de contaminantes, por ejemplo) con otra serie de tiempo que refleja los cambios en la frecuencia de la enfermedad en la población de un área geográfica. La inferencia causal de este tipo de análisis de series de tiempo puede ser limitada debido a cambios en los criterios diagnósticos de enfermedad y por dificultades provenientes de los periodos de latencia entre la exposición y los efectos o de la medición de la exposición. Loomis y colaboradores⁹ estudiaron la influencia de la contaminación del aire en la ciudad de México con los cambios diarios en un periodo de cuatro años. En este estudio documentaron una asociación entre las concentraciones de contaminantes y el riesgo de muerte prematura por enfermedades respiratorias.

Estudios mixtos. En esta categoría se incluyen los estudios de series de tiempo combinadas con la evaluación de grupos múltiples. Otra variante son los estudios que evalúan la exposición en el ámbito grupal y, en contraste, los datos de los efectos o enfermedad, así como posibles covariables, se encuentran disponibles en el ámbito individual. Hopenhayn-Rich y colaboradores¹⁰ compararon las tendencias de la mortalidad infantil en dos áreas geográficas de Chile: Antofagasta, con historia de contaminación natural por arsénico de las fuentes de abastecimiento de agua, y Valparaíso, una ciudad de exposición baja. La comparación de las variaciones en las fuentes de agua y la tendencia en la mortalidad infantil de las ciudades por medio de técnicas gráficas y modelos de regresión sugieren una influencia de la exposición a arsénico en el incremento del riesgo de muerte fetal y mortalidad infantil.

Inferencia causal en los estudios ecológicos

De manera clásica, en las escuelas de salud pública se enseña que no se debe confiar en los estudios ecológicos puesto que estos diseños son más susceptibles de presentar sesgos que los estudios basados en individuos y, más específicamente, están sujetos a la falacia ecológica. Este problema de hacer inferencias individuales a partir de estudios de grupo, clasificada por Selvin¹¹ como "la falacia ecológica", ha ocasionado un problema de actitud en los epidemiólogos y un descrédito de estos estudios. Sin embargo, algunos practicantes de la salud pública defienden su utilidad real en las decisiones de prevención y control de enfermedades. Ejemplos clásicos, como los estudios de Snow y Farr sobre el cólera,^{12,13} muestran las bondades de estos estudios, a pesar de haber tenido limitantes en el entendimiento biológico detallado de la enfermedad.

Entre otras cosas, este tipo de análisis recientemente ha mostrado, por ejemplo, el inicio de una posible epidemia de cáncer relacionada con la contaminación ambiental.¹⁴

Algunos autores como Susser¹⁵ sugieren que la justificación principal para el enfoque ecológico es estudiar la salud en un contexto ambiental, que la salud de un grupo es más que la suma de la salud de los miembros individuales, y que la perspectiva brindada por los estudios cuya unidad de análisis es el grupo más que los individuos, se debe entender como una manera de abordaje de la epidemiología y la salud pública, cuyo objeto de interés son los grupos. Este autor argumenta que la diferencia entre los diseños y enfoques descansa en los niveles de organización, ya que, en un nivel, las unidades individuales se integran en grupos y estos grupos se convierten en las unidades de análisis en el siguiente nivel. Cada nivel adquiere propiedades colectivas que son más que la suma de las propiedades de sus miembros individuales. El problema surge cuando se pretende extrapolar directamente de un nivel de análisis a otro, ya sea hacia arriba –de un gen a una molécula, a la célula, el tejido, el órgano, la persona, el grupo–, o en sentido inverso.¹⁵ Así, en el análisis de grupos las dimensiones de contexto o integrales son únicas y no pueden ser explicadas en el ámbito individual. Un ejemplo de lo anterior es la inmunidad de grupo en las enfermedades infecciosas, que se refiere a que la susceptibilidad, infección y transmisión de la infección no es independiente entre los individuos de un colectivo. Otro ejemplo es la modificación del efecto de variables individuales en el contexto de grupo, como pueden ser las conductas riesgosas para enfermedades de transmisión sexual en contextos con prevalencias altas o bajas de las enfermedades.

La falacia ecológica se refiere al error que se comete cuando se aceptan asociaciones entre eventos cuando en realidad no existen, siguiendo el supuesto de que los resultados obtenidos a partir de un estudio ecológico serían los mismos que se obtendrían de un estudio basado en observaciones de individuos (un supuesto implícito aquí es que el estándar de oro son los estudios de base individual). Este término en realidad cubre varias fuentes de sesgos¹⁶ que pueden afectar más a los estudios ecológicos que a los otros estudios con base individual. La mayor parte del sesgo ecológico se debe a la dificultad de controlar factores de confusión en este tipo de estudio.¹⁷ La confusión, en este diseño, se refiere a variables que no son de interés intrínseco pero que se encuentran relacionadas con las medidas agregadas de la enfer-

medad o exposición. La confusión en el ámbito individual y ecológico pueden ser independientes. Puede haber confusión en el ámbito individual entre los grupos por variables que no son confusoras a nivel ecológico, o viceversa. Por ejemplo, entre estados o municipios, puede ser que los ancianos tiendan a vivir en casas más contaminadas, lo que hace que se requiera controlar la edad como un confusor. Sin embargo, puede ser que el promedio de edad sea igual entre los distintos municipios, por lo que la edad no es un confusor en el ámbito ecológico.¹⁶

Adicionalmente, muchas veces los estudios ecológicos están sujetos a confusión potencial por la falta de medición de algunas covariables. Estas pueden ser covariables que rutinariamente no son registradas en los reportes oficiales. Por ejemplo, rutinariamente no se registra el número de fumadores en una comunidad, factor que podría ser relevante en el estudio de casi todas las enfermedades respiratorias.

Otro problema importante que se presenta en los estudios ecológicos es la determinación de la secuencia temporal entre la exposición y la enfermedad. La mayor parte de los estudios ecológicos son transversales, es decir, tanto la enfermedad como la exposición son medidas al mismo tiempo. Por ejemplo, si se desea evaluar la influencia del tabaco en una comunidad o de las medidas de prevención, se debe esperar un tiempo adecuado para observar sus efectos; es decir, tomar en cuenta el periodo de latencia entre la exposición y el efecto. Frecuentemente se asume que la exposición actual refleja la exposición en el pasado. El mismo problema ocurre en los estudios con base en individuos; sin embargo, en los estudios ecológicos se presenta la migración como problema adicional, es decir, la población afectada podría haber emigrado antes de la medición de la enfermedad, o algunos otros podrían haber inmigrado a la población estudiada, ocasionando un sesgo de selección.¹⁸

El problema de inadecuada especificación del modelo de análisis es otro factor de discrepancias cuando se espera que los resultados de los estudios ecológicos sean los mismos que de estudios individuales. Teóricamente, los resultados podrían ser los mismos cuando se emplean modelos lineales aditivos, sin términos multiplicativos.¹⁹ Sin embargo, en los estudios individuales la medida de asociación más frecuentemente empleada es una estimación de la razón de tasas o un equivalente como el riesgo relativo, mismo que no se puede estimar directamente en modelos de regresión con medidas de exposición continuas y ser interpretado como una medida de razón entre dos grupos de exposición.

Análisis estadístico de los estudios ecológicos

La manera usual de evaluación de la asociación en estudios de grupos múltiples es mediante modelos lineales de regresión. Comúnmente se emplean los métodos de mínimos cuadrados para producir la ecuación de predicción: $Y = \beta_0 + \beta_1 X_1$, donde β_0 es el intercepto y β_1 la pendiente estimada o razón de cambio en la frecuencia de la enfermedad por unidad de cambio en la exposición. Las tasas o frecuencia de la enfermedad de cada región se convierten en la variable independiente, mientras que los promedios de las variables de exposición son las variables independientes. Cuando se cuenta con la medición de otras covariables (Z) éstas son incluidas en el modelo de regresión como una variable adicional para controlar por el posible efecto confusor de la misma, de tal manera que el modelo es $Y = \beta_0 + \beta_1 X_1 + \beta_2 Z_2$.

Dependiendo del diseño y la distribución de los datos se pueden emplear otros modelos no lineales o no aditivos. Como las tasas de morbilidad y mortalidad en las regiones geográficas que se comparan comúnmente son eventos raros o que ocurren a bajas frecuencias, éstos semejan una distribución Poisson; así que la regresión Poisson puede ser usada para hacer las comparaciones entre los grupos y poder incluir variables potencialmente confusoras en modelos de regresión múltiple. El modelo se puede expresar de la siguiente manera $\text{Log}(E[Y]) = \beta_0 + \beta_1 X_1 + \beta_2 Z_2$.

Un problema adicional es el poco desarrollo de modelos matemáticos que integren distintos niveles de análisis en estudios ecológicos donde se cuenta con variables integrales o contextuales medidas a nivel de grupo y con características en el ámbito individual. El análisis contextual, frecuentemente usado en ciencias sociales, es un ejemplo de estos modelos, y más recientemente los modelos de efectos mixtos permiten la evaluación de características combinadas en varios niveles, ya sea regiones, conglomerados e individuos.²⁰

El caso de la contaminación del aire y sus efectos en la mortalidad

El caso de los estudios de los efectos de la contaminación del aire sobre la mortalidad puede ilustrar la aplicación de los diseños ecológicos. Uno de los primeros reportes que llamaron la atención sobre estos posibles efectos se produjo en Londres, en 1952. Las condiciones meteorológicas provocaron una inversión térmica y una acumulación de contaminantes, principalmente de material particulado, que se asoció con un rápido aumento en el número de muertes en la ciudad, como se

puede observar en la figura 1. Este patrón no dejó duda de que el incremento en la mortalidad se relacionó con la contaminación del aire en ausencia de epidemias por causas infecciosas.²¹ Por fortuna, las acciones de control de la calidad del aire han evitado que estos eventos de contaminación severa se presenten en épocas recientes. Sin embargo, durante la última década se han publicado un gran número de trabajos de estudios de series de tiempo en diversas ciudades que evalúan el efecto de la contaminación del aire a niveles mucho más bajos presentes en áreas urbanas de casi todo el mundo.²²

En la ciudad de México se han realizado varios análisis de series de tiempo para evaluar este efecto.²³⁻²⁶ La figura 2 muestra el comportamiento de la mortalidad en la zona suroeste de la ciudad. Como se puede observar se presenta un componente estacional en el número de muertes, con picos en los meses fríos, de la misma manera que se comporta temporalmente la concentración de material particulado respirable. Sin embargo, la temperatura también tiene una variación similar, lo que podría indicar un efecto confusor importante, que ha llevado al desarrollo de modelos estadísticos que permiten evaluar el efecto de la calidad del aire, independientemente del efecto de la temperatura y otras condiciones ambientales que presentan variación estacional semejante.²⁷

Este efecto de la contaminación del aire también ha sido evaluado con estudios ecológicos de comparación de grupos múltiples. Bobak y Leon²⁸ condujeron un estudio ecológico sobre mortalidad infantil y contaminación del aire en la República Checa. Analizaron datos rutinarios de mortalidad infantil y contaminación del aire de 1986 a 1988 para 46 distritos de este país. Emplearon regresión logística para estimar los efectos de partículas suspendidas totales, dióxido

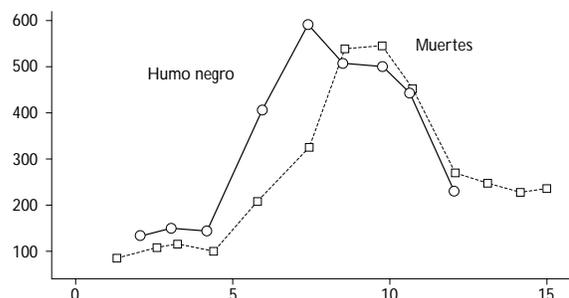


FIGURA 1. ASOCIACIÓN ENTRE EL AUMENTO DE CONTAMINACIÓN (MATERIAL PARTICULADO) E INCREMENTO DE NÚMERO DE MUERTES. LONDRES, 1952²¹

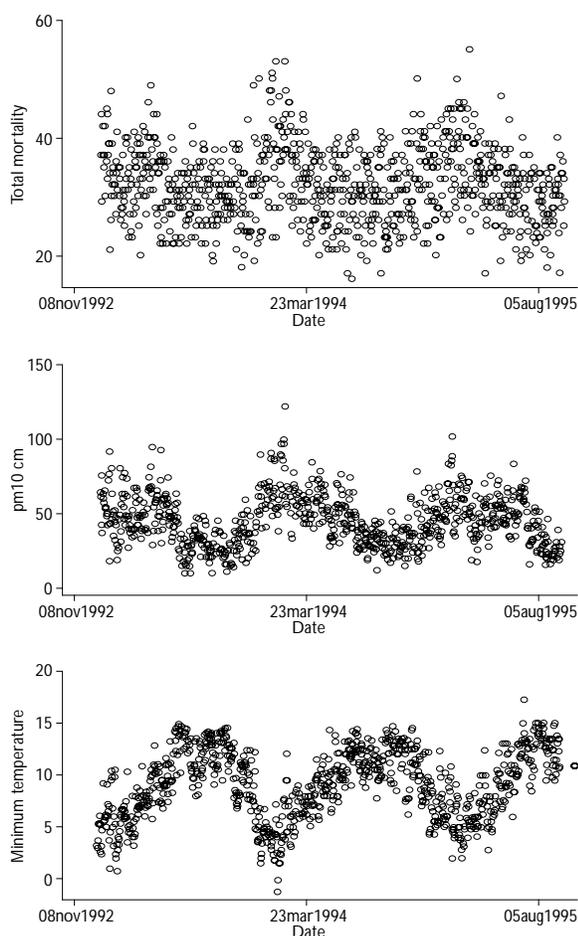


FIGURA 2. COMPORTAMIENTO DE LA MORTALIDAD EN LA ZONA SUROESTE DE LA CIUDAD DE MÉXICO²⁴

de azufre y óxidos de nitrógeno ajustando por características socioeconómicas de los distritos, tales como ingreso, posesión de automóviles y tasa de abortos. Los autores encontraron débiles asociaciones de partículas suspendidas menores a 10 micras con la mortalidad neonatal, pero una asociación más fuerte con la mortalidad posneonatal por causas respiratorias. Estos hallazgos fueron confirmados en un estudio de casos y controles más reciente.²⁹ Loomis y colaboradores evaluaron también esta asociación en un estudio de series de tiempo en la ciudad de México, con resultados que apoyan las asociaciones iniciales del estudio de comparación de grupos múltiples.⁹

Uno de los problemas de estos estudios ecológicos es la medición de la exposición a los contaminantes. Por un lado, estas series de tiempo se basan en mediciones de la calidad del aire hechas con monito-

res fijos, que no necesariamente reflejan las variaciones individuales en la exposición, derivadas del tipo de vivienda, tiempo dedicado a actividades extramuros o intensidad de las actividades y, por otro lado, la medición de las partículas suspendidas se basa en su peso más que en los componentes químicos de las mismas. Sin embargo, aunque estos estudios ecológicos no pueden por sí mismos confirmar la causalidad de los efectos observados, han impulsado una gran cantidad de otros estudios epidemiológicos y de experimentos en condiciones controladas con animales que cada día dan más pistas sobre los mecanismos que conducen a un aumento en el riesgo de morir en ambientes contaminados. A la fecha, tres estudios de cohorte han encontrado disminución en la esperanza de vida en ciudades con mayor concentración de partículas respirables.³⁰⁻³²

Referencias

1. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research. Nueva York: Van Nostrand Reinhold, 1982.
2. Yang CY, Cheng BH, Tsai SS, Wu TN, Lin MC, Lin KC. Association between chlorination of drinking water and adverse pregnancy outcome in Taiwan. *Environ Health Perspect* 2000;108:765-768.
3. Law MR, Frost CD, Wald NJ. By how much does dietary salt reduction lower blood pressure? I. Analysis of observational data among populations. *BMJ* 1991;302:811-824.
4. Morgenstern H. Ecologic studies. En: *Modern epidemiology*. Rothman K, Greenland S, ed. Lippincott: Raven Publishers, 1998;459-480.
5. Morgenstern H. Uses of ecologic analysis in epidemiologic research. *Am J Public Health* 1982;72(12):1336-1344.
6. Lee JAH. Melanoma and exposure to sunlight. *Epidemiol Rev* 1982;4:110-136.
7. Croen LA, Shaw GM, Wasserman CR, Tolarova MM. Racial and ethnic variations in the prevalence of orofacial clefts in California, 1983-1992. *Am J Med Genet* 1998;79(1):42-47.
8. Hatch M, Susser M. Background gamma radiation and childhood cancers within 10 miles of a U.S. nuclear plant. *Int J Epidemiol* 1990;19:546-552.
9. Loomis DP, Castillejos M, Gold DR, McDonnell W, Borja-Aburto VH. Air pollution and infant mortality in Mexico City. *Epidemiology* 1999;10:118-123.
10. Hopenhayn-Rich C, Browning SR, Hertz-Picciotto I, Ferreccio C, Peralta C, Gibb H. Chronic arsenic exposure and risk of infant mortality in two areas of Chile. *Environ Health Perspect* 2000;108:667-673.
11. Selvin H. Durkheim's suicide and problems of empirical research. *Am J Sociol* 1958;63:607-619.
12. Snow J. On the mode of communication of cholera. 2a. edición. Londres: J. Churchill, 1855. Reprinted as Snow on cholera. Nueva York (NY): The common-wealth Fund, 1956.
13. Farr W. Report on the mortality of cholera in England in 1848-1849. Londres: His Majesty's Stationery Office, 1852.
14. Poole C. Ecologic analysis as outlook and method. *Am J Public Health* 1994;84(5):715-716.

15. Susser M. The logic in ecological: I. The logic of analysis. *Am J Public Health* 1994;84(5):825-829.
16. Steenland K, Deddens JA. Design and analysis of studies in environmental epidemiology. En: *Topics in environmental epidemiology*. Nueva York: Oxford University Press, 1997.
17. Greenland S, Robins J. Ecologic studies-biases, misconceptions, and conterexamples. *Am J Epidemiol* 1994;139(8):747-760.
18. Polissar L. The effect of migration on comparison of disease rates in geographic studies in the United States. *Am J Epidemiol* 1980;111(2):175-182.
19. Piantadosi S. Ecologic biases. *Am J Epidemiol* 1994;139(8):761-764.
20. Goldstein H. *Multilevel statistical models*. 2a. edición. Londres: Edward Arnold, 1995.
21. Schwartz J. Air pollution and daily mortality: A review and meta-analysis. *Environ Res* 1994;64:36-52.
22. Levy JI, Hammitt JK, Spengler JD. Estimating the mortality impacts of particulate matter: What can be learned from between-study variability? *Environ Health Perspect* 108:109-117(200a).
23. Borja-Aburto VH, Loomis D, Shy C, Bangdiwala S, Rascón-Pacheco RA. Ozone, suspended particulates and daily mortality in Mexico City. *Am J Epidemiol* 1997;145:258-268.
24. Borja-Aburto VH, Castillejos M, Gold DR, Bierzwinski S, Loomis D. Mortality and ambient fine particles in Southwest Mexico City, 1993-1995. *Environ Health Perspect* 1998; 106:849-855.
25. Castillejos M, Borja-Aburto VH, Dockery DW, Gold DR, Loomis D. Airborne coarse particles and mortality. *Inhal Toxicol* 2000;12 (Suppl 1): 61-72.
26. Téllez-Rojo MM, Romieu I, Velasco S. Daily respiratory mortality and pollution in Mexico City. The importance of considering place of death. *Eur Respir J* 2000;16(3):391-396.
27. Samet JM, Zeger SL, Berhane K. Particulate air pollution and daily mortality. Health Effects Institute, August 1995.
28. Bobak M, Leon DA. Air pollution and infant mortality in the Czech Republic, 1986-88. *Lancet* 1992;340:1010-1014.
29. Bobak M, Leon DA. The effect of air pollution on infant mortality appears specific for respiratory causes in the postneonatal period. *Epidemiology* 1999;10(6):666-670.
30. Abbey DE, Nishino N, McDonnell WF, Burchette RJ, Knutsen SF, Beeson WL *et al*. Long-term inhalable particles and other air pollutants related to mortality in nonsmokers. *Am J Respir Crit Care Med* 1999; 159: 373-382.
31. Dockery DW, Pope CA III, Xu X, Spengler JD, Ware JH, Fray ME *et al*. An association between air pollution and mortality in six US cities. *N Engl J Med* 1993;329:1753-1759.
32. Pope CA III, Thun MJ, Namboodiri MM, Dockery DW, Evans JS, Speizer FE *et al*. Particulate air pollution as a predictor of mortality in a prospective study of US adults. *Am J Respir Crit Care Med* 1995;141:669-674.

Encuestas transversales

Bernardo Hernández, D.Sc.,⁽¹⁾
Héctor Eduardo Velasco-Mondragón, M. en C.⁽¹⁾

La encuesta transversal es un diseño de investigación epidemiológica de uso frecuente. Se trata de estudios observacionales, también llamados encuestas de prevalencia.¹⁻⁴ El diseño de una encuesta transversal debe considerar aspectos relacionados con la población que se estudiará, los sujetos de quienes se obtendrá información y la información que se busca captar.

En epidemiología las encuestas transversales se dirigen primordialmente al estudio de la frecuencia y distribución de eventos de salud y enfermedad (estudios descriptivos), aunque también se utilizan para explorar y generar hipótesis de investigación (estudios analíticos). En el primer caso, las encuestas tienen como fin medir una o más características o enfermedades (variables) en un momento dado de tiempo; por ejemplo: el número de enfermos con diabetes en la población en un momento dado; el número de integrantes de las familias en un periodo de tiempo determinado; el promedio de edad de hombres y mujeres que utilizaron o no utilizaron servicios de salud por trimestres del año; el nivel de satisfacción de pacientes atendidos por médicos familiares el mes previo, o la intención en hombres y mujeres de cesar de fumar en los meses siguientes.

Las encuestas transversales son de gran utilidad por su capacidad para generar hipótesis de investigación, estimar la prevalencia de algunos padecimientos (esto es, la proporción de individuos que padece alguna enfermedad en una población en un momento

determinado), así como identificar posibles factores de riesgo para algunas enfermedades.

Cuando el fin es explorar hipótesis de investigación, la característica distintiva de este tipo de estudios es que la variable de resultado (enfermedad o condición de salud) y las variables de exposición (características de los sujetos) se miden en un mismo momento o periodo definido. Como ejemplos se tiene el análisis de: la relación entre alcoholismo (exposición) y violencia intrafamiliar (evento); la relación entre actividad física (exposición) y obesidad (evento), y la relación entre estado marital (exposición) y mortalidad (evento).

A diferencia de otros diseños epidemiológicos, como los estudios de cohorte, en los cuales se realiza un seguimiento de sujetos expuestos y la ocurrencia de eventos nuevos por un periodo determinado de tiempo, en las encuestas transversales se obtiene únicamente una medición de las exposiciones y eventos en los sujetos de estudio en un momento dado. Debido a esto, no es posible determinar si el supuesto factor de exposición precedió al aparente efecto y establecer causalidad entre exposición y efecto, salvo en el caso de exposiciones que no cambian con el tiempo. Su limitación para establecer causalidad se compensa por su flexibilidad para explorar asociaciones entre múltiples exposiciones y múltiples efectos.

Las encuestas transversales son utilizadas para estudiar enfermedades de larga duración o cuyas ma-

(1) Dirección de Epidemiología, Centro de Investigación en Salud Poblacional, Instituto Nacional de Salud Pública. Cuernavaca, Morelos, México.

nifestaciones se desarrollan lentamente, como es el caso de enfermedades crónicas, problemas de desnutrición o mala nutrición por exceso, etcétera. Estas encuestas no son adecuadas para el estudio de enfermedades (o exposiciones) que se presentan con poca frecuencia en una población (enfermedades raras o con baja prevalencia) o que son de corta duración, debido a que sólo captarían información sobre un número reducido de individuos que las padezcan. A pesar de sus limitaciones, los estudios transversales son comunes y útiles, ya que su costo es relativamente inferior al de otros diseños epidemiológicos, como los estudios de cohorte, y proporcionan información importante para la planificación y administración de los servicios de salud.

En el presente artículo se discuten algunos aspectos relevantes para el diseño, conducción, análisis, interpretación y aplicaciones de las encuestas transversales, así como sus ventajas y limitaciones.

Población y muestra

En las encuestas transversales se obtiene información sobre una población definida para fines del estudio. A diferencia de otros diseños, como los estudios de casos y controles, en las encuestas transversales se recolectan datos de los sujetos sin estratificarlos *a priori* de acuerdo con la presencia o ausencia de la variable de resultado que se desee estudiar.

Se define como población base del estudio aquella a la que el estudio hace referencia.⁵ En muchas ocasiones una encuesta transversal no obtiene información de todos los sujetos que integran la población bajo estudio, sino sobre un grupo de ellos llamado muestra.

Al realizar una encuesta transversal es necesario definir la unidad de observación del estudio, esto es, la unidad básica sobre la cual se captará información, como, por ejemplo, los individuos, familias, hogares o escuelas.

El proceso de selección de informantes es muy importante en estos estudios. La muestra seleccionada debe reflejar las características de la población base que se busca estudiar; por ejemplo, para obtener la media de edad o la distribución de edades de la población general se deberá incluir sujetos de todas las edades. En ocasiones el investigador puede estar interesado en estudiar características de algún subgrupo específico de su población y, por lo mismo, puede aumentar la proporción de sujetos en la muestra que pertenecen a ese subgrupo.

Existen diversos métodos de selección de sujetos para participar en el estudio llamados métodos de muestreo. Algunos métodos de uso frecuente son el muestreo por conveniencia y formas de muestreo

probabilístico (aleatorio simple, estratificado, sistemático y por conglomerados). Los métodos de muestreo se presentan en otro artículo de esta serie, por lo que solamente se destacan los sesgos potenciales de las encuestas transversales.

Las encuestas transversales suelen sobrerrepresentar a los casos con larga duración de la enfermedad y a subrepresentar aquéllos de corta duración (muestreo con sesgo de duración). Por ejemplo, en el caso de una enfermedad de duración muy variable, una persona que contrae una enfermedad que dura desde los 20 a los 70 años tiene una gran posibilidad de ser incluida durante los 50 años de duración de la enfermedad; en cambio, una persona que contrae la enfermedad a los 40 años y que muere al día siguiente, difícilmente será incluida en el grupo prevalente.

En cuanto a la exposición, se debe considerar si ésta cambia con el tiempo (por ejemplo, en un periodo de tiempo una persona puede dejar de fumar o fumar más) o no cambia (por ejemplo, el grupo sanguíneo, el sexo).³ Además, si la exposición produce enfermedad leve y de larga duración –aun cuando no produzca riesgo de enfermar– la frecuencia de exposición será elevada en los casos y, por lo tanto, de aparente mayor riesgo. En cambio, si la exposición produce una alta letalidad de la enfermedad, entonces la frecuencia de exposición será muy baja entre los casos y la asociación exposición-enfermedad puede resultar negativa, aun cuando en realidad la exposición no resulte en menor riesgo de enfermar.

A los sesgos inherentes de los estudios transversales se agregan otros: si el modo de selección de los sujetos de estudio (participación voluntaria) está relacionado con menor o mayor exposición, o con menor o mayor enfermedad en comparación con la población base, entonces los resultados que obtengamos no serán válidos. A esto se le conoce como sesgo de selección.⁶ Una de las estrategias para tratar de evitar este sesgo es realizar un muestreo probabilístico o aleatorio en el que todos los individuos que conforman la población bajo estudio tengan la misma probabilidad de ser incluidos en el estudio, o que se conozca la probabilidad que tiene cada sujeto de la población de ser incluido en la muestra.

Otros sesgos son el sesgo de cortesía (la persona trata de complacer al entrevistador dándole la respuesta que cree será aprobada); el sesgo de vigilancia (la enfermedad o evento se confirma mejor en la población de estudio que en la población general), y el sesgo de información (debido a datos poco verídicos o incompletos o a la no participación o no respuesta de los individuos seleccionados como población de estudio, lo cual se puede relacionar con características de interés

que hagan que la población participante sea diferente a la no participante.) Esto afectará las estimaciones de prevalencias o de asociaciones entre exposiciones y efectos y afectará la validez del estudio. Con el fin de analizar y corregir el efecto de dicha falta de participación es necesario conocer las razones de no participación o no respuesta y las características de los sujetos no participantes, para saber si se trata de valores perdidos al azar o de manera sistemática y cómo esto afecta las mediciones.⁷ La comparación entre los estimadores obtenidos con y sin datos de estos sujetos nos permitirá estimar la magnitud del sesgo. La inclusión de los sujetos no participantes permite obtener el estimador más conservador, o sea, el peor escenario, sesgado contra la prevalencia o hipótesis de interés.

Cuando se hacen preguntas sobre exposiciones o eventos pasados, las personas que han sufrido una experiencia traumática (enfermedad, aborto, accidente) pueden recordar las exposiciones más que los que no tuvieron dicha experiencia, produciendo el llamado sesgo de memoria.

La ausencia de sesgos en la selección de los sujetos de estudio y en la medición de variables en la población de estudio constituye su validez interna; esto es, los resultados obtenidos son ciertos para la población o muestra estudiada. Si la muestra es representativa de la población base, esto aumentará la validez externa del estudio; esto es, la posibilidad de inferir dichos resultados a la población base de la cual se obtuvo la muestra, así como a poblaciones similares.

Generalmente, en las encuestas transversales el tamaño de muestra se calcula de tal forma que permita estimar, con un determinado poder y nivel de confianza, la prevalencia de alguna enfermedad, alguna característica de la población, o bien, la diferencia en nuestra variable de resultado de acuerdo con la variable de exposición. Un tamaño de muestra pequeño no permitirá que el estudio tenga el poder suficiente para encontrar asociaciones significativas entre las variables de exposición y resultado, y un tamaño excesivo ocasionará dispendio de recursos y tiempo. Las características (exposiciones y efectos), el nivel de confianza estadística y el poder del estudio se deberán establecer al inicio del mismo. El cálculo de nivel de confianza y poder de un estudio obtenidos *a posteriori* en una muestra no son válidos.^{8,9}

Definición de variables en estudios transversales

Es importante definir, antes de iniciar el estudio, las variables de estudio, de resultado, de exposición y

potenciales variables confusoras o modificadores de efecto que se desea estudiar de manera teórica y operacional. La definición operacional consiste en determinar la forma en que se medirá una variable. Esta, junto con los indicadores e instrumentos que se utilizarán, definirá el tipo de análisis de dichas variables.

Conducción de encuestas transversales

Las encuestas transversales, si bien son logísticamente más sencillas que los estudios que implican un seguimiento de sujetos, tienen dificultades importantes por los tamaños de muestra que pueden alcanzar. Una vez diseñado el estudio, definidas la población y muestra, así como las variables que se investigarán es necesario definir los instrumentos que se emplearán para recolectar la información. Las encuestas transversales con frecuencia utilizan cuestionarios que pueden ser aplicados a los informantes por un(a) entrevistador(a), o bien empleando cuestionarios autoadministrados; en otras ocasiones se tomarán muestras biológicas (sangre, orina, saliva) o mediciones antropométricas (peso, talla, pliegues cutáneos).

En el caso de la utilización de cuestionarios es necesario definir al informante ideal para proporcionar la información necesaria para el estudio. En general, la persona que responde debe ser capaz de entender el vocabulario utilizado para hacer la pregunta. Los datos de ingreso económico familiar los deberá proporcionar el jefe o jefa de familia. Asimismo, es necesario definir si los cuestionarios serán administrados por una entrevistadora o entrevistador (ya sea en una entrevista cara a cara o por vía telefónica), o si los cuestionarios serán contestados por escrito por los informantes, ya sea en presencia de personal del proyecto o bien por medios como el envío de cuestionarios por correo. Ante el avance de los recursos computacionales, también es posible que la información de la entrevista se capte en medios electrónicos al momento de la entrevista. Esta decisión dependerá de las características de los informantes y de la naturaleza de la información a recolectar. Las preguntas sobre conducta sexual, toxicomanías y otras sobre la vida privada de las personas suelen producir respuestas incompletas o evasivas por parte de los sujetos de estudio, por lo que se deberán probar de antemano y hacerse con particular cuidado.

Es necesario cuidar la integración de los instrumentos de recolección de información. Los cuestionarios deben estar adaptados a la forma en que serán administrados y a la población bajo estudio. Los cuestionarios deben tener un formato que permita su aplicación y, posteriormente, su fácil codificación

(transcripción de respuestas a códigos numéricos) y captura de información en medios electrónicos. Se debe evaluar la confiabilidad y validez de un cuestionario antes de su utilización en un estudio transversal. La confiabilidad se refiere a la capacidad de un instrumento para dar resultados similares en distintos momentos en el tiempo. La validez es la capacidad de un instrumento para medir la variable que realmente desea medir.¹⁰ La confiabilidad y validez de las distintas secciones de un cuestionario pueden ser evaluadas en un estudio específico (piloto) con una muestra menor de sujetos.

Al integrar un cuestionario no es necesario que todas las preguntas que contiene sean creadas específicamente para él. Es posible integrar en un cuestionario preguntas o escalas que ya han sido utilizadas y validadas con esa población o con poblaciones similares. El uso de instrumentos de medición similares y estandarizados permitirá comparar los resultados del estudio con los de otros estudios. En todos los casos es necesario llevar a cabo una prueba piloto de los instrumentos de recolección de información. Esta prueba, llevada a cabo con una submuestra de la población bajo estudio, permitirá corregir errores y problemas en el cuestionario y su procedimiento de aplicación.

Asimismo, si en dicha submuestra es posible obtener y verificar las respuestas correctas al cuestionario o instrumento de medición, se podrá estimar el sesgo entre las respuestas del cuestionario inicial (cuestionario a evaluar) y las del cuestionario validado (estándar de oro).¹¹

El trabajo de campo es una etapa crucial del estudio. Es en ese momento cuando la información que se empleará en los análisis será recolectada, y los errores u omisiones en esta etapa serán difíciles de corregir. Por esta razón, es importante llevar a cabo un riguroso entrenamiento y supervisión del personal de campo. Habitualmente, el equipo de trabajo en campo se integra por supervisores(as) que revisan y coordinan la recolección de información, y encuestadores(as) que propiamente recogen la información (sea administrando cuestionarios, tomando muestras biológicas o antropométricas, etc.). El personal de campo debe tener un conocimiento general del proyecto que le permita recolectar la información necesaria, pero cuidando que no les lleve a sesgar la información que se recolectará. El entrevistador o recolector de datos también puede ser fuente de sesgo, cuando un mismo entrevistador obtiene mediciones diferentes de la característica o atributo de interés (variabilidad intraobservador), y cuando una misma medición se obtiene de manera diferente entre un observador y otro (variabilidad entre observadores).¹²

El personal de campo debe conocer a fondo los instrumentos de recolección de información, el equipo a utilizar para la obtención de muestras, y debe estar estandarizado en la toma de las muestras a recolectar en el estudio. Del mismo modo, debe conocer el esquema de muestreo a utilizar y los procedimientos a seguir cuando un informante no se encuentre, cuando una entrevista no se pueda realizar, etcétera.

Una vez recolectada la información ésta debe ser cuidadosamente revisada por los supervisores de campo y por personal de verificación. En caso de detectar omisiones o anomalías se aconseja regresar con la persona entrevistada para completar o corregir la información, la cual, una vez verificada, debe ser codificada para permitir su captura en medios electrónicos y, posteriormente, su análisis. El sesgo del observador o el llenado fraudulento de datos se pueden evaluar mediante la identificación de repetición de ciertos dígitos en una variable registrada por un observador en comparación con los demás observadores. La distribución de respuestas a dicha variable deberá ser similar a la obtenida por los otros observadores.

Para minimizar errores en el proceso de captura de información es necesario emplear programas de captura validada, en los cuales la información es capturada dos veces para identificar discrepancias. Dos tipos de verificación son comunes: la verificación por rangos (no debe haber valores no plausibles o fuera de límite de los valores posibles de esa variable) y la contingencial o lógica (no debe haber valores incoherentes, por ejemplo, sexo masculino y número de abortos, o profesionistas analfabetas). Actualmente es posible llevarlas a cabo en los programas computacionales de bases de datos relacionales (DBase, Paradox, Fox-Pro) y en algunos paquetes estadísticos que contienen rutinas para la elaboración de programas de captura validada de fácil utilización (EpiInfo, CDC, Atlanta). El uso de tecnologías como lectores ópticos puede agilizar el proceso de captura, aunque requiere de equipo y formatos de captura de información específicos.

El diseño, métodos y procedimientos del estudio deberán estar debidamente documentados, de tal manera que exista información disponible en caso de necesidad de replicar y comparar el estudio.

Análisis de encuestas transversales

El análisis de datos de una encuesta transversal depende de los objetivos de la misma y de la escala de medición de las variables de estudio. Si el fin es caracterizar o describir a la población, se miden las variables una vez y se presentan los valores de cada una de ellas o por grupos.

El análisis de información de encuestas transversales suele iniciar con la obtención de estadísticas descriptivas de variables de interés (figura 1). Este análisis permitirá conocer las características generales de la población bajo estudio y estimar prevalencias de las exposiciones y variables de resultado. Por ejemplo, una encuesta permitirá conocer la frecuencia y distribución de edades, escolaridad, ingreso económico, género, uso de servicios de salud, motivos de consulta médica, tabaquismo, cefalea, opinión sobre el estado de salud, etcétera. Para datos categóricos (presencia o ausencia de enfermedad, número de hijos, nivel socioeconómico bajo, medio, alto, etc.) la descripción se hace por medio de distribución de frecuencias (número de sujetos u observaciones dentro de cada categoría de la variable), frecuencias relativas (distribución porcentual de las observaciones dentro de las categorías de la variable) y proporciones. La prevalencia de una enfermedad se obtiene dividiendo el número de casos encontrados durante el periodo de estudio entre el total de la población en riesgo de presentar el evento de estudio.

Si se quiere estimar la prevalencia de una enfermedad, por ejemplo, de diabetes, los datos se presentan como proporción (número de diabéticos sobre el total de la población) por 100, 1 000, 10 000, etcétera. Si la medición se realiza en un periodo corto de tiempo se le llama prevalencia puntual; si se mide en un periodo mayor de tiempo se le llama prevalencia lápsica.

En el caso de variables continuas, como el peso y la talla de niños menores de cinco años, se presentan medidas de tendencia central (media, mediana, moda) y de dispersión (rangos, desviación estándar, varian-

za, percentiles). Este tipo de variables se puede categorizar en escalas de acuerdo con el interés de la investigación, esto es, convertirse en variables ordinales (por ejemplo, bajo peso, normal y sobrepeso, o talla baja, normal y alta).

En los estudios transversales analíticos se explora la relación o asociación entre variables de exposición y de resultado. En el caso de variables dicotómicas, la clasificación se hace en cuatro grupos: enfermos expuestos, enfermos no expuestos, sanos expuestos y sanos no expuestos (figura 2).

La asociación entre la ocurrencia de una enfermedad (variable de resultado) y algunos factores de riesgo (variables independientes o de exposición) se hace a través del cálculo de medidas de asociación. La mejor medida de riesgo de pasar del estado sano al de enfermo en una población es la tasa de incidencia (cuyo numerador es el número de casos nuevos de enfermedad y el denominador es el tiempo persona en riesgo); la razón de tasas (tasa de incidencia en expuestos entre la tasa de incidencia en no expuestos) indica el exceso de riesgo de enfermedad entre expuestos y no expuestos. A falta de este estimador ideal, se puede obtener la incidencia acumulada (número de casos nuevos de enfermedad, entre la población de riesgo al inicio del estudio) y la razón de riesgos o riesgo relativo (incidencia acumulada en expuestos entre incidencia acumulada en no expuestos). En los estudios transversales, como sólo se tiene la prevalencia (casos ya existentes y casos nuevos, con distintos periodos de duración de enfermedad, entre la población de estudio), las medidas de asociación que se pueden obtener son la razón de prevalencias (RP) y la razón de momios de prevalencia (RM).¹³

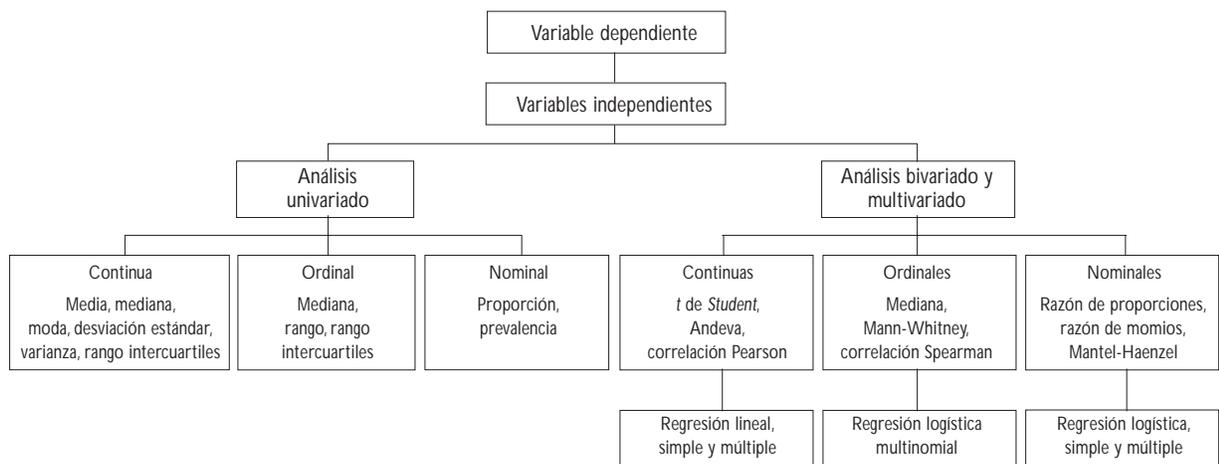


FIGURA 1. ANÁLISIS ESTADÍSTICO POR TIPO DE VARIABLE

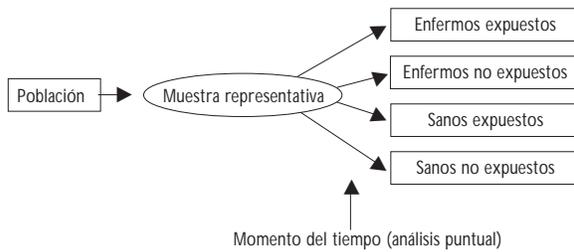


FIGURA 2. DISEÑO DE ENCUESTA TRANSVERSAL

Para la interpretación de estas medidas de asociación es necesario considerar la relación entre prevalencia, incidencia y duración de la enfermedad, la cual se expresa como:

$$P = I \cdot D \times (1 - P)$$

Donde:

- P= prevalencia
- I= incidencia
- D= duración promedio de la enfermedad

Para calcular estas medidas de asociación se construye una tabla de cuatro celdas (cuando se trata de variables dicotómicas) donde las columnas registran el número de enfermos y no enfermos y los renglones el número de expuestos y no expuestos:

	Enfermos	No enfermos	
Expuestos	a	b	a+b
No expuestos	c	d	c+d
	a+c	b+d	a+b+c+d

donde:

- a+c= número de enfermos en la población
- a+c/a+b+c+d= prevalencia de enfermedad en la población
- a/a+b= prevalencia de enfermedad en los expuestos
- c/c+d= prevalencia de enfermedad en los no expuestos
- (a/a+b) / (c/c+d)= razón de prevalencias de enfermedad.

Un valor de uno se interpreta como igual prevalencia de enfermedad entre expuestos y no expuestos. Un valor mayor de uno significa que la prevalencia es mayor en los expuestos que en los no expuestos. Un valor menor a uno significa que la prevalencia es menor en los no expuestos que en los expuestos. La depen-

dencia de la prevalencia respecto a la duración de la enfermedad antes mencionada, hace que estas medidas se aproximen a la razón de riesgos sólo bajo ciertas condiciones. La razón de prevalencias se aproxima o es buen estimador de la razón de riesgos o riesgo relativo (RR) cuando la duración de la enfermedad es igual entre expuestos y no expuestos y cuando no hay migración hacia dentro o fuera de estos grupos.

$$\text{Razón de prevalencias} = \text{RR} \times \frac{D \text{ en expuestos}}{D \text{ no expuestos}} \times \frac{[1.0 - \text{prevalencia en expuestos}]}{[1.0 - \text{prevalencia no expuestos}]}$$

Así, la razón de prevalencias puede ser un buen estimador del riesgo relativo en función de la razón de duración de la enfermedad en expuestos y no expuestos y de la razón de los complementos de la prevalencia en expuestos y no expuestos.

Alternativamente, se puede calcular la razón de momios de prevalencia de enfermedad, con la fórmula:

$$(a/b) / (c/d), \text{ o con la razón de productos cruzados } (a \cdot d) / (b \cdot c)$$

La razón de momios de prevalencia también se relaciona con la incidencia y duración de la enfermedad, ya que:

$$\text{Momios de prevalencia: } \frac{\text{prevalencia}}{1 - \text{prevalencia}} = \text{incidencia} \times \text{duración}$$

Así, la razón de momios de prevalencia es igual a:

$$\frac{\text{Momios de prevalencia en expuestos}}{\text{Momios de prevalencia en no expuestos}}$$

Su valor es aproximado al de la razón de prevalencias (y razón de riesgos) cuando la prevalencia de la enfermedad es baja (menor que 5 a 10%) (los momios [a/b o c/d] son casi iguales a las proporciones [a/a+b o c/c+d] cuando éstas son pequeñas, como en el caso de una prevalencia menor a 0.1) y su interpretación es similar; un valor de uno se interpreta como momios iguales o igual posibilidad de enfermar entre expuestos y no expuestos. Un valor mayor de uno significa que la posibilidad de enfermar es mayor en los expuestos que en los no expuestos. Un valor menor a uno significa que la posibilidad de enfermar es menor en los no expuestos que en los expuestos.

Para la interpretación de estas razones deben considerarse los sesgos inherentes a las encuestas transversales.

Cabe mencionar que, a pesar de la similitud en el análisis estadístico de los estudios transversales y de los de casos y controles, estos diseños difieren en cuanto a interpretación de medidas, sesgos y aproximación a la causalidad. Por ejemplo, un estudio transversal sobre el tabaquismo como factor etiológico de cáncer pulmonar comparando enfermos con no enfermos, equivaldría a un estudio de casos y controles con un grupo control excesivo (habría pocos casos de cáncer pulmonar en la población), con información sobre tabaquismo en un periodo inapropiado (la historia previa de tabaquismo sería más apropiada que el hábito tabáquico actual) y con verificación de casos sesgada (se captarían más casos de larga duración que de corta duración).³

La significancia estadística de la asociación entre las variables se hace a través del cálculo de intervalos de confianza y prueba de hipótesis de no asociación (RP o RM= 1). El lector interesado puede encontrar información complementaria en un artículo de revisión sobre estudios transversales, publicado en 1998, por García de la Torre y colaboradores.¹⁴

Debido a que frecuentemente la selección de sujetos se obtiene de la población general, el control de variables confusoras suele hacerse en el análisis de los datos (y no por diseño). Para ajustar el efecto de otras variables confusoras (definidas como variables asociadas con la variable de resultado y de exposición, que no se encuentran en la cadena causal entre éstas),¹⁵ es posible realizar análisis estratificado o emplear técnicas de análisis multivariado –regresión lineal múltiple, regresión logística– (figura 1). En el caso de análisis estratificado, el ajuste por variables confusoras se hace con el método de Mantel-Haenszel.⁹

La razón de momios de prevalencia ajustada se interpreta de manera similar a la razón de momios cruda, pero se le refiere como ajustada por o controlando el efecto de una tercera o más variables.

En el caso de variables de resultado continuas y variables de exposición categóricas podrá hacerse una comparación de medias con pruebas como la de *t* de Student o con análisis de varianza (Andeva). También es posible analizar con métodos estadísticos no paramétricos, algunos de los cuales se incluyen en la figura 1, aunque no serán descritos en este trabajo.

En el caso de variables continuas de resultado y de exposición, la relación lineal entre dos variables continuas puede llevarse a cabo con la estimación de la correlación de Pearson, o bien, mediante análisis de regresión lineal simple. Este último método permite establecer una relación lineal entre variables de exposición (predictora, independiente x_i) y una de resulta-

do (efecto, dependiente y); así, $y = a + b \cdot x_{i...}$, donde a y b son las constantes estimadas a partir de los datos.¹⁶

Para el ajuste de variables confusoras o para evaluar modificación del efecto o interacción es posible también emplear técnicas de análisis multivariado, ya sea usando regresión lineal múltiple (para variables de exposición categóricas y continuas y variables de resultado continuas) o regresión logística (para variables de resultado dicotómicas) y multinomial (para variables ordinales y nominales con más de dos categorías). Es importante mencionar que la construcción de modelos multivariados –la selección de variables a incluir en el modelo– para el análisis de encuestas transversales se debe guiar por la teoría, por el tipo de muestreo y por análisis cuidadosos de bondad de ajuste y del efecto de cada variable en el modelo. Por esta razón, los métodos de regresión gradual (*stepwise*) anterógrada o retrógrada, los cuales se basan únicamente en el incremento del valor explicativo del modelo, deben utilizarse con sumo cuidado.

La regresión logística permite obtener razones de momios crudas y ajustadas por potenciales confusores. Para estos estimadores crudos y ajustados es posible hacer pruebas de hipótesis para evaluar si son iguales al valor nulo (1), calcular intervalos de confianza, así como realizar pruebas de bondad de ajuste.¹⁶

El valor p convencionalmente aceptado para rechazar una hipótesis nula es $\alpha = 0.05$ o menor y su interpretación (frecuentista) es de que, si hiciéramos 100 estudios similares, se obtendrían por azar los mismos resultados o resultados más extremos en cinco estudios, por lo que la probabilidad de que éstos se deban al azar sería muy baja. Se debe considerar que el valor de p es función del estimador a obtener y del tamaño de muestra. Con un valor de confianza generalmente de 95%, los intervalos de confianza (IC) nos dan valores mínimos y máximos del estimador obtenido (prevalencia, razón de prevalencias, razón de momios, diferencia de medias, correlación de Pearson, coeficiente beta de la regresión lineal); se interpretan como un 95% de confianza de obtener cuando menos el límite inferior del estimador, hasta un límite máximo del mismo. Si el intervalo no incluye el valor nulo del estimador dado, existe asociación estadística y el valor p será significativo a un α de 0.05. La amplitud del intervalo es función del tamaño muestral. La información que proporciona el valor p y los intervalos de confianza es complementaria, aunque el valor límite de 0.05 para p es arbitrario y el IC más informativo; una p mayor a 0.05 en presencia de un intervalo de confianza para una razón de momios de 0.7 a 20 significa que aunque no tenemos p

estadísticamente significativa, la razón de momios podría ser hasta de 20 si se tuviera un tamaño de muestra mayor.

Las pruebas de bondad de ajuste evalúan la capacidad probabilística del método estadístico para ajustar o explicar la relación y variabilidad entre las variables incluidas en el modelo explicativo de exposición y efectos. Su significancia estadística se evalúa con valores de *p*.

En ocasiones, una muestra no refleja la composición de la población de la cual fue extraída directamente. Algunos grupos pueden estar sub o sobrerrepresentados. Una alternativa es emplear ponderadores, o valores que hacen variar el peso de cada observación en la muestra para que tenga una composición similar a la de la población base. En estos casos, es indispensable tomar en cuenta los ponderadores antes de proceder a cualquier análisis de la muestra. Otro problema relacionado con la no representatividad de la muestra es el conocido como falacia de Berkson (cuadro I).

En el ejemplo mostrado en el cuadro I es notorio que una mayor proporción de pacientes obesos con cardiopatía coronaria (CC) y de no obesos con CC, participó en el estudio. Esta falta de representatividad de la muestra produjo una asociación espuria entre la obesidad y la CC, derivada de una selección inadecuada de los sujetos de estudio. La selección aleatoria hubiera tendido a obtener proporciones de sujetos participantes en el estudio en cada celda similares a las de la población base, lo cual habría evitado este problema.

Frecuentemente, en las encuestas transversales se exploran múltiples exposiciones y variables, muchas veces no consideradas *a priori*; esto implica el análisis de múltiples subgrupos. En las bases de datos de encuestas transversales frecuentemente se analizan diversas hipótesis que no fueron establecidas al inicio del estudio. Con cada nueva hipótesis que se analiza aumenta la probabilidad de encontrar asociaciones estadísticamente significativas por azar. Existen métodos (por ejemplo el de Bonferroni)⁹ para tratar de corregir dicho problema, sin embargo, la naturaleza de los estudios transversales de por sí se limita a la exploración y generación de hipótesis.

El análisis de subgrupos dentro de un estudio transversal puede reducir mucho el número de individuos analizados, con la consiguiente pérdida de poder, sobre todo si se recuerda que la muestra se obtiene con base en una hipótesis en el grupo total. Su correcta interpretación deberá ser en el sentido de falta de significancia estadística al nivel convencional

Cuadro I
EJEMPLO DE LA FALACIA DE BERKSON

Un cardiólogo desea estudiar a pacientes en alto riesgo de enfermedad coronaria, por lo que solicita a los médicos de consulta externa recluten pacientes con hipertensión arterial. Los médicos de consulta externa identifican a 3 000 pacientes con diagnóstico de hipertensión arterial y los invitan a participar en el estudio. Por diversos motivos, sólo 300 pacientes acceden a ser hospitalizados por un día para participar en el estudio.

Con el fin de estudiar la relación entre obesidad y cardiopatía coronaria (CC), el cardiólogo clasifica a los 300 pacientes de la manera siguiente:

	Obesos	No obesos	Total
CC	10	40	50
No CC	20	230	250
Total	30	270	300

Razón de momios 2.8; IC 95% 1.27 - 6.5; $\chi^2(1) = 6.67$ *p*<0.05

Como se puede ver, la posibilidad (momios de prevalencia) de CC en los sujetos obesos es casi tres veces mayor (RM= 2.87) en los obesos, en comparación con los no obesos. Esta relación es estadísticamente significativa a un alfa de 0.05 y el intervalo de confianza a 95% muestra valores de RM que excluyen el valor nulo.

Posterior a esto, el cardiólogo de alguna manera logra estudiar a la población de 3 000 pacientes hipertensos y encuentra lo siguiente:

	Obesos	No obesos	Total
CC	25	175	200
No CC	250	2 550	2 800
Total	275	2 725	3 000

Razón de momios 1.45; IC 95% 0.94 - 2.25; $\chi^2(1) = 2.86$ *p*>0.05

Para su sorpresa, la razón de momios disminuye casi a la mitad y desaparece la significancia estadística encontrada previamente. Además, el intervalo de confianza incluye el valor nulo de no asociación y señala que la razón de momios puede ser incluso de 0.94, esto es, que la obesidad pudiera "proteger" con un 6 % (1-0.94= 0.06) de menor posibilidad de padecer CC.

Esta falacia se debe a que las proporciones o porcentajes en los cuatro subgrupos de participantes en el estudio diferían de las proporciones de los subgrupos de la población total de 3 000 pacientes.

Comparando las tablas, note que:

- de los 25 pacientes obesos con CC, 10 participaron en el primer estudio (40%)
- de los 250 pacientes obesos sin CC, 20 participaron en el primer estudio (8%)
- de los 175 pacientes no obesos con CC, 40 participaron en el primer estudio (23%)
- de los 2 550 pacientes no obesos sin CC, 230 participaron en el primer estudio (9%)

por falta de poder, y no de ausencia de efecto o de asociación, además de la consideración de sesgos.

En encuestas con grandes muestras (como las encuestas nacionales donde se obtienen datos sobre miles

de sujetos) es fácil encontrar asociaciones estadísticamente significativas; sin embargo, se deberá considerar la significancia conceptual (etiología) de efectos pequeños, aunque estadísticamente significativos, y de efectos grandes, estadísticamente no significativos.

Tipos de encuestas transversales

Existen diversos tipos de encuestas transversales de salud. En México se han llevado a cabo diversos tipos de éstas, como las encuestas nacionales de salud,¹⁷ las de seroprevalencia,¹⁸ de adicciones,¹⁹ de nutrición y de enfermedades crónicas.²⁰

Consideraciones éticas y de bioseguridad

El participante en una encuesta deberá estar enterado y de acuerdo con el uso que se le dará a la información que proporcione. Se deberá garantizar la seguridad, confidencialidad y de ser posible el anonimato de la persona que proporciona los datos. Se debe evitar el uso de datos para fines diferentes a los que autorizó el sujeto de estudio. Las muestras biológicas de sangre, material genético, tejidos y otras, deberán ser utilizadas exclusivamente para los fines autorizados por el sujeto de estudio y los residuos biológicos se deben manejar de manera adecuada con apego a las normas establecidas para ello. El uso de este material con objetivos de investigación distintos a los autorizados, aun años después de almacenamiento, requiere del consentimiento del donador y algunos comités de ética ni siquiera permiten solicitar nuevamente autorización al sujeto de estudio para usar sus datos con otros fines de investigación.

Por último, es responsabilidad del investigador asegurarse de la calidad de los datos, tanto de aquellos obtenidos a través de entrevistas o cuestionarios como de los correspondientes a mediciones de laboratorio, por medio de sistemas de control de calidad. Una vez recolectados los datos, su manejo, análisis e interpretación se deben realizar de acuerdo con el protocolo de estudio y se deberá evitar la manipulación de los mismos hasta obtener resultados "interesantes" o convenientes.

Conclusiones

Las encuestas transversales son un diseño de investigación ampliamente utilizado. Entre sus ventajas podemos mencionar su bajo costo y rapidez, ya que no requieren del seguimiento de los sujetos de estudio. Este diseño permite explorar múltiples exposiciones y efectos, generar hipótesis y datos útiles para la pla-

neación y gerencia de los servicios de salud, así como realizar mediciones de carga de la enfermedad. No obstante, este diseño también tiene algunas limitaciones como son la imposibilidad de establecer causalidad, la falta de temporalidad de la asociación exposición-efecto (salvo en algunos casos donde por razones teóricas es obvio que la exposición antecede a la variable de resultado), la dificultad para establecer valores basales para su comparación entre poblaciones y periodos de tiempo y su limitada utilidad para estudiar enfermedades de corta duración o poco frecuentes.

Referencias

- Gordis L. Epidemiology. Filadelfia: WB Saunders Co., 1996.
- Hennekens CH, Buring JE. Epidemiology in medicine. Boston: Little, Brown and Co., 1987.
- Rothman KJ, Greenland S. Modern epidemiology. 2a. edición: Lippincott-Raven, 1998.
- Dos Santos-Silva I. Estudios transversales. En: Dos Santos-Silva I. Epidemiología del cáncer: principios y métodos. Lyon, (Francia): Agencia Internacional de Investigación sobre el Cáncer/Organización Mundial de la Salud, 1999:225-244.
- Walker AM. Observation and inference: An introduction to the methods of epidemiology. Boston (MA): Epidemiology Resources, Inc., 1991.
- Gordis L. Epidemiology. Filadelfia: WB Saunders Co., 1996.
- Campbell D, Stanley J. Diseños experimentales y cuasi-experimentales en la investigación social. Buenos Aires: Amorrortu, 1982.
- Altman D. Practical statistics for medical research. Londres (UK): Chapman & Hall Ed., 1997.
- Pagano M, Gauvreau K. Principles of biostatistics. Wadsworth, Belmont (CA): Duxbury Press, 1993.
- Babbie E. The practice of social research. 3a. edición. Belmont (CA): Wadsworth, Inc., 1983.
- Aday L. Designing and conducting health surveys: A comprehensive guide. San Francisco: Jossey Bass, 1989.
- Fowler FJ. Survey research methods. Applied social research methods series. 2a. edición. Londres: Sage Publications, 1993; vol. 1.
- Kelsey JL, Whittemore AS, Evans AS, Thompson WD. Methods in observational epidemiology. Monographs in epidemiology and biostatistics. Nueva York: Oxford University Press, 1996.
- García-de la Torre GS, Huerta-Alvarado SG. Consideraciones metodológicas y análisis simple de los estudios transversales. Bol Med Hosp Infant Mex 1998;55:348-356.
- Lilienfeld DE, Stolley PD. Foundations of epidemiology. 3a. edición. Nueva York: Oxford University Press, 1994.
- Selvin S. Statistical analysis of epidemiologic data. 2a. edición. Nueva York: Oxford University Press, 1996.
- Instituto Nacional de Salud Pública, Centro de Investigaciones en Salud Pública. Encuesta Nacional De Salud II, 1994. Salud Publica Mex 1994;36:562.
- Tapia-Conyer R, Gutiérrez G, Sepúlveda J. Metodología de la Encuesta Nacional de Salud. Salud Publica Mex 1992;34:124-135.
- Tapia-Conyer R, Cravioto P, Borges-Yáñez A, De la Rosa B. Consumo de drogas médicas en población de 60 a 65 años en México. Encuesta Nacional de Adicciones 1993. Salud Publica Mex 1996;38:458-465.
- Guerrero-Romero JF, Rodríguez-Morán M. Prevalencia de hipertensión arterial y factores asociados en la población rural marginada. Salud Publica Mex 1998;40:339-346.

Estudios epidemiológicos de casos y controles. Fundamento teórico, variantes y aplicaciones

Eduardo Lazcano-Ponce, Dr. en C.,⁽¹⁾ Eduardo Salazar-Martínez, Dr. en C.,⁽¹⁾
Mauricio Hernández-Avila, Ph. D.⁽¹⁾

“El objetivo principal de un estudio de casos y controles es proveer una estimación válida y razonablemente precisa, de la fuerza de asociación de una relación hipotética causa-efecto”.

Philip Cole¹

Los estudios de casos y controles fueron empíricamente empleados por John Snow en el siglo XIX durante sus investigaciones sobre las causas de la epidemia de cólera, cuando comparó casos y no casos en cuanto a su lugar de residencia y fuente de agua potable.² Este tipo de diseño también fue utilizado por Lane-Clayton, en 1926, en un reporte sobre factores reproductivos y cáncer de mama.³ Sin embargo, es hasta los años cincuenta cuando se identifica como un diseño epidemiológico específico, en los trabajos reportados por Cornfield,⁴ y Mantel y Haenszel;⁵ estos autores proporcionaron las primeras bases metodológicas y estadísticas para su aplicación y análisis. Finalmente, en los años setenta, Miettinen⁶ establece la concepción moderna de este tipo de estudios, presentando las bases teóricas que establecen la estrecha relación que existe entre este tipo de diseño y los estudios tradicionales de cohorte.

Con estos antecedentes es posible afirmar que información derivada de diferentes estudios de casos y controles ha sido notoriamente útil para modificar políticas de salud y avanzar en el conocimiento médico. A este respecto, los estudios de casos y controles se han empleado exitosamente para evidenciar la asociación entre consumo de cigarrillos y el riesgo de cáncer de pulmón,⁷ su interacción con la exposición al asbes-

to con la elevada frecuencia de mesoteliomas,⁷ así como el antecedente de consumo de estrógenos (dietilestilbestrol), durante el primer trimestre del embarazo, por las madres de adolescentes en las que se identificó cáncer de vagina.⁸ Si bien se pudiera pensar que el diseño de cohorte conjunta los factores idóneos para la observación epidemiológica, su realización está seriamente limitada por la ausencia de poblaciones especiales en quienes construirla y, frecuentemente, por la carencia de tiempo o recursos financieros necesarios para estudiar los grandes grupos poblacionales que se requieren para el estudio de enfermedades poco frecuentes; por esta razón, los estudios de casos y controles se constituyen en una alternativa costo-efectiva para identificar factores de riesgo y generar hipótesis para estudios subsecuentes; al respecto, tiene diversas ventajas y desventajas que se resumen en el cuadro I.

Los estudios de casos y controles representan una estrategia muestral, en la que de manera característica se selecciona a la población en estudio con base en la presencia (caso) o ausencia (control o referente) del evento de interés. Es común que se utilicen sistemas de registro de eventos relacionados con la salud, registros de padecimientos, listados de pacientes hospitalizados, etcétera, para identificar y seleccionar de

(1) Centro de Investigación en Salud Poblacional. Instituto Nacional de Salud Pública. Cuernavaca, Morelos, México.

Cuadro I
**VENTAJAS Y DESVENTAJAS DE LOS ESTUDIOS
 DE CASOS Y CONTROLES**

Ventajas

1. Útiles para estudiar problemas de salud poco frecuentes
2. Indicados para el estudio de enfermedades con un largo periodo de latencia
3. Suelen exigir menos tiempo y ser menos costosos que los estudios de cohorte
4. Caracterizan simultáneamente los efectos de una variedad de posibles factores de riesgo del problema de salud que se estudia
5. No es necesario esperar mucho tiempo para conocer la respuesta
6. Requiere de menor número de sujetos en quienes se puede profundizar
7. Estima cercanamente el riesgo relativo verdadero, si se cumplen los principios de representatividad, simultaneidad y homogeneidad

Desventajas

1. Especialmente susceptible a sesgos porque:
 - La población en riesgo a menudo no está definida (a diferencia de los estudios de cohorte)
 - Los casos seleccionados por el investigador se obtienen a partir de una reserva disponible
 - Es difícil asegurar la comparabilidad de factores de riesgo poco frecuentes
 - Pueden generar frecuentemente sesgos de información, debido a que la exposición –en la mayoría de los casos– se mide, se reconstruye o se cuantifica, después del desarrollo de la enfermedad
 - Se puede introducir un sesgo de selección, si la exposición de interés determina diferencialmente la selección de los casos y los controles
2. El riesgo o la incidencia de la enfermedad no se puede medir directamente, porque los grupos están determinados no por su naturaleza sino por los criterios de selección de los investigadores
3. Si el problema de salud en estudio es muy prevalente (mayor de 5%), la razón de momios no ofrece una estimación confiable del riesgo relativo
4. No sirven para determinar otros posibles efectos de una exposición sobre la salud, porque se ocupan de un solo resultado
5. Inapropiados cuando la enfermedad bajo estudio se mide en forma continua

manera costo-efectiva los casos; también, que una vez delimitada la población fuente –definida como aquella de donde se originan los casos–, se utilice esta misma para la selección de los controles; estos últimos deberán representar de manera adecuada a los miembros de la población fuente que no desarrollaron el evento en estudio. Una vez seleccionados los casos y los controles se compara la exposición relativa de cada grupo a diferentes variables o características que pueden tener relevancia para el desarrollo de la condición o enfermedad.

En teoría, los estudios de casos y controles se basan en la identificación de los casos incidentes en una determinada población durante un periodo de observación definido, tal y como se lleva a cabo en los es-

tudios de cohorte. La diferencia estriba en que en el estudio de casos y controles se identifica la cohorte, se identifica a los casos y se obtiene una muestra representativa de los individuos en la cohorte que no desarrollaron el evento en estudio; esto último tiene el propósito de estimar la proporción de individuos expuestos y no expuestos en la cohorte o población base, evitando de esta manera la necesidad de determinar la presencia de la exposición en todos los miembros de la población o cohorte en estudio.⁹

En este sentido, la principal diferencia entre los diseños de cohorte y de casos y controles se encuentra en la selección de los sujetos de estudio. Un estudio de cohorte selecciona sujetos a partir de la exposición; se parte de un grupo de individuos inicialmente exentos de la enfermedad o evento de estudio y se les sigue en el tiempo, con el fin de registrar la ocurrencia del evento. En contraste, en el estudio de casos y controles se selecciona a los sujetos de estudio en función de la presencia o ausencia de la enfermedad o evento en estudio. Esto es lo que constituye el paradigma de este tipo de diseños, y repercute ampliamente en su interpretación, aplicación y principales limitantes.

En los estudios de cohorte se comparan dos o más grupos de exposición y se estima la posibilidad o riesgo de tener el evento o enfermedad en función de la condición o exposición estudiada, es decir, se parte de la causa al efecto. A diferencia de los estudios de casos y controles, en los que se compara un grupo de sujetos con el evento en estudio y un grupo control o referente (sin el evento en estudio) y, posteriormente, se estiman en estos grupos las diferencias en la exposición (figura 1); consecuentemente, en este tipo de estudios se parte del efecto en busca de la posible causa, por lo que se considera que no cuentan con una relación de causa-efecto correcta y, por esta razón, pueden ser susceptibles a diferentes errores o sesgos en su interpretación. Otra limitación de estos estudios es que, en general, no se pueden estimar de manera directa las medidas de incidencia o prevalencia que tradicionalmente se obtienen en los estudios de cohorte o transversales.

En los estudios de casos y controles sólo es posible estimar seudotasas, conocidas como momios, mismas que indican la frecuencia relativa de la exposición o condición en estudio entre los casos y los controles; la seudotasa de exposición en los casos se estima dividiendo los casos expuestos sobre los no expuestos; de manera similar, la seudotasa de exposición en los controles se estima dividiendo los controles expuestos entre los no expuestos, el cociente de estas seudotasas se conoce como la razón de momios (RM) o momios relativos. La razón de momios bajo ciertas suposiciones que se detallan más adelante, puede ser

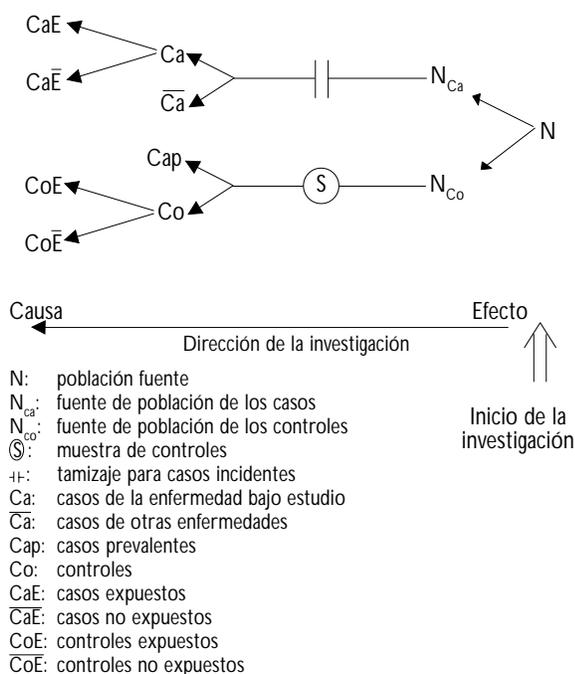


FIGURA 1. DISEÑO CLÁSICO DE UN ESTUDIO DE CASOS Y CONTROLES

un estimador no sesgado de la razón de tasas de incidencia o del riesgo relativo –la medida de asociación que tradicionalmente se deriva de los estudios de cohorte y que se utiliza para valorar la asociación entre una exposición y un evento en este tipo de estudios.

Estimación del riesgo relativo y fuente de obtención de casos

Actualmente se reconoce que la razón de tasas de incidencia es la mejor medida de riesgo, esto es, la probabilidad de sufrir un evento; y que la razón de incidencia acumulada es una buena medida para conocer en términos “relativos” lo que aumenta o disminuye dicho riesgo en presencia o ausencia de cierta exposición o condición. Sin embargo, con fines de reconocimiento de causas, resulta de suma importancia práctica la estimación de la razón de momios que se obtiene en los estudios de casos y controles, que en circunstancias específicas se considera como un buen estimador del riesgo relativo (RR). Los estudios de casos y controles se pueden conceptualizar como una estrategia metodológica para estudiar una cohorte; dentro de este contexto pueden existir diferentes alternativas para la selección de los casos. La situación ideal es la que

más se acerca a la cohorte, siguiendo el paradigma de los estudios longitudinales; en este sentido, se recomienda seleccionar a los casos conforme se diagnostican y aparecen en el sistema de registro utilizado, es decir, la población de casos queda compuesta principalmente por casos incidentes o casos nuevos de diagnóstico reciente. Así, en el extremo opuesto del paradigma, lo que se podría considerar como menos recomendable es seleccionar a los casos existentes en un punto en el tiempo, es decir, a los casos prevalentes o casos sobrevivientes al momento de iniciar el estudio. A continuación se describen algunas de las diferentes opciones en la selección de casos.

Utilización de casos incidentes con periodos de exposición o latencia prolongados. La razón de momios tiende a parecerse al RR cuando los casos del estudio son incidentes y la exposición que la precede es de larga duración. Este tipo de casos tiene tres ventajas en comparación con los casos prevalentes: a) se puede disminuir el sesgo de memoria porque el sujeto puede recordar mejor la experiencia pasada por ser más reciente; b) además, la supervivencia no está condicionada por los factores de riesgo como pudiera ocurrir en los casos prevalentes, y c) es menos probable que el estatus de enfermedad pueda modificar la exposición que se está estudiando; por ejemplo, asociación entre infección por virus de papiloma humano y cáncer cervical.¹⁰

Utilización de casos prevalentes con periodos de exposición prolongados. La razón de momios se parece al riesgo relativo si, a pesar de utilizar casos prevalentes, el periodo de exposición es muy largo y la enfermedad no afecta el estado de exposición. Los casos prevalentes pueden ser incluidos especialmente cuando no se dispone de casos nuevos porque la enfermedad es muy rara y tiene baja letalidad, y cuando la exposición no modifica el curso clínico (sobrevivida) de la enfermedad, como es el caso de enfermedades de predisposición genética. Ejemplo, gen BC y cáncer de mama.¹¹

Utilización de casos incidentes y periodos de exposición muy cortos. La razón de momios es similar al riesgo relativo cuando el periodo de riesgo es muy corto, y se utilizan casos incidentes. Ejemplo, un brote de intoxicación alimentaria.¹²

Utilización de casos prevalentes. La razón de momios se aproxima al riesgo relativo cuando la prevalencia de casos es muy pequeña, sólo si el evento resultado no está relacionado con la sobrevivida antes de la selección, condición o exposición, y si la enfermedad no afecta el estado de exposición. Ejemplo, mesotelioma y exposición al asbesto.¹³

Utilización de casos o controles fallecidos. La inclusión de sujetos muertos sólo se justifica en algunas exposi-

ciones que se puedan cuantificar mediante el uso de fuentes secundarias de datos de alta calidad, como pudiera tratarse de historias clínicas o registros de fuentes de información ocupacional. En este caso, como no están siendo estudiados factores pronósticos y de sobrevida –como pudiera ser un objetivo de un estudio de cohorte–, si el individuo está vivo o muerto carece de relevancia.

Selección de los controles

El grupo control o referente se utiliza fundamentalmente para estimar la proporción de individuos expuestos y no expuestos en la población base que da origen a los casos. Por esta razón, los procedimientos para la selección de los controles quedarán definidos en la medida que se expliciten claramente los criterios de selección de los casos, así como de la población de donde se originan éstos. Generalmente, el grupo control más apropiado corresponde a la subpoblación de individuos que está en riesgo de desarrollar el evento en estudio y de los cuales se puede asegurar, con relativa certeza, que en el caso teórico de que desarrollaran el evento o enfermedad en estudio quedarían incluidos, dentro de la población en estudio, en el grupo de casos.

Cuando los casos se obtienen de una población claramente definida en tiempo, espacio y lugar, y éstos constituyen un censo de los eventos en estudio o una muestra representativa de los mismos, la selección de controles se puede realizar mediante un muestreo aleatorio simple de la población base. En este caso, la selección es un simple procedimiento técnico que no introduce ningún sesgo más allá de los errores muestrales que pudieran existir al utilizar la totalidad de la base poblacional como marco muestral para la selección de los controles. Como ejemplo de esta situación podemos citar el estudio realizado en la ciudad de México por Romieu y colaboradores.¹⁴ En el mencionado estudio, los autores conformaron el grupo de casos con una muestra representativa de los casos de cáncer de mama diagnosticados en la ciudad de México durante 1990 a 1992. Para la selección del grupo control seleccionaron una muestra representativa de las mujeres residentes en la ciudad de México durante el mismo periodo en el que se seleccionaron los casos. La suposición en lo que se refiere al grupo control es que, en caso de que las mujeres seleccionadas en este grupo hubiesen sido diagnosticadas con cáncer de mama, éstas habrían sido detectadas y podrían haber sido estudiadas como casos. Si esta suposición se cumple, podemos entonces afirmar que el grupo control seleccionado representa adecuadamente a la po-

blación base y puede ser utilizado para hacer inferencias válidas sobre la proporción relativa de sujetos expuestos y no expuestos en la población base que dio origen a los casos.

En otro estudio, también realizado en la ciudad de México, Pérez-Padilla y colaboradores¹⁵ reportaron resultados sobre los factores de riesgo para enfermedad pulmonar obstructiva crónica (EPOC). Los casos en este estudio fueron definidos como mujeres con diagnóstico reciente de EPOC, que acuden para atención médica al Instituto Nacional de Enfermedades Respiratorias (INER). Puesto que el INER es un centro nacional de referencia para padecimientos pulmonares en México, la población que acude a este centro hospitalario en busca de atención médica no está bien caracterizada, razón por la cual los autores no lograron definir una población base en tiempo, espacio y lugar, que les permitiera realizar directamente la selección de los controles. Para superar este problema los autores eligieron como controles a otros pacientes que acuden al INER en busca de atención médica, pero por otros padecimientos no relacionados con el EPOC. En este caso, la suposición de que el grupo control representa adecuadamente la base de donde se originaron los casos de EPOC, se cumple siempre y cuando se pueda suponer que, en caso hipotético de haber desarrollado EPOC, en lugar del padecimiento que los llevó al INER, los controles también habrían acudido al INER y habrían sido incluidos en la lista de casos. En la medida en que esta suposición se cumpla, el grupo control será adecuado para estimar la proporción de expuestos y no expuestos en la población base que dio origen a los casos.

Por lo tanto, las consideraciones básicas para la selección de los controles incluyen:

1. Los controles deben ser seleccionados de la misma base poblacional (de la cohorte imaginaria) de donde se originaron los casos. Operacionalmente este último concepto quiere decir que en el supuesto de que el control desarrollara la enfermedad o evento en estudio, éste necesariamente tendría que aparecer en la lista de casos.
2. Los controles deben ser seleccionados independientemente de su condición de expuestos o no expuestos para garantizar que representen adecuadamente a la población base. Esto último se logra siempre y cuando la condición de exposición no determine la posibilidad de que un individuo sea o no incluido en el estudio como control, lo que implica que las fracciones muestrales para los controles expuestos y no ex-

- puestos, aunque la mayor parte de las veces desconocidas por el investigador, deben ser las mismas. En la medida en que éstas difieran, se introducirá un sesgo de selección y se comprometerá la validez interna de la investigación.
3. La probabilidad de selección para los controles debe ser proporcional al tiempo que el sujeto permaneció elegible para desarrollar el evento o enfermedad en estudio. Así, un individuo que migró o que falleció durante el estudio dejará de ser elegible como control. Una manera de operacionalizar este concepto es seleccionando un control del grupo de individuos elegibles cada vez que se detecta o selecciona un caso; esto se conoce como selección por grupo en riesgo. En teoría, utilizando este esquema de selección se asegura que los controles están en riesgo de desarrollar el evento en el momento en que son seleccionados. Este esquema también indica que un sujeto seleccionado como control en una etapa temprana del estudio, podría también ser seleccionado como caso en etapas posteriores del estudio.
 4. En la selección de los controles se debe evitar, en la medida de lo posible, los factores de confusión. Se espera que el grupo control sea similar al grupo de casos en lo que se refiere a otras variables que pudieran ser factores de riesgo para el desarrollo del evento y al mismo tiempo estar asociados con la exposición. Una estrategia frecuentemente utilizada para lograr este requisito es el pareamiento o igualación de atributos. Este esquema de selección implica un segundo requisito de elegibilidad para el control. Por ejemplo, si al seleccionar el control se decidiera parear por edad y género, además de cumplir el criterio de estar en riesgo deberá cumplir el criterio de ser del mismo género y grupo de edad.
 5. La medición de variables debe ser comparable entre los casos y los controles. Todos los procedimientos para medir la exposición o los factores de confusión potenciales deben ser aplicados, reportados y registrados de la misma manera en casos y controles.

En este contexto, existen diferentes posibilidades y fuentes de obtención de controles, a saber:

Con base poblacional. Si los casos representan una muestra de todos los casos que ocurren en una población identificada y definida claramente en tiempo y espacio, y los controles se muestrean directamente de

esta misma población, los controles son definidos como base poblacional. Este tipo de controles son más factibles de utilizar cuando se toman casos de registros poblacionales o se cuenta con los suficientes recursos para obtenerlos directamente, como se ejemplifica en el cuadro II.

Controles vecindarios. Este tipo de controles puede ser utilizado cuando los investigadores deciden comparar los casos pareando con sus respectivos controles vecindarios. Es decir, después de que se identifica un caso, se seleccionan al azar uno o más controles que viven en la misma zona de residencia que los casos. Este tipo de controles puede generar sesgos de selección, ya que no siempre se puede asumir que son una fuente confiable de representación de la base poblacional de donde se originaron los casos.

Controles hospitalarios. Son sujetos que acuden al mismo hospital donde se realizó la selección de los casos, pero acuden a este sitio por un padecimiento diferente. La principal limitación de este grupo control es que puede existir una relación entre la exposición en estudio y la causa de demanda de atención. Es decir, en este tipo de controles los investigadores raramente pueden estar seguros de que la exposición no está re-

Cuadro II

EJEMPLO DE CASOS Y CONTROLES CON BASE POBLACIONAL

Objetivo

Evaluar la calidad del Programa de Detección Oportuna de Cáncer Cervical (PDOC) en el área metropolitana de la ciudad de México

Material y métodos

Casos y controles con base poblacional. Selección de 513 casos de Cáncer de Otero de ocho hospitales del área metropolitana de la ciudad de México y 1 007 mujeres seleccionadas de un muestreo aleatorio de 6 220 viviendas de la ciudad de México

En teoría:

- La muestra de estudio se obtuvo a partir de la población de la que se obtienen todos los casos incidentes
- Los controles se eligieron aleatoriamente entre los miembros de la misma población sin la enfermedad
- Aun cuando los casos se identificaron en ocho hospitales es razonable suponer que representan los casos en el área geográfica de estudio

Resultados

El PDOC en la ciudad de México carece de efecto (RM = 0.95; IC 95% 0.76 - 1.19).

El PDOC en la ciudad de México tiene un efecto si las mujeres asisten sin síntomas ginecológicos (RM= 0.51; IC 95% 0.39 - 0.67)

Fuente: referencia 16

lacionada con la enfermedad o el motivo de hospitalización. Por ello se necesita utilizar una variedad de diagnósticos posibles cuando se decide utilizar este tipo de controles, esto con el fin de eliminar sesgos de selección ocasionados al atenuar los efectos de incluir un grupo diagnóstico específico que puede estar relacionado con la exposición en estudio. Es decir, es probable que al padecer algún tipo de enfermedad difieran de los individuos sanos en una serie de factores que tienen relación con el proceso de enfermar, como pueden ser mayor prevalencia en el consumo de tabaco, alcohol o deficientes hábitos dietéticos, y estos factores pueden estar relacionados directa o indirectamente con la exposición en estudio.

Controles seleccionados aleatoriamente de números telefónicos. Frecuentemente utilizado en países desarrollados, es un muestreo de viviendas basado en una selección aleatoria del listado de números telefónicos de un área geopolítica. Cuenta con diversas limitaciones que incluyen la probabilidad diversa de contactar los sujetos elegibles, porque en las viviendas varía el número de personas que residen en la misma y es variable la cantidad de tiempo para que algún sujeto esté en casa. Otras limitaciones importantes son la cantidad de tiempo invertido para contactar a la población objetivo, que puede requerir adicionalmente de varias llamadas, así como la inclusión de números no residenciales que puede afectar la tasa de respuesta. La existencia de diversos números telefónicos en una misma casa y que utilicen contestadoras automáticas plantea problemas adicionales. En nuestro país la limitante más importante es el gran número de viviendas sin teléfono.

Controles con otras enfermedades de un registro poblacional. Para la obtención de este tipo de controles se pueden utilizar registros de base poblacional como los registros de tumores, de sistemas de vigilancia epidemiológica o de estadísticas vitales (fallecidos), de hospitales, de centros de atención primaria, de empresas y compañías de seguros, entre otros. La ventaja es la procedencia de una misma base poblacional, pero se debe tener cuidado para excluir algunos diagnósticos que puedan estar relacionados con la exposición.

Controles de amigos o familiares. Son personas relacionadas con los casos. Este grupo presenta la ventaja de reducir los costos y tienen una elevada probabilidad de que provengan de una misma base poblacional que los casos. Sin embargo, su principal inconveniente es el potencial riesgo de sobrepareamiento por algunos factores de exposición, dado que algunos hábitos de vida son compartidos, por lo que frecuentemente puede estar subestimado el posible efecto. Sin embargo, para trabajos en los que se estudian factores genéticos pueden constituir un buen grupo control.

Controles obtenidos del registro de mortalidad. En este caso los controles no son seleccionados directamente de la base poblacional, que necesariamente son personas vivas, pero siendo otro tipo de muertes que se presentan en la base poblacional, este tipo de controles pueden ser útiles si la distribución de la exposición entre los grupos es similar a la de la población fuente. Consecuentemente, este tipo de controles debe restringirse a las categorías de muerte que no están relacionadas con la exposición. Como había sido señalado previamente, estos controles son útiles cuando se ponen en práctica estudios de mortalidad proporcional.

Uso de controles del mismo tipo o controles de diferentes tipos. Múltiples controles de una misma base poblacional, tales como dos o más controles por cada caso, se pueden utilizar para incrementar el poder de estudio. Sin embargo, se acepta que se gana incremento en el poder solamente hasta un índice de un caso por cada cuatro controles.¹⁷ Otra posible situación es la de utilizar *múltiples controles de diferentes tipos*. Se puede uno encontrar frecuentemente con la disyuntiva de que la exposición de los controles hospitalarios usados en el estudio no representa la exposición "esperada" para personas no enfermas. En este caso, se puede elegir utilizar un grupo adicional de controles vecindarios o poblacionales, en espera de que los resultados obtenidos cuando los casos se comparan con controles hospitalarios sean similares a los resultados obtenidos cuando se comparan con otro tipo de controles. El problema es que si los hallazgos difieren, la razón de la discrepancia no se puede identificar fácilmente.

Métodos en la selección de casos y controles

Pareamiento. Es la selección de controles en función de una o varias características comunes a los casos, como el sexo, la edad y la condición socioeconómica, entre otras. Tiene como ventajas aumentar la eficiencia estadística, y disminuir el sesgo asociado a factores de confusión conocidos. Puede asegurar la homogeneidad por edad y sexo, y facilitar la comparación de casos y controles en presencia de exposiciones que varían con el tiempo. Dentro de las ventajas prácticas se encuentra la mayor facilidad de identificar a los controles. Asimismo, existen dos tipos fundamentales de pareamiento: individual o grupal (pareamiento por grupos de frecuencia). Esta estrategia facilita la detección de una interacción entre la exposición y el factor por el que se pareo, siempre que éste tenga un gran efecto sobre el riesgo de padecer la enfermedad y sea poco frecuente en la población.

Dentro de las desventajas de parear en un estudio de casos y controles se encuentra la de no poder analizar el posible efecto de riesgo de una variable de pareamiento, porque, por definición, son iguales para casos y controles. Asimismo, otra posible desventaja es la de sobrepareamiento, que consiste en reducción de la eficiencia del estudio, y se genera porque se para por una variable que es una condición intermedia en el camino causal entre exposición y enfermedad.

Es necesario destacar que los controles brindan información acerca de la distribución de la exposición en la cohorte de donde se originan los casos, y la exposición del control es relevante sólo para el día en que se haga un muestreo. Sin embargo, podemos afirmar que, en relación con la medición de la exposición, este tipo de estudio puede ser evaluado en forma retrospectiva o prospectiva. En forma retrospectiva, cuando los casos y controles son identificados en el presente y los factores de riesgo son determinados de la historia de los sujetos bajo estudio (historias clínicas o entrevistas para cuantificar eventos pasados).

En casos especiales, este diseño de estudio puede ser prospectivo; por ejemplo, cuando los casos y controles potenciales fueron identificados en el pasado y la exposición se determina posteriormente (utilizando, por ejemplo, un banco de sueros). Finalmente, si ambos, el evento resultado y los factores de exposición, son determinados al mismo tiempo, y los investigadores no conocen la condición de caso o no caso, se puede considerar como un estudio transversal.

Variantes del diseño de casos y controles

Los casos como los controles, en teoría, deben tener características de representatividad, simultaneidad y homogeneidad. Representatividad significa que los casos deben representar a todos los casos existentes en un tiempo determinado, y que los sujetos que se

seleccionen como controles debieran representar en el estudio a los sujetos que se pueden convertir en casos, y proceder de la misma base poblacional. Asimismo, simultaneidad significa que los controles deben obtenerse en el mismo tiempo de donde surgieron los casos. Finalmente, homogeneidad significa que los controles se deben obtener de la misma cohorte de donde surgieron los casos e independientemente de la exposición bajo estudio.

Estas tres características se describen en el cuadro III, donde se observan las limitaciones inherentes que por definición tienen las diversas variantes de diseños de casos y controles, y donde cualquier factor que se aleje de estos principios producirá que la medida de efecto –RM– se sobre o subestime, afectando la validez del estudio.

Durante los últimos años se ha avanzado considerablemente sobre los aspectos metodológicos de los estudios de casos y controles; en particular se ha conceptualizado más claramente la estrecha relación que existe entre estos estudios y los de cohorte, lo que ha permitido el desarrollo de esquemas de muestreo del tipo de casos y controles, con el fin de reducir los costos que implica la realización de los grandes estudios de cohorte. Recientemente se han propuesto y formalizado, desde el punto de vista estadístico, variantes que inciden sobre la selección de los casos o la definición del grupo control; a continuación se describen las más utilizadas.

- *Estudios caso-cohorte.* En esta variante, la definición de casos y controles se encuentra anidada en una cohorte fija, bien definida en tiempo, espacio y lugar, en la cual existe el interés de estimar la razón de incidencia acumulada y es razonable asumir que todos los miembros de la cohorte tendrán el mismo tiempo de seguimiento. Para realizar un estudio de caso-cohorte se requiere llevar a

Cuadro III
CONDICIONES DE LOS ESTUDIOS DE CASOS Y CONTROLES

Tipo de estudio de casos y controles	Representatividad		Simultaneidad	Homogeneidad
	Casos	Controles		
Caso-cohorte	Sí	Sí	Asegurada	Definitiva
Caso-caso	Azar	Azar	Asegurada	Definitiva
Casos y controles anidado en una cohorte	Sí	Sí	Asegurada	Definitiva
Casos y controles con base poblacional	Azar	Sí	Posible	Posible
Casos hospitalarios y controles con base poblacional	No	Sí, siempre que se obtengan de un marco muestral de la población de la que surgen los casos	Desconocida	Desconocida
Casos hospitalarios y controles vecindarios	No	No	Desconocida	Desconocida
Casos y controles hospitalarios	No	No	Desconocida	Desconocida

cabo los pasos que se describen en la figura 2. En un primer tiempo se define la cohorte o población en estudio; en un segundo paso se selecciona el grupo control que se utilizará para estimar la proporción de individuos expuestos y no expuestos que se encuentran en riesgo de desarrollar el evento al inicio del estudio y, posteriormente, se realiza el seguimiento de la cohorte, con el fin de detectar los eventos (casos incidentes) que se desarrollan a lo largo del tiempo y caracterizarlos en términos de su pertenencia al grupo expuesto o no expuesto. Es evidente que al usar este tipo de selección, un sujeto inicialmente identificado como control podría desarrollar el evento de interés durante el seguimiento y ser seleccionado como caso. Esta última situación, cuando ocurre con frecuencia, puede convertirse en una limitante importante y comprometer el poder estadístico del estudio. Por esta razón este tipo de estrategia se recomienda para el estudio de enfermedades poco frecuentes, en cohortes fijas, claramente definidas, donde la determinación de la exposición en todos los miembros de la cohorte resultaría muy costosa (cuadro IV).

- *Estudios de casos y controles anidado o de grupo de riesgo.* En esta variante se utiliza un esquema de muestreo conocido como de grupo de riesgo, ya que la elegibilidad de un individuo como control depende de que éste se encuentre en riesgo,

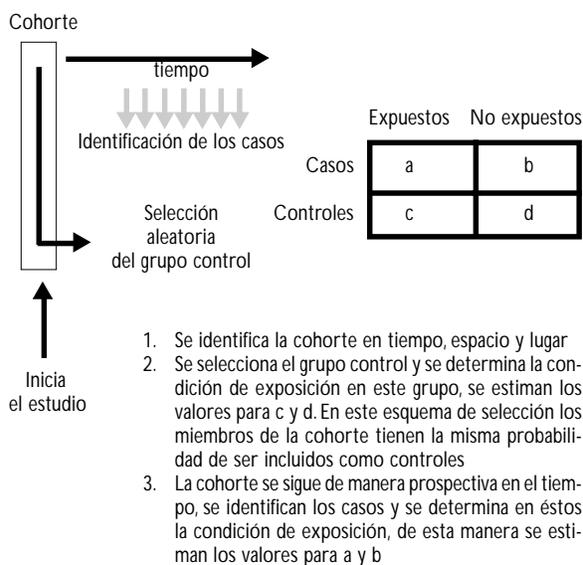


FIGURA 2. DISEÑO DE CASOS Y CONTROLES DE TIPO CASO-COHORTE

**Cuadro IV
EJEMPLO DE ESTUDIO DE CASO-COHORTE**

Objetivo
Investigar si las características antropométricas se relacionan con cáncer de próstata (CP)

Material y métodos
Estudio de cohorte con una medición basal en 1986 de 58 279 hombres entre 55-69 años. Después de 6.3 años de seguimiento, analizaron 681 casos, de CP y 1 565 miembros de la subcohorte. Se excluyó a hombres con medidas antropométricas incompletas

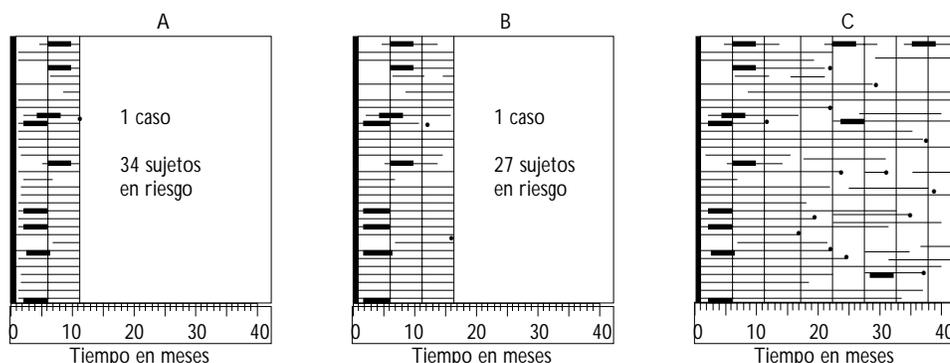
Exposición
Se estimó el índice de composición corporal a la edad de 20 años en relación con la medición basal

Resultados
El índice de masa corporal (IMC) no se asoció significativamente con cáncer de próstata. Sujetos con IMC a los 20 años <19 fueron categorizados como referencia. Con IMC entre 19-20, la RM fue de 1.06; entre 21-22 fue de 1.09; entre 23-24, de 1.39 y, finalmente, IMC >25 tuvo una RM de 1.33. La tendencia de riesgo, sin embargo, fue significativa ($p=0.02$)

Conclusiones
Los hallazgos sugieren que el IMC en adultos jóvenes puede ejercer un efecto sobre el riesgo posterior de cáncer de próstata

* Fuente: referencia 18

es decir, sea miembro de la cohorte en el momento en que se selecciona o identifica el caso. Los casos y el conjunto de individuos en riesgo que no desarrollaron el evento constituyen el grupo de riesgo. En esta variante de casos y controles es frecuente asumir que la selección de los casos y controles se realiza anidada dentro de una cohorte dinámica, donde los sujetos de estudio permanecen en la cohorte durante tiempos variables y en los que la exposición puede tomar valores diferentes en el tiempo. Para realizar un estudio de casos y controles anidado se requiere llevar a cabo los pasos que se describen en la figura 3. En un primer tiempo se define de manera conceptual la cohorte o población en estudio; se realiza el seguimiento de la misma con el fin de detectar los eventos que ocurren (casos incidentes) a lo largo del tiempo, y cada vez que se selecciona o se identifica un caso se selecciona uno o varios controles de la población que en ese momento particular se encontraba en riesgo de desarrollar el evento en estudio. Es evidente que al usar este tipo de selección, un sujeto inicialmente identificado como control podría desarrollar el evento de interés durante el seguimiento y posteriormente ser seleccionado como caso. Esta situación



En los paneles A, B y C se representa gráficamente la realización de un estudio de casos y controles anidado en una cohorte dinámica. En el tiempo 0 se definen los criterios de elegibilidad para la cohorte. En el panel A se observa que el primer caso incidente se detecta a los 10 meses de haber iniciado el estudio; en ese momento en la cohorte existían en riesgo 34 sujetos; este grupo forma la población base para la selección de los controles, es decir, el grupo de riesgo. En el panel B se observa que el segundo caso se detecta a los 16 meses de seguimiento y que el grupo de riesgo ha variado en su composición; en este punto en el tiempo existen 27 sujetos elegibles como controles. Finalmente en el panel C se presenta el resultado final de la definición variable de los grupos de riesgo, que cambia conforme ocurren y se selecciona a los casos.

FIGURA 3. DISEÑO DE CASOS Y CONTROLES ANIDADADO O DE GRUPO DE RIESGO

en general ocurre con poca frecuencia. Sin embargo, en ese diseño en particular podría ocurrir que un individuo fuera inicialmente seleccionado como control y posteriormente como caso. Esta situación no es fuente de error o de sesgo, ya que en los estudios de cohorte un mismo individuo puede contribuir tanto al numerador como al denominador y esta misma situación se mantiene en este tipo de estrategia. El diseño de casos y controles anidado, o de grupo de riesgo, se recomienda para el estudio de enfermedades poco frecuentes, en cohortes dinámicas en las que la determinación de la exposición y sus cambios en el tiempo, en todos los miembros de la cohorte, resultaría muy costosa (cuadros V y VI).^{19,20}

- *Estudios de caso-caso.* Este tipo de diseño es una manera científica de cuestionar y responder la pregunta ¿estuvo el sujeto haciendo algo inusual justo antes del inicio de la enfermedad o evento de estudio? Para responder esta pregunta, se necesita hacer la comparación entre el mismo individuo, esto es, en este diseño se compara la exposición a ciertos agentes durante el intervalo en que el evento no ha ocurrido (periodo de control), con la exposición durante el intervalo en que el evento ocurre (periodo de riesgo). En realidad este tipo de diseño se puede conceptualizar como un estudio de casos y controles pareado, en donde cada uno de los individuos sirve como su propio control.²¹ Pueden existir diversos eventos de exposición también conocidos como periodo de efecto

Cuadro V EJEMPLO DE CASOS Y CONTROLES ANIDADADOS EN UNA COHORTE

Objetivo	Evaluar la posible asociación entre infección por <i>Helicobacter pylori</i> y cáncer de estómago en una cohorte de 5 908 hombres japoneses-estadounidenses de Hawaii
Periodo de estudio	1967 - 1970
Banco de sueros	Alicuotas obtenidas en cada sujeto al inicio del estudio
Número de casos reportados en 1989	109 casos de cáncer gástrico diagnosticados histopatológicamente 20 años después
Exposición	Presencia de anticuerpos IgG a <i>Helicobacter pylori</i>
Selección de controles	Muestreo aleatorio pareado por edad
Resultados	94% de prevalencia en casos 76% de prevalencia en controles RM= 6.0; IC 95% 2.1 - 17.3
Conclusiones	Infección con <i>H. pylori</i> se asocia estrechamente a un incremento de riesgo de cáncer gástrico
* Fuente: referencia 19	

de la exposición, que constituyen la fase de alteración de riesgo en una población, y que es la diferencia entre el tiempo mínimo transcurrido antes del impacto o evento y el tiempo máximo

remanente. En esta variante de estudios de casos y controles, la longitud del periodo de efecto y su periodo de riesgo puede ser decidida empíricamente. Sin embargo, este periodo es crítico, porque la sobre o subestimación de la duración puede diluir la posible asociación. Un ejemplo de este tipo de diseño²² se presenta en el cuadro VII y figura 4. Para el análisis se utiliza el mismo método de un estudio de casos y controles pareado, pero en lugar de caso se utiliza el periodo de riesgo en com-

Cuadro VI
EJEMPLO DE CASOS Y CONTROLES ANIDADO EN
UNA COHORTE DINÁMICA

Objetivo	Evaluar la relación entre el consumo de estrógenos en la posmenopausia y la incidencia de infarto agudo del miocardio
Periodo de estudio	1978-1984
Población fuente	Mujeres entre 50 y 64 años de edad aseguradas por el Grupo de Cooperación en Salud (<i>Group Health Cooperative, GHC</i>) en Seattle, Washington. Este grupo provee a todos sus miembros cobertura total para cualquier evento en salud
Casos	Casos incidentes diagnosticados entre 1978 y 1984 120 mujeres con diagnóstico de infarto agudo del miocardio de primera vez, incluyendo aquellas fallecidas durante la hospitalización
Controles	Mujeres aseguradas durante el periodo de estudio; se seleccionaron al azar siete controles por caso entre las mujeres que estuvieron aseguradas entre 1978 y 1984 y que utilizaron la farmacia durante el mismo periodo que los casos
Exposición	El consumo de estrógenos como terapia de remplazo hormonal en la posmenopausia fue establecido en mujeres que lo utilizaron al menos 12 meses en forma continua desde la prescripción médica. La exposición se obtuvo de la lista de usuarias de la farmacia
Resultados	Consumo de estrógenos conjugados en los casos: 15% Consumo de estrógenos conjugados en los controles: 18% Incidencia de infarto agudo del miocardio en mujeres consumidoras de estrógenos conjugados: 6.8 Incidencia de infarto agudo del miocardio en mujeres no consumidoras de estrógenos conjugados: 9.7 Razón de momios para infarto agudo del miocardio entre consumidoras y no consumidoras de estrógenos conjugados fue: RM 0.7; IC 95% 0.4-1.3
Conclusiones	Los datos sugieren que la terapia de remplazo hormonal con estrógenos conjugados se asocia en forma inversa con el riesgo de sufrir infarto agudo del miocardio (estadísticamente no significativa) y es consistente con otros estudios
Fuente:	referencia 20

paración con el periodo control. Asimismo, los datos también se pueden analizar utilizando unidades tiempo-persona.

- *Estudios de mortalidad proporcional.* Los casos son definidos como aquellas muertes que ocurren de una fuente poblacional, mientras que los controles no son seleccionados directamente de personas vivas de esta fuente poblacional, sino que son obtenidos de otros tipos de muerte que ocurren en la población base. Rothman²³ establece que este tipo de controles son aceptables sólo si la distribución de la exposición entre los grupos es similar a la que presenta la base poblacional. Consecuentemente, la serie seleccionada como control se deberá restringir a las categorías de muerte que no están relacionadas con la exposición.

Sesgos

El hecho de que casos y controles se seleccionen utilizando diferentes esquemas y que la información se

Cuadro VII
EJEMPLO DE ESTUDIO DE CASO-CASO

Antecedente	La intuición y diversos estudios con limitaciones metodológicas sugieren que eventos de máximo estrés y "enojo" preceden inmediatamente al desarrollo de un infarto agudo del miocardio (IAM)
Material y métodos	Entrevista a 1 623 sujetos, en promedio cuatro días después de un infarto agudo del miocardio
Exposición	1. Características del evento clínico 2. Frecuencia de enojo y estrés durante el año previo 3. Intensidad del enojo, estrés y otros factores desencadenantes 26 horas antes del IAM 4. "Enojo" fue cuantificada con una escala de estrés (autorreporte de siete niveles) 5. Consumo de aspirina
Diseño	Caso-caso, mediante comparación de la ocurrencia de enojo intenso dos horas previas a la ocurrencia del IAM en relación con dos auto-controles ("self-matched") pareados
Resultados	La razón de momios de IAM dos horas después de un episodio de enojo fue: RM=2.3; IC 95% 1.7 - 3.2 Uso regular de aspirina tuvo un menor riesgo (RM=1.4; IC 95% 0.8-2.6) que los no usuarios (RM 2.9; IC 95%= 2.0 - 4.1; $p < 0.05$)
Conclusiones	Episodios de máximo estrés y enojo son capaces de desencadenar un evento de IAM, pero la aspirina puede reducir este riesgo
Fuente:	referencia 22

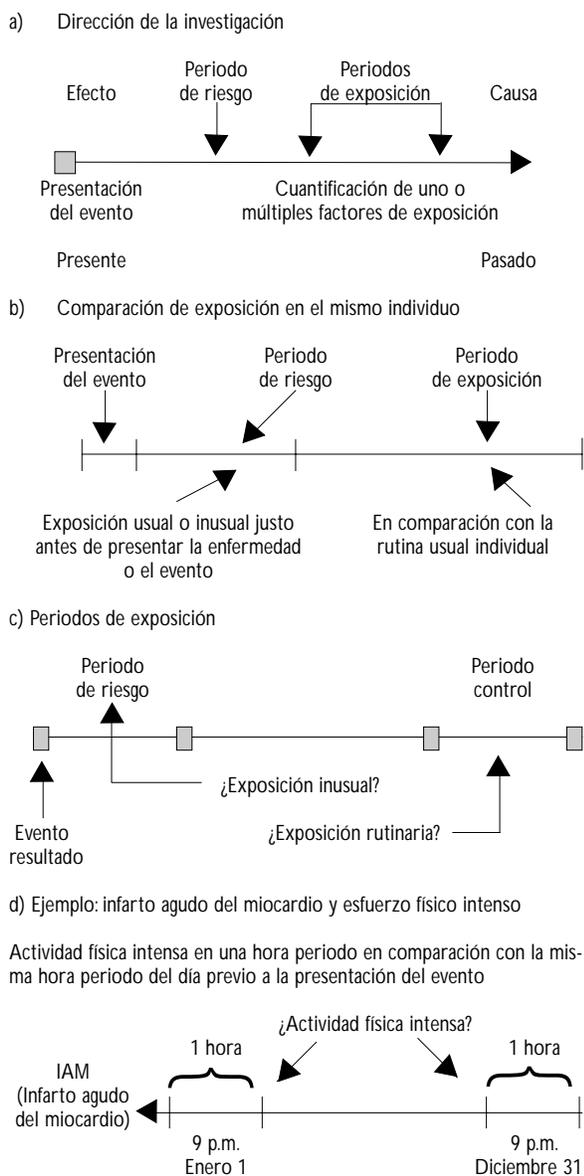


FIGURA 4. DISEÑO CASO-CASO ("CASE-CROSSOVER")

obtenga en la mayoría de las veces de forma retrospectiva, es decir, una vez que ocurrió el evento, hace que este tipo de estudios sean más vulnerables al efecto de diferentes sesgos. La característica que diferencia a los sesgos de los errores aleatorios es que los primeros se refieren a errores que ocurren de diferente manera entre los casos y los controles, lo que hace imposible distinguir entre diferencias reales que pudieran ser atribuidas a la exposición y diferencias espurias, atribuibles a errores metodológicos. En esta sección hare-

mos una breve revisión de los sesgos que se presentan con mayor frecuencia en los estudios de casos y controles.

Sesgos de selección. Puesto que en los estudios de casos y controles se selecciona a los participantes sobre la base de la ocurrencia del evento, este tipo de estudio epidemiológico es particularmente vulnerable a los sesgos de selección, por esta razón es recomendable trabajar con casos incidentes y evitar que la exposición o variables asociadas a ésta determinen o condicionen la participación en el estudio, ya sea como caso o como control. Como ejemplo de este tipo de sesgo se puede citar el estudio que reportó la asociación entre uso de estrógenos de remplazo y aumento en la frecuencia de cáncer endometrial. El sesgo de selección en el mencionado estudio se originó del hecho de que las mujeres que recibían estrógenos de remplazo también estaban sujetas a una vigilancia médica más estrecha, lo que producía que el diagnóstico de cáncer endometrial se realizara con mayor frecuencia en este grupo, en comparación con el grupo que no recibía estrógenos.²⁴ Como resultado de este sobrediagnóstico en el grupo que era usuario de estrógenos, se obtuvo una selección preferencial de los casos de cáncer endometrial expuestos a hormonales de remplazo, lo que condicionó una sobrestimación de la asociación real entre la exposición y la enfermedad.

Los estudios de casos y controles que se basan en poblaciones hospitalarias, pueden también estar sujetos a sesgos de selección con relativa frecuencia. Esto ocurre cuando se selecciona como control otro grupo de padecimientos. Si la exposición en estudio se encuentra asociada con el grupo de padecimientos seleccionado para el grupo control siempre existirá la posibilidad de incurrir en sesgos de selección. Por ejemplo, si analizáramos la relación entre el consumo de alcohol y el riesgo de enfermedad coronaria en un estudio, en el que los casos fueran sujetos que ingresan al servicio de urgencia por enfermedad coronaria, y los controles, sujetos que ingresan al mismo servicio, pero por trauma agudo, se podría presentar un sesgo de selección. Este podría ocurrir si el consumo de alcohol fuera un factor que estuviera relacionado con la ocurrencia de trauma agudo. Esta asociación condicionaría una sobrestimación de sujetos expuestos al alcohol en el grupo control, ocasionando una diferencia espuria entre casos y controles.

Otra fuente de sesgo de selección puede ser la *no respuesta* en alguno de los grupos. Cuando la tasa de *no respuesta* es diferente entre casos y controles y, además, ésta difiere entre expuestos y no expuestos,

se puede introducir un sesgo. En estudios donde los casos son obtenidos de fuentes hospitalarias, las condiciones de invitación y convencimiento son más favorables y se logran mejores tasas de respuesta de las que generalmente se logran para el grupo control; sin embargo, esta tasa diferencial de participación para casos y controles será una fuente de sesgo, siempre y cuando las tasas de participación para sujetos expuestos y no expuestos sean diferenciales. Este tipo de sesgo se puede introducir cuando el personal de campo conoce la hipótesis en estudio y realiza un esfuerzo mayor para lograr la participación de casos o controles con la exposición de interés.

Otro tipo de sesgo de selección es el que se puede introducir al estudiar casos prevalentes. Los casos prevalentes representan los sujetos con la enfermedad en estudio, que sobrevivieron hasta el momento en que se inicia la investigación y, en general, en este grupo hay alta proporción de sujetos que cursaron con la forma más benigna de la enfermedad. Si la exposición estuviese asociada no sólo con la ocurrencia de la enfermedad, sino también con la sobrevida, el uso de casos prevalentes podría conducir a conclusiones erróneas sobre la relación entre exposición y enfermedad.

Sesgos de información. Los estudios de casos y controles están propensos a sesgos que se introducen en el proceso de medición de la exposición, porque ésta se mide en forma retrospectiva, después del inicio de la enfermedad o del evento en estudio, esto es: a) la existencia del evento puede tener un efecto directo sobre la exposición (causalidad inversa); b) la existencia del resultado afecta la calidad de la medición, y c) la existencia del resultado afecta la determinación o registro de la exposición.

La información en los estudios de casos y controles frecuentemente es obtenida por medio de cuestionarios aplicados por entrevistadores, o mediante otra fuente como pueden ser los registros médicos; por lo tanto, es frecuente que las personas tengan problemas para recordar la información exacta sobre alguna exposición pasada. Sin embargo, es importante minimizar estas diferencias entre los casos y los controles. Por ejemplo, es probable que una mujer que tuvo un hijo con algún defecto congénito, y que es estudiada como caso, trate y haga un esfuerzo por recordar cualquier exposición durante el embarazo. En comparación, una mujer que tuvo un niño sano y que es estudiada como control, no tendrá la misma motivación para recordar, y es probable que consigne una información con un grado mayor de inexactitud. En este ejemplo, la respuesta es diferen-

cial entre ambos grupos, lo que podría introducir un sesgo de información conocido en la literatura como sesgo de *recordatorio*. En este ejemplo en particular, es probable que la asociación entre la exposición y el evento esté sobrestimada, debido al mayor grado de error que existe en la determinación de la exposición entre los controles.

Otro de los sesgos potenciales surge cuando los entrevistadores conocen la condición de caso y control, lo que puede conducir a que la entrevista se realice de manera diferencial entre los grupos. Por ejemplo, un entrevistador mal estandarizado podría inducir respuestas positivas sobre la exposición preferencialmente en el grupo de casos, aquí el sesgo de información es introducido por *el entrevistador*.

Otro de los sesgos de información puede ocurrir al clasificar a los individuos como expuesto o no expuesto, utilizando información sobre su condición de caso o control. Esto suele ocurrir durante la medición de la exposición basada en exámenes médicos, o en resultados de exámenes de laboratorio. La definición de exposición se debe sujetar a criterios estrictos y estandarizados, definidos *a priori* y las decisiones o revisiones siempre deben llevarse a cabo sin el conocimiento de la condición de caso o control. Esto último se podrá lograr siempre que el laboratorio que determina la exposición se mantenga ciego o enmascarado a la información sobre la condición del evento.

Análisis e interpretación

A diferencia de los estudios de cohorte, donde se puede calcular la razón de riesgos y la razón de tasas de incidencia, en los estudios de casos y controles no se puede estimar directamente la incidencia de la enfermedad en sujetos expuestos y no expuestos, debido primordialmente a que los individuos son seleccionados con base en la presencia o ausencia del evento de estudio y no por el *estatus* de exposición (una excepción son los estudios anidados y de caso-cohorte), donde podrá estimarse la incidencia, si se conocen las fracciones muestrales de exposición tanto en los casos como en los controles. Por esta razón, un buen estimador para medir la asociación entre una exposición y una enfermedad, es la razón de momios (también llamada por diversos autores: razón de ventajas, de productos cruzados, de suertes, de oportunidad relativa, entre otras).

A continuación se presentan consideraciones básicas del análisis de casos y controles, así como de sus diferentes modalidades.

Análisis sin pareamiento. Se basa en la tabla tradicional de 2 x 2: es indispensable recordar que se parte de que

el evento ya ocurrió y que se medirá el antecedente de exposición. Se puede entonces calcular los momios de exposición tanto en los casos como en los controles, esto es, comparar la posibilidad de ocurrencia de un evento con la posibilidad de que no ocurra bajo las mismas condiciones, y se expresa de la siguiente manera:

Momios de exposición en el grupo de los casos: a / b

Momios de exposición en el grupo de los controles: c / d

Comparando los momios de ocurrencia del evento de los expuestos y los no expuestos obtenemos la razón de momios:

$$RM = \frac{\text{momios de exposición en los casos}}{\text{momios de exposición en los controles}} = \frac{a/b}{c/d} = \frac{a \cdot d}{b \cdot c}$$

La interpretación de los resultados es la siguiente: si la RM es igual a uno, la exposición no está asociada con el evento o enfermedad; si la RM es menor de uno, la exposición está asociada de manera inversa con el evento, esto es, la exposición disminuye la posibilidad de desarrollar el evento; si la RM es mayor de uno, la exposición se encuentra asociada positivamente con el evento, lo que quiere decir que la exposición aumenta la posibilidad de desarrollar el evento.

Para cuantificar la precisión de la asociación se realiza el cálculo de los intervalos de confianza, normalmente estimados para un nivel de confianza del 95%, como se observa en el cuadro VIII; esto es, si se repitiera el mismo estudio n veces, bajo las mismas suposiciones estadísticas, en 95% de los casos el estimador puntual de la RM estará contenido dentro de los límites estimados.

Cuando la RM tiene valor por arriba del valor nulo (uno) y los intervalos de confianza no abarcan dicho valor, se puede calcular el impacto de la exposición mediante el riesgo atribuible (llamado también fracción etiológica, fracción atribuible). En otras palabras, la proporción de la enfermedad que se evitaría si se lograra erradicar la exposición. Consideremos como ejemplo un estudio de casos y controles realizado en la ciudad de México para evaluar la asociación entre la obesidad y el cáncer de endometrio;²⁵ 84 casos confirmados histopatológicamente fueron comparados con 626 controles, seleccionados aleatoriamente de la misma fuente de obtención de los casos. Para efecto de este ejercicio, la obesidad se definió como índice de masa corporal (IMC) mayor de 30 puntos (cuadro IX).

La obesidad se presentó en 35 mujeres en el grupo de los casos y en 166 en el grupo control. El resultado

Cuadro VIII
ANÁLISIS CLÁSICO DE UN ESTUDIO
DE CASOS Y CONTROLES NO PAREADO
PARA EVALUAR RAZÓN DE MOMIOS

	Exposición		Total
	Si	No	
Casos	a	b	n_1
Controles	c	d	n_0
Total	m_1	m_0	N

Casos: sujetos que desarrollaron el evento (enfermedad)
Controles: sujetos que no desarrollaron el evento

Resultado

Prevalencia de exposición en los casos:	a / n_1
Prevalencia de exposición en los controles:	c / n_0
Momios de exposición en los casos:	a / b
Momios de exposición en los controles:	c / d
Razón de momios (RM):	$a \cdot d / b \cdot c$
IC 95%:	$e^{\ln(RM) \pm 1.96 \cdot DE}$

Desviación estándar (DE):	$\sqrt{1/a+1/b+1/c+1/d}$
Riesgo atribuible poblacional (Rap):	$a/n_1(RM - 1) / RM$
Riesgo atribuible en los expuestos (Rae):	$RM - 1 / RM$

Categoría de referencia

- a: sujetos que desarrollaron el evento y estaban expuestos
- b: sujetos que desarrollaron el evento y que no estaban expuestos
- c: sujetos que no desarrollaron el evento y estaban expuestos
- d: sujetos que no desarrollaron el evento y no estaban expuestos
- m_1 : total de sujetos expuestos
- m_0 : total de sujetos no expuestos
- n_1 : total de casos
- n_0 : total de controles
- N: total de la población en estudio

ln: logaritmo natural

es una prevalencia de obesidad de 41% y 26%, respectivamente. Al evaluar la relación entre la obesidad y el cáncer endometrial se encuentra una asociación positiva: el riesgo de padecer cáncer endometrial es 1.98 veces mayor en las mujeres con IMC >30 (con obesidad), comparado con mujeres que tienen un IMC ≤30 (sin obesidad), la cual es significativa, ya que los IC 95% no abarcan el valor nulo, teniendo una variabilidad de esta asociación 1.24 veces mayor, o tan grande como 3.16 veces mayor.

Puesto que la asociación es positiva y estadísticamente significativa, se puede evaluar el impacto de la obesidad sobre el cáncer de endometrio. En este caso, el riesgo atribuible en la población es 0.21; en otras palabras, la obesidad en la población general, es responsable de 21% de los casos de cáncer de endometrio.

Análisis con pareamiento individual. Cuando el diseño de los estudios de casos y controles contempla un pa-

Cuadro IX
**CÁLCULO DE LA RAZÓN DE MOMIOS
 PARA EVALUAR EL RIESGO DE CÁNCER ENDOMETRIAL
 CON RELACIÓN A OBESIDAD**

	Índice de masa corporal		Total
	Expuestos >30	No expuestos ≤30	
Casos	35	49	84
Controles	166	460	626
Total	201	509	710

Hipótesis
 La obesidad (índice de masa corporal, IMC, >30) es un factor de riesgo para cáncer endometrial

Diseño
 Un grupo de mujeres con edades entre 22 y 79 años con cáncer endometrial se compara con un grupo de mujeres sin la enfermedad. Todas las mujeres fueron categorizadas con obesidad de acuerdo con el IMC: obesidad, IMC>30; sin obesidad, IMC≤30

Resultados
 Prevalencia de obesidad en los casos: 35 / 84 = 0.41
 Prevalencia de obesidad en los controles: 166 / 626 = 0.26
 Momios del grupo de los casos: 35 / 49 = 0.71
 Momios del grupo de los controles: 166 / 460 = 0.36
 Razón de momios (RM): 35*460 / 49*166= 1.98
 IC 95% para la razón de momios: $e^{\ln 1.98 \pm 1.96 * 0.2391} = 1.24 - 3.16$

Desviación estándar (DE): $\sqrt{1/35+1/49+1/166+1/460}=0.2391$

Riesgo atribuible poblacional (Rap): 35/84*1.98- 1 / 1.98= 0.21
 Riesgo atribuible en los expuestos (Rae): 1.98-1 / 1.98= 0.49

Fuente: referencia 25

reamiento individual por algún factor de confusión, el análisis tiene particularidades diferentes al de los estudios tradicionales de casos y controles. La tabla de 2 x 2 adquiere una connotación diferente. Dada esta condición, la razón de momios pareada (RMP) se puede calcular tomando en consideración las parejas con casos expuestos y controles no expuestos y dividirlos entre las parejas con casos no expuestos y controles expuestos (b/c), es decir, se utilizan únicamente las parejas discordantes en cuanto a la exposición. Distinga que aunque la notación en la tabla es la misma que los estudios no pareados (a, b, c, d), el contenido de cada celda difiere debido a que se estudian parejas. Este cálculo de la RMP considera solamente a pares discordantes y se explica por el hecho de que los pares en los que caso y control estuvieron expuestos, o en los que ambos no estuvieron expuestos, no contribuyen con información acerca de la posible asociación entre la exposición y la enfermedad (cuadro X).

Consideremos como ejemplo un estudio de casos y controles realizado en la ciudad de México, con 28

Cuadro X
**ANÁLISIS DE UN ESTUDIO DE CASOS
 Y CONTROLES PAREADO INDIVIDUALMENTE
 PARA EVALUAR RAZÓN DE MOMIOS**

	Controles		Total
	Expuestos	No expuestos	
Expuestos	a	b	a + b
No expuestos	c	d	c + d
Total	a + c	b + d	n = a + b + c + d

Casos
 Sujetos que desarrollaron el evento (enfermedad)
 Controles
 Sujetos que no desarrollaron el evento

Resultados
 Razón de momios pareada: RMP= b / c

IC 95%: $e^{\ln(b/c) \pm 1.96 \sqrt{(1/b + 1/c)}}$

Categoría de referencia
 a: parejas con caso expuesto y control expuesto
 b: parejas con caso expuesto y control no expuesto
 c: parejas con caso no expuesto y control expuesto
 d: parejas con caso no expuesto y control no expuesto
 a + c: total de parejas con controles expuestos
 b + d: total de parejas con controles no expuestos
 a + b: total de parejas con casos expuestos
 c + d: total de parejas con casos no expuestos
 n: total de parejas en el estudio

casos de cáncer de ovario no epitelial confirmado histopatológicamente, y 84 controles sin la enfermedad, pareados individualmente por edad, seleccionados aleatoriamente de la fuente de la cual fueron obtenidos los casos.²⁶ La exposición estudiada fue la paridad (por lo menos un parto a término), y como no expuestas aquellas mujeres que nunca habían tenido un parto a término.

Las parejas formadas por caso y control expuesto (concordantes), sumaron 30 (celda a); y las parejas formadas por caso y control no expuestos (concordantes), sumaron 24 (celda d); éstas se excluyen para el cálculo, y se utilizan exclusivamente las parejas discordantes por lo que se obtiene una RMP= 0.25. Esto quiere decir que la paridad se encuentra asociada de manera inversa al cáncer de ovario no epitelial, interpretándose como que las mujeres que tuvieron al menos un parto a término, tienen un riesgo cuatro veces menor de padecer cáncer ovárico no epitelial, al compararlas con aquellas mujeres que nunca tuvieron un parto a término (el valor resulta de invertir la RMP solamente para facilitar la interpretación: 1/0.25= 4; por lo tanto, la interpretación se realiza con base en el cambio en la exposición).

Los intervalos de confianza al no abarcar el valor nulo, hacen que la asociación sea significativa (RMp= 0.25; IC 95%= 0.10 – 0.61) (cuadro XI).

Análisis de mortalidad proporcional. Una variante de diseños de casos y controles son los estudios de mortalidad proporcional. En este caso se utiliza la información de registros o censos sobre causas de mortalidad por alguna enfermedad, por ejemplo, de cáncer; sin embargo, los mismos registros no permiten obtener información completa sobre una característica que puede ser la de interés para el estudio (denominadores) o puede faltar información sobre los datos de los casos (numeradores). En tal situación, principalmente cuando no se cuenta con información acerca de los denominadores, la opción más apropiada es realizar un estudio de mortalidad proporcional, donde el cálculo de la razón de momios de la mortalidad proporcional (RMmp) puede otorgar información

Cuadro XI
CÁLCULO DE LA RAZÓN DE MOMIOS PAREADA
PARA EVALUAR EL RIESGO DE CÁNCER DE OVARIO
NO EPITELIAL CON RELACIÓN A PARIDAD

	Controles		Total
	Paridad	Sin paridad	
Casos			
	Paridad	6	36
	Sin paridad	24	48
	Total	30	84

Hipótesis

La paridad es un factor de riesgo para el cáncer de ovario de origen no epitelial

Diseño

Un grupo de mujeres con cáncer de ovario no epitelial es comparado con un grupo de mujeres sin cáncer, se parearon individualmente por edad tres controles por cada caso. Todas las mujeres son categorizadas con *paridad*, cuando habían tenido por lo menos un parto a término, y *sin paridad* cuando nunca lo habían tenido

Resultados

Razón de momios pareada: (RMp): $6 / 24 = 0.25$
IC 95%: $e^{\ln 0.25 \pm 1.96 \sqrt{0.208}} = 0.10 - 0.61$

Categoría de referencia

a:	parejas con caso expuesto y control expuesto=	30
b:	parejas con caso expuesto y control no expuesto=	6
c:	parejas con caso no expuesto y control expuesto=	24
d:	parejas con caso no expuesto y control no expuesto=	24
a + c:	total de parejas con controles expuestos=	54
b + d:	total de parejas con controles no expuestos=	30
a + b:	total de parejas con casos expuestos=	36
c + d:	total de parejas con casos no expuestos=	48
n:	total de parejas en el estudio=	84

Fuente: referencia 26

valiosa de registros o estudios donde no hay denominadores.²⁷ Esta se puede calcular, por ejemplo, al estimar el riesgo de morir por un tipo de cáncer (casos) en relación con otros tipos de cáncer (controles), en un grupo expuesto de la población comparado con un grupo no expuesto de la misma población. Tal como se mostró previamente, la razón de momios sería en este caso un buen estimador de la razón de tasas de incidencia.

Conclusiones

En este artículo se han descrito las diversas alternativas y variantes de los diseños de casos y controles, cuya elección depende del tipo de información disponible y del conocimiento metodológico y analítico, así como de la imaginación e intuición del investigador para plantear un diseño de estudio y dar una respuesta válida a la pregunta de la investigación. A este respecto es necesario destacar que si las variables en estudio son dependientes del tiempo, los diseños de estudio anidados pueden resultar ideales. Por el contrario, si se requiere estudiar múltiples enfermedades con el mismo grupo de comparación, se puede utilizar un estudio de caso-cohorte. Adicionalmente, cuando se tiene la posibilidad de evaluar una exposición de riesgo inusual previa a un evento de estudio, se puede utilizar un estudio de caso-caso; o cuando se dispone de registros poblacionales de enfermedad podrían ser eficientes los estudios de mortalidad proporcional.

Finalmente, podemos mencionar que los estudios de casos y controles, igual que otros estudios observacionales, están sujetos a la acción de diferentes sesgos, por lo que no tienen como principal objetivo el de generalizar sus hallazgos, sino el de apoyar relaciones causa efecto, que tendrán que ser verificadas mediante estudios analíticos con un mayor poder en la escala de causalidad.

Referencias

1. Breslow NE, Day NE. Statistical methods in cancer research. The design and analysis of cohort studies. Vol. II. Lyon, Francia: International Agency for Research on Cancer, 1994.
2. Snow J. On the mode of communication of cholera. Churchill 1855. En: Snow on cholera, Commonwealth Fund 1936, reimpresión en Hafner Press, 1965.
3. Lane-Clayton JE. A further report on cancer of the breast. Reports on Public Health and Medical Subjects 32. Londres: Her Majesty's Stationery Office, 1926.

4. Cornfield J. A method of estimating comparative rates from clinical data. Application to cancer of the lung, breast and cervix. *J Natl Cancer Inst* 1951; 11:1269-1275.
5. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959;22(4):719-748.
6. Miettinen OS. Estimability and estimation in case-referent studies. *Am J Epidemiol* 1976;104:609-620.
7. Doll R, Hill AB. A study of the aetiology of carcinoma of the lung. *BMJ* 1952;2:1271-1286.
8. Herbst AL, Ulfelder H, Poskanzer DC. Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. *N Engl J Med* 1971;284:878-881.
9. Ahlbom A, Staffan N. Fundamentos de epidemiología. 4ª edición. México, D.F.: Siglo XXI editores, 1993.
10. Muñoz N, Bosch FX, de Sanjose S, Tafur L, Izruga I, Gili M *et al*. The causal link between human papillomavirus and invasive cervical cancer: A population-based case-control study in Colombia and Spain. *Int J Cancer* 1992;52:743-749.
11. Narod SA, Brunet JS, Gharidian P, Robson M, Heimdal K, Neuhausen SL *et al*. Tamoxifen and risk of contralateral breast cancer in BRCA1 and BRCA2 mutation carriers: A case-control study. *Hereditary breast cancer clinical study group. Lancet* 2000;356:1876-1881.
12. Weber JT, Hibbs RG Jr, Darwish A, Mishu B, Corwin AL, Rakha M *et al*. A massive outbreak of type E botulism associated with traditional salted fish in Cairo. *J Infect Dis* 1993;167(2):451-454.
13. Magnani C, Agudo A, Gonzalez CA, Andron A, Calleja A, Chellini E *et al*. Multicentric study on malignant pleural mesothelioma and non-occupational exposure to asbestos. *Br J Cancer* 2000;83(1):104-111.
14. Romieu I, Hernandez AM, Lazcano PE, Lopez L, Romero JR. Breast cancer and lactation history in Mexican women. *Am J Epidemiol* 1996; 143:543-552.
15. Perez-Padilla R, Regalado J, Vedal S, Pare P, Chapela R, Sansores R *et al*. Exposure to biomass smoke and chronic airway disease in Mexican women. A case-control study. *Am J Respir Care Med* 1996;154:701-706.
16. Hernández M, Lazcano E, Alonso P, Romieu I. Evaluation of the cervical cancer screening programme in Mexico: A population based case control study. *Int J Epidemiol* 1998; 27: 370-376.
17. Gordis L. Epidemiology. Philadelphia: W.B. Sanders Company, 1996.
18. Schuurman AG, Goldbohm RA, Dorant E, van den Brandt P. Anthropometry in relation to prostate cancer risk in the Netherlands Cohort Study. *Am J Epidemiol* 2000;151(6):541-549.
19. Nomura A, Stemmermann GN, Chyou PH, Kato I, Perez-Perez GI, Blaser MJ. *Helicobacter pylori* infection and gastric carcinoma among Japanese Americans in Hawaii. *N Engl J Med* 1991;325(16):1132-1136.
20. Hernández AM, Walker AM, Jick H. Use of replacement and the risk of myocardial infarction. *Epidemiology* 1990;1:128-133.
21. Maclure M. The case-crossover design: A method for studying transient effects on the risk of acute events. *Am J Epidemiol* 1991;133:144-153.
22. Mittleman MA, Maclure M, Sherwood JB, Mulry RP, Tofler GH, Jacobs SC *et al*. Triggering of acute myocardial infarction onset by episodes of anger. Determinants of myocardial infarction onset study investigators. *Circulation* 1995;92(7):1720-1725.
23. Rothman KJ, Greenland S. Modern epidemiology. 2a. edición. Filadelfia: Lippincott-Raven Publishers, 1998:93-114.
24. Horwitz RI, Feinstein AR. Alternative analytic methods for case-controls studies of estrogens and endometrial cancer. *N Engl J Med* 1978;299:1089-1094.
25. Salazar ME, Lazcano PE, Gonzalez LG, Escudero RP, Salmeron CJ, Hernández AM. Case-control study of diabetes, obesity, physical activity and risk of endometrial cancer among Mexican women. *Cancer Causes Control* 2000;11:707-711.
26. Sánchez ZL, Salazar ME, Lazcano PE, González LG, Escudero RP, Hernández AM. Factores de riesgo asociados a cáncer de ovario no epitelial en una muestra de mujeres mexicanas (tesis de maestría). Cuernavaca, México: Instituto Nacional de Salud Pública, 2001.
27. Miettinen OS, Wang JD. An alternative to the proportionate mortality ratio. *Am J Epidemiol* 1981;114:144-148.

Estudios de cohorte.

Metodología, sesgos y aplicación

Eduardo Lazcano-Ponce, Dr. en C.,⁽¹⁾ Esteve Fernández, Dr. en C.,⁽²⁾
Eduardo Salazar-Martínez, Dr. en C.,⁽¹⁾ Mauricio Hernández-Avila, Dr. en C.⁽¹⁾

cohorte

Del latín cohors, cohortis: séquito, agrupación. Entre los romanos, cuerpo de infantería que comúnmente constaba de 500 hombres, y era la décima parte de una legión.

Por lo general, los veteranos ocupaban la primera y última fila de la cohorte.

Puede provenir del verbo latino cohortari, arengar, toda vez que la fuerza de la cohorte se ajustó generalmente al número de hombres que podían escuchar juntos la voz del jefe que les dirigía la palabra

La característica que define a los estudios de cohorte es que los sujetos de estudio se eligen de acuerdo con la exposición de interés; en su concepción más simple se selecciona a un grupo expuesto y a un grupo no-expuesto y ambos se siguen en el tiempo para comparar la ocurrencia de algún evento de interés. Los estudios de cohorte prospectivos son semejantes a los ensayos clínicos aleatorizados en el sentido de que los sujetos de estudio se siguen en el curso de la exposición hasta la aparición del evento que interesa, pero a diferencia del ensayo clínico aleatorizado, donde el investigador asigna la exposición, en los estudios de cohorte el investigador observa a los sujetos después de ocurrida la exposición.

Anteriormente, los estudios de cohorte eran referidos como longitudinales, porque los sujetos eran seguidos con al menos dos mediciones a través del tiempo; también eran conocidos como prospectivos, que implican la dirección hacia el futuro en la cual eran seguidos los sujetos; o de incidencia, porque se relacionan con la medida básica de los nuevos casos del evento de estudio a lo largo del tiempo. Actualmente, sin em-

bargo, el término correcto para referirse a este diseño es de estudios de cohorte.

Los estudios de cohorte se han utilizado de manera clásica para determinar la ocurrencia de un evento específico en un grupo de individuos inicialmente libres del evento o enfermedad en estudio. Es decir, toda la población en estudio se sigue a través del tiempo y se compara la incidencia del evento de estudio en individuos expuestos con la de los no-expuestos. El seguimiento de la población en estudio se continúa hasta que ocurre una de las siguientes condiciones: a) se manifiesta el evento de estudio (en razón de salud o enfermedad), cuando ocurre esta condición el individuo deja de contribuir a la cohorte pero puede reingresar si se trata de un evento recurrente o que puede ocurrir varias veces; b) los sujetos de estudio mueren; c) los sujetos se pierden durante el seguimiento, o d) el estudio termina.

Clasificación de los estudios de cohorte

Dependiendo de la relación temporal del inicio del estudio respecto a la ocurrencia del evento, los estu-

(1) Centro de Investigación en Salud Poblacional, Instituto Nacional de Salud Pública. Cuernavaca, Morelos, México.

(2) Departamento de Salud Pública, Universidad de Barcelona. Barcelona, España.

dios de cohorte se han clasificado como: prospectivos, y retrospectivos (o históricos). Los estudios de cohorte histórica reconstruyen la experiencia de la cohorte en el tiempo, por esta razón dependen de la disponibilidad de registros para establecer exposición y resultado. Una aplicación frecuente de una cohorte histórica son los estudios de exposición ocupacional, donde se reconstruye la exposición y frecuencia de eventos en un grupo ocupacional o empresa durante un periodo determinado. La validez del estudio dependerá en gran medida de la calidad de los registros utilizados. En contraste, en las cohortes prospectivas es el investigador quien documenta la ocurrencia del evento en tiempo venidero en la población en estudio, por lo que la exposición y resultado se valoran de manera concurrente, y la calidad de las mediciones puede ser controlada por los investigadores.

Con relación al tipo de población, las cohortes construidas pueden ser fijas, también llamadas cerradas, y dinámicas (figura 1). Cerradas o fijas son las cohortes que por diseño de estudio no consideran la inclusión de población en estudio mas allá del periodo de reclutamiento fijado por los investigadores, por ejemplo: estudiantes de la Escuela de Salud Pública de México, generación 1945.

Dinámicas son aquellas cohortes que consideran la entrada y salida de nuevos sujetos de estudio durante la fase de seguimiento, por lo que el número de miembros puede variar a través del tiempo. Los participantes entran o salen de la cohorte cuando cumplen criterios de elegibilidad, incorporando la aportación años-persona desde el momento de inclusión en el estudio; frecuentemente están definidos por unidades geográficas y grupos poblacionales, por ejemplo: población asegurada por el Instituto Mexicano del Seguro Social, durante el periodo de 1995 a 2000.

Diseño de estudio

En un estudio de cohorte se recluta a un grupo de individuos, ninguno de los cuales manifiesta en ese momento el evento de estudio, pero todos los integrantes están en riesgo de padecer o presentar dicho evento. Para su inclusión en el estudio los individuos de la cohorte son clasificados de acuerdo con las características (factores de riesgo) que podrían guardar relación con el resultado. Posteriormente, estos individuos son observados a lo largo del tiempo para cuantificar cuáles de ellos manifiestan el resultado.

Los dos grupos de comparación (expuestos y no expuestos) pueden ser seleccionados de poblaciones diferentes; sin embargo, la inferencia causal (la validez del estudio) depende del supuesto de que ambos grupos son comparables respecto a otros factores asociados con la exposición o evento de interés. La principal ventaja metodológica de un diseño de cohorte es que los niveles del factor de estudio son observados a través del periodo de seguimiento antes de que la enfermedad o evento de interés sea detectado. Consecuentemente, el investigador puede razonablemente postular la hipótesis de que la causa precede la ocurrencia de la enfermedad y que el estatus de enfermedad no influyó diferencialmente en la selección de sujetos o en la determinación de la exposición. Los estudios de cohorte tienen ciertas ventajas y desventajas respecto a otro tipo de estudios epidemiológicos (cuadro I), pero en general son menos susceptibles de sesgos de selección como se describe más adelante.

Como puede observarse en la figura 2, en un estudio de cohorte la información acerca del factor de exposición o de estudio es conocida para todos los sujetos al inicio del periodo de seguimiento. La población en riesgo de desarrollar el evento es seguida por un perio-

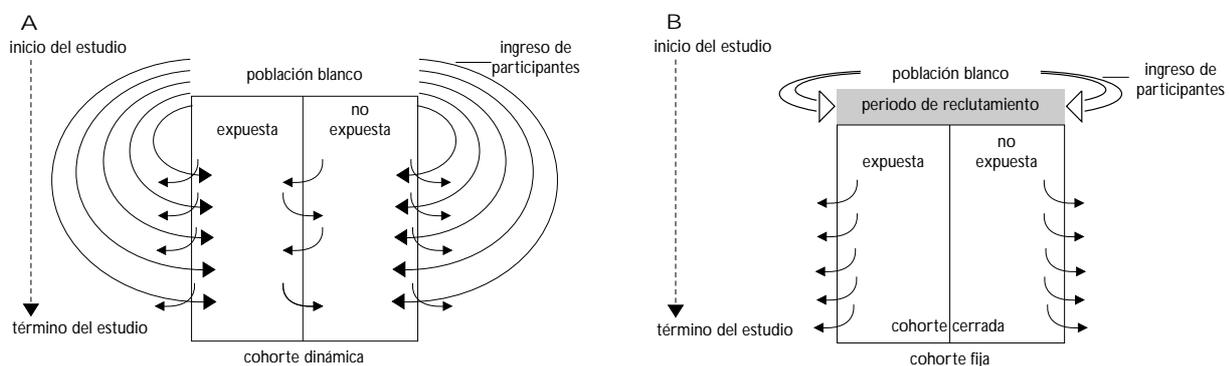


FIGURA 1. CLASIFICACIÓN DE ESTUDIOS DE COHORTE

Cuadro I

VENTAJAS Y DESVENTAJAS DE LOS ESTUDIOS DE COHORTE

Ventajas

- Es el único método para establecer directamente la incidencia
- La exposición puede determinarse sin el sesgo que se produciría si ya se conociera el resultado; es decir, existe una clara secuencia temporal de exposición y enfermedad
- Brindan la oportunidad para estudiar exposiciones poco frecuentes
- Permiten evaluar resultados múltiples (riesgos y beneficios) que podrían estar relacionados con una exposición
- La incidencia de la enfermedad puede determinarse para los grupos de expuestos y no-expuestos
- No es necesario dejar de tratar a un grupo, como sucede con el ensayo clínico aleatorizado

Desventajas

- Pueden ser muy costosos y requerir mucho tiempo, particularmente cuando se realizan de manera prospectiva
- El seguimiento puede ser difícil y las pérdidas durante ese periodo pueden influir sobre los resultados del estudio
- Los cambios de la exposición en el tiempo y los criterios de diagnóstico pueden afectar a la clasificación de los individuos
- Las pérdidas en el seguimiento pueden introducir sesgos de selección
- Se puede introducir sesgos de información, si la identificación de la enfermedad puede estar influenciada por el conocimiento del estado de exposición del sujeto
- No son útiles para enfermedades poco frecuentes porque se necesitaría un gran número de sujetos
- Durante mucho tiempo no se dispone de resultados
- Evalúan la relación entre evento del estudio y la exposición a sólo un número relativamente pequeño de factores cuantificados al inicio del estudio

do de tiempo determinado mediante nuevos exámenes o cuantificando su periodo de supervivencia, hasta que el evento de estudio o muerte sean identificadas.

Selección de la cohorte

Antes de que pueda identificarse a las personas expuestas es necesario definir explícitamente los niveles y duración mínima de la exposición. Adicionalmente, en el estudio pueden fijarse criterios de elegibilidad, pero lo más importante es que los sujetos estén libres al menos momentáneamente del evento de estudio.

En estudios de cohorte se puede contar con más de un grupo de exposición. La decisión de incluir o excluir a determinados sujetos de la población de estudio dependerá de la exposición y del resultado que interesa, así como de la medida en que se influye sobre la heterogeneidad al restringir la admisión a ciertos grupos.

La exposición y las covariables en estudio pueden ser dependientes del tiempo, por lo que es necesario considerar los factores por los cuales el nivel de exposición varía con el tiempo, así como variación de tasas y la posibilidad de que los confusores y modificadores de efecto varíen en el tiempo. En las exposiciones fijas los factores no cambian a través del tiempo, tal es el caso del sexo y el lugar de nacimiento, entre otros.

Existen otras múltiples aplicaciones en que pueden utilizarse los estudios de cohorte, como puede observarse en el cuadro II, de acuerdo con su dimensión tiempo. Entre ellos la edad que es un determinante de incidencia; el tiempo calendario, donde los antecedentes y exposición pueden variar con el tiempo; el seguimiento propiamente dicho que puede ser un índice de exposición. Asimismo, la duración de la expo-

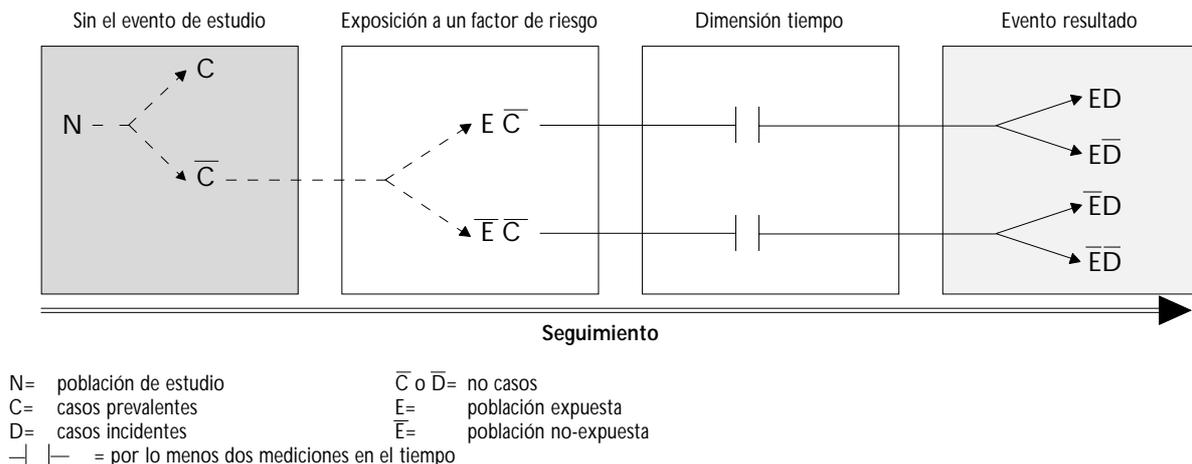


FIGURA 2. DISEÑO CLÁSICO DE UN ESTUDIO DE COHORTE

sición refleja un índice de exposición acumulada. El tiempo transcurrido desde la exposición útil al evento es también referido como latencia y, finalmente, el tiempo desde el final de la exposición puede reflejar declinación en el riesgo.

Los sujetos no-expuestos han de ser similares a los expuestos en todos sus aspectos excepto en que no han estado sometidos a la exposición que se estudia. La captación de los sujetos no-expuestos debe ser la misma que se aplicó a los sujetos expuestos. Deben tener, además, el mismo riesgo potencial de presentar el evento de estudio, y tener las mismas oportunidades que los expuestos de ser diagnosticados del evento resultado en estudio.

Las opciones utilizadas para la conformación de las cohortes varían según la exposición que es objeto de estudio. Así, por ejemplo, el estudio de la dieta o estilos de vida se ha estudiado en cohortes conformadas con muestras de la población general, mientras que exposiciones poco frecuentes se han estudiado en cohortes ocupacionales.

Medición de evento resultado y seguimiento

Los eventos de estudio pueden ser: a) evento simple (fijo en el tiempo) o evento raro (muerte o incidencia

de enfermedad). En ambos casos, al observar el evento en cada unidad de análisis el seguimiento termina; b) eventos múltiples o raros (enfermedades recurrentes, sintomatología o eventos fisiológicos). Al presentar el evento el individuo deja de estar en riesgo por lo que ya no cumple con el criterio de permanencia en la cohorte. Se puede reiniciar el seguimiento cuando se restablece el riesgo, es decir, cuando hay curación y el individuo vuelve a estar en riesgo de presentar el evento; c) modificación de medida eje (función broncopulmonar en el tiempo, modificación de la función pulmonar hacia un aumento o disminución; crecimiento), que son evaluados mediante tasa de cambio, y d) marcadores intermedios del evento (cuenta de apolipoproteínas A y B como marcadores de predisposición a enfermedad cardiovascular).

El periodo de seguimiento puede abarcar años, meses, semanas o días, dependiendo de la frecuencia del evento estudiado. Dos momentos definen el periodo de seguimiento: el examen inicial (medición basal) y el final del seguimiento. El inicio del seguimiento depende del tipo de cohorte: si es cerrada o dinámica, ya que en el caso de esta última, el inicio del seguimiento se define para cada participante a través de un largo periodo de tiempo. El seguimiento, dependiendo del evento de interés, puede ser activo o pasivo. Activo es aquel en el que se utilizan contactos repetidos por diversos medios; nueva entrevista y obtención de muestras, cuestionarios autoaplicables o llamadas telefónicas. El seguimiento pasivo es el que se realiza mediante búsqueda sistemática de sistemas de información en registros preestablecidos (registros de cáncer, hospitalarios, registro civil, entre otros).

Pérdidas en el seguimiento

Las pérdidas en el seguimiento pueden originarse principalmente por tres razones: a) abandono del estudio, b) muerte por otra causa al evento de interés, y c) pérdidas, llamadas "administrativas", originadas por la terminación temprana del estudio por razones ajenas a las que se plantearon originalmente (ejemplo: agotamiento de la fuente de financiamiento).

Cuantificar las causas que producen pérdidas en el seguimiento es importante para evaluar la validez del estudio.

Análisis estadístico

La base del análisis de un estudio de cohorte es la evaluación de la ocurrencia de un evento (en términos de salud o enfermedad) resultante del seguimiento en el tiempo, como consecuencia de haber estado expuesto

Cuadro II
OBJETIVOS DE LOS ESTUDIOS DE COHORTE

Características en común	Evaluar el efecto de:	Ejemplo
Edad	Edad	Esperanza de vida de los individuos de 70 años (con independencia de cuándo nacieron)
Fecha de nacimiento	Cohorte	Tasa de cáncer cervical para mujeres nacidas en 1910
Exposición	Factor de riesgo	Cáncer de pulmón en individuos que fuman
Enfermedad	Pronóstico	Tasa de supervivencia de mujeres con cáncer cervical
Intervención preventiva	Prevención	Disminución de la incidencia de cáncer de hígado después de la vacunación vs. hepatitis B
Intervención terapéutica	Tratamiento	Supervivencia similar para mujeres con cáncer de ovario epitelial unilateral, a las que se les practicó cirugía conservadora para mantener su fertilidad

o no (grupos de comparación) a una determinada exposición (factor de riesgo). Esto es, el investigador selecciona un grupo de sujetos expuestos y otro grupo de sujetos no-expuestos y los sigue en el tiempo para comparar la incidencia de algún evento (incidencia de la enfermedad, o según sea el caso, tasa de muerte de la enfermedad). Es indispensable considerar que para poder analizar adecuadamente un estudio de cohorte se necesitará información sobre la fecha de inicio, fecha en que ocurran los eventos y de terminación del estudio, así como la información completa de los datos de los sujetos participantes, la escala de medición y el motivo de terminación del seguimiento (pérdida, muerte u ocurrencia del evento en estudio).

Cuando existe una asociación positiva entre la exposición y el evento se esperaría que la proporción del grupo expuesto que desarrolló la enfermedad sea mayor que la proporción del grupo no-expuesto que también presentó el evento (incidencia del grupo expuesto vs. incidencia del grupo no-expuesto). Partiendo de un grupo expuesto donde "a" sujetos desarrollan el evento y "c" sujetos no desarrollan el evento, tenemos entonces, que la incidencia de la enfermedad entre los expuestos es: $a / a + c$. De la misma manera, en el grupo de sujetos no-expuestos, "b" y "d", el evento ocurre en "b" sujetos, pero no en "d" sujetos, tenemos entonces, que la incidencia de la enfermedad entre los no-expuestos es: $b / b + d$ (cuadro III).

Para calcular la razón de incidencia acumulada (RIA) se estima la incidencia del grupo expuesto entre la incidencia del grupo no-expuesto:

$$RIA = \frac{a / a + c}{b / b + d}$$

La RIA es una medida de asociación entre el evento y la exposición. Consideremos un ejemplo con datos hipotéticos de un estudio de cohorte donde se investiga la asociación entre el estado nutricional y el riesgo de muerte en pacientes con diagnóstico de leucemia, seleccionando un grupo de 17 sujetos con bajo estado nutricional (expuestos) y otro grupo de 15 sujetos con estado nutricional normal (no-expuestos), quienes se encontraban libres de enfermedad al inicio del estudio; ambos grupos fueron seguidos hasta que se presentó el evento. El evento (muerte) se registró en 14 sujetos del grupo con déficit nutricional y en ocho en el grupo sin déficit. El resultado es una incidencia de 0.82 en el grupo de bajo estado nutricional y de 0.53 en los del grupo de estado nutricional normal. Al estimar el efecto obtenemos un riesgo relativo positivo; por lo tanto, los sujetos con bajo estado nutricional tienen 1.54 veces mayor riesgo de presentar el evento

Cuadro III
ANÁLISIS DE UN ESTUDIO DE COHORTE PARA EVALUAR RAZÓN DE RIESGOS

Evento	Exposición	
	Sí	No
	a	b
	c	d
Total	m_1	m_0

Incidencia en el grupo de expuestos (m_1) = $a / (a + c)$
 Incidencia en el grupo de no-expuestos (m_0) = $b / (b + d)$
 Razón de incidencia acumulada = m_1 / m_0
 Diferencia de incidencia acumulada* = $m_1 - m_0$

* Si la exposición es protectora, la diferencia de riesgos debe calcularse como $m_0 - m_1$

donde: a= Sujetos con la exposición que desarrollaron el evento
 b= Sujetos sin la exposición que desarrollaron el evento
 c= Sujetos con la exposición que no desarrollaron el evento
 d= Sujetos sin la exposición que no desarrollaron el evento
 m_1 = Total de sujetos expuestos
 m_0 = Total de sujetos no-expuestos

al ser comparados con los sujetos con estado nutricional normal (cuadro IV).

Razón de tasas de incidencia

Cuando el estudio presenta pérdidas en el seguimiento (abandono, cambio de domicilio, muerte por causa diferente al evento del estudio, finalización del estudio por llegada al tiempo establecido por el protocolo o por agotamiento de recursos financieros) los tiempos de seguimiento son desiguales. Una forma de tratar periodos de seguimiento variables es con el análisis basado en tiempo-persona. En estos casos, se puede utilizar el promedio de tiempo contribuido por la totalidad de sujetos de la cohorte; es decir, se agregan seis meses por el año de entrada, seis meses más por el año de salida y doce meses por todos los años de entrada y salida.¹

Cuando se ha cuantificado el tiempo-persona de seguimiento para cada sujeto, el denominador cambia a una dimensión de tiempo (las unidades son por ejemplo, años-persona, días-persona, horas-persona). Esto nos permite estimar la tasa de los casos incidentes en una unidad de tiempo determinada.²

Consideremos una cohorte donde se conoce el tiempo que cada individuo ha permanecido en el seguimiento, se puede calcular la tasa de incidencia de acuerdo con el estado de exposición de cada sujeto. Así, partiendo de un grupo de sujetos "a" que presentan el evento y la exposición, y el tiempo-persona de

Cuadro IV
CÁLCULO DE LA RAZÓN DE RIESGOS PARA SUJETOS CON
BAJO ESTADO NUTRICIONAL CON RELACIÓN A LEUCEMIA

		Estado nutricional		Total
		Bajo	Normal	
Leucemia	Sí	14	8	22
	No	3	7	10
	Total	17	15	32

Incidencia en el grupo de expuestos (m_1) = $a / (a + c) = 0.82$
 Incidencia en el grupo de no-expuestos (m_0) = $b / (b + d) = 0.53$
 Razón de incidencia acumulada = $m_1 / m_0 = 1.54$
 Diferencia de incidencia acumulada = $m_1 - m_0 = 0.29$
 IC 95% para la razón de incidencia acumulada* = $0.91 - 2.6$

* IC 95% = $e^{\ln RIA \pm 1.96 \cdot \sqrt{(c / am_1 + d / bm_0)}}$

seguimiento "tp_e" de estos sujetos expuestos, se puede calcular la tasa de incidencia para los expuestos: $TI_1 = a / tp_e$. De la misma manera, consideremos un grupo "b" de sujetos con el evento, pero sin la exposición y el tiempo-persona de seguimiento de estos sujetos "tp_{ne}", en los cuales se puede calcular la tasa de incidencia para los sujetos no-expuestos: $TI_0 = b / tp_{ne}$ (cuadro V).

El cálculo de la razón de tasas de incidencia (RTI) se deriva de la siguiente manera:

$$RTI = TI_1 / TI_0$$

El producto de la estimación es una medida de asociación que nos permite evaluar la diferencia entre los grupos expuesto y no-expuesto.

Tomando el mismo ejemplo hipotético de estado nutricional y leucemia, en el cual 14 sujetos con bajo estado nutricional (contribución de 571 días-persona de seguimiento) y ocho sujetos del grupo con estado nutricional normal (contribución de 1 772 días-persona de seguimiento) desarrollaron el evento (muerte). La tasa de mortalidad para los sujetos con bajo estado nutricional fue de 24.5/1 000 días-persona y para los sujetos con estado nutricional normal la tasa de mortalidad fue de 4.5/1 000 días-persona. Por lo tanto, la velocidad de ocurrencia del evento es 5.4 veces más alta en el grupo de sujetos con bajo estado nutricional que en el grupo de sujetos con estado nutricional normal (cuadro VI). Como se puede observar, los resultados de la estimación de la razón de riesgos y la razón de tasas de incidencia es diferente, esto es debido, probablemente, a que el evento es frecuente (14 de 17 su-

Cuadro V
ANÁLISIS DE UN ESTUDIO DE COHORTE PARA ESTIMAR
RAZÓN DE TASAS DE INCIDENCIA

	Expuestos	No expuestos
Casos	a	b
Tiempo-persona	tp _e	tp _{ne}
Tasas	TI ₁	TI ₀

Tasa de incidencia en el grupo de expuestos (TI_1) = a / tp_e
 Tasa de incidencia en el grupo de no-expuestos (TI_0) = b / tp_{ne}
 Razón de tasas = TI_1 / TI_0
 Diferencia de tasas* = $TI_1 - TI_0$

* Si la exposición es protectora, las diferencias de tasas deben calcularse como $TI_0 - TI_1$

donde: a = Sujetos con la exposición que desarrollaron el evento
 b = Sujetos sin la exposición que desarrollaron el evento
 tp_e = Tiempo-persona de seguimiento de los sujetos expuestos que desarrollaron el evento
 tp_{ne} = Tiempo-persona de seguimiento de los sujetos no-expuestos que desarrollaron el evento
 TI₁ = Tasa de incidencia de los sujetos expuestos
 TI₀ = Tasa de incidencia de los sujetos no-expuestos

jetos lo presentaron) y el periodo de seguimiento es prolongado.

Este método permite realizar análisis cuando existe un cambio en el estado de exposición, de tal manera que un mismo sujeto puede contribuir en el denominador de los expuestos en un periodo y entre los no-expuestos en otro momento.

Finalmente, aunque hemos descrito de manera simple la forma como se analizan los datos provenientes de una cohorte, cuando queremos conocer el efecto de la variable estudiada controlando variables potencialmente confusoras, requerimos de un análisis múltiple ajustando simultáneamente diferentes variables mediante la regresión de Poisson.¹

Otra estrategia para el análisis de un estudio de cohorte es el análisis de supervivencia.³ Esta estrategia permite el análisis de eventos frecuentes en poblaciones pequeñas, a diferencia del análisis de tiempo-persona en el cual el evento es generalmente poco frecuente y se realiza en poblaciones más grandes. El cuadro VII presenta las principales diferencias entre las dos estrategias de análisis.⁴

Sesgo y validez en los estudios de cohorte

Aunque se reconoce que los estudios de cohorte representan un diseño menos sujeto a error sistemático o sesgo en comparación con otros estudios observa-

Cuadro VI
CÁLCULO DE LA RAZÓN DE TASAS DE INCIDENCIA PARA SUJETOS CON BAJO ESTADO NUTRICIONAL CON RELACIÓN A LEUCEMIA

	Estado nutricional	
	Bajo	Normal
Leucemia	14	8
Tiempo-persona*	571	1772
Tasas	0.0245	0.0045

Tasa de incidencia en el grupo de expuestos (TI_1) = $a / tp_e = 0.0245$
 Tasa de incidencia en el grupo de no-expuestos (TI_0) = $b / tp_{ne} = 0.0045$
 Razón de tasas = $TI_1 / TI_0 = 5.4$
 Diferencia de tasas = $TI_1 - TI_0 = 0.2$
 IC 95% para la razón de tasas[†] = 2.12 - 14.9

* días-persona

† IC 95% = $e^{-lnRT \pm 1.96 \cdot \sqrt{(1/a + 1/b)}}$

cionales,⁵ no es menos cierto que se deben tener en consideración algunas fuentes que pueden distorsionar los resultados que se deriven de ellos (figuras 3, 4 y 5). Existen, en efecto, sesgos de selección e información en los estudios de cohorte que deben ser rigurosamente considerados, sobre todo por lo que se refiere a pérdidas en el seguimiento (de los pacientes, de los participantes, etc.), al modo en que se obtiene la información sobre la exposición estudiada y al modo en que se determina en la población en estudio la ocurrencia de la enfermedad o condición de interés durante el seguimiento. Por lo que se refiere a sesgos de confusión, en los estudios de cohorte es importante considerar factores que se asocian independientemente tanto con la exposición como con la condición o evento estudiado, que no sean pasos intermedios en el proceso causal, ya que éstos pueden hacer aparecer una asociación ficticia entre la exposición y el factor en estudio.⁶ Dado que el procedimiento para la identificación y el control de la confusión (mediante el análisis estratificado y modelos multivariados, principalmente) es conceptualmente similar al usado en los estudios de casos y controles, vamos a centrar el resto de la discusión en las principales fuentes de sesgos.

Clásicamente se clasifican los sesgos en los estudios epidemiológicos como sesgos de selección (cuando los errores derivan de cómo se constituye la población en estudio) y sesgos de información (cuando los errores se originan durante el proceso de recolección de la información). Una fuente de error con entidad propia, aunque podría considerarse un tipo especial de sesgo de información, es la llamada mala-clasificación o error

Cuadro VII
COMPARACIÓN DE LAS PRINCIPALES ESTRATEGIAS PARA EL ANÁLISIS DE LOS ESTUDIOS DE COHORTE*

	Análisis de supervivencia	Análisis basado en tiempo-persona
Tamaño de muestra	Relativamente pequeño	Relativamente grande
Número de eventos	Frecuentes	Raros
Escala temporal	Única	Única o múltiple
Tipos de medida de incidencia	Probabilidad (condicional y acumulada)	Tasa (densidad)
Análisis unifactorial	<ul style="list-style-type: none"> • Comparación de curvas de supervivencia • Prueba de log-rank • Razón de riesgos 	<ul style="list-style-type: none"> • Comparación de tasas • Razón de tasas (densidades) • Razón estandarizada de mortalidad (REM)
Análisis multifactorial	Regresión de Cox	Regresión de Poisson

* Modificado de Nieto GJ⁴

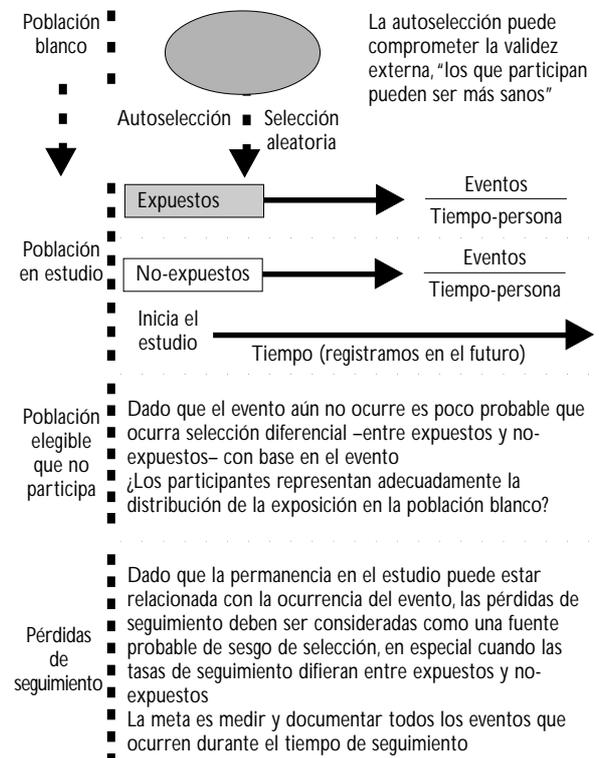


FIGURA 3. SESGOS DE SELECCIÓN EN ESTUDIOS DE COHORTE PROSPECTIVOS. POBLACIÓN ELEGIBLE QUE NO PARTICIPA

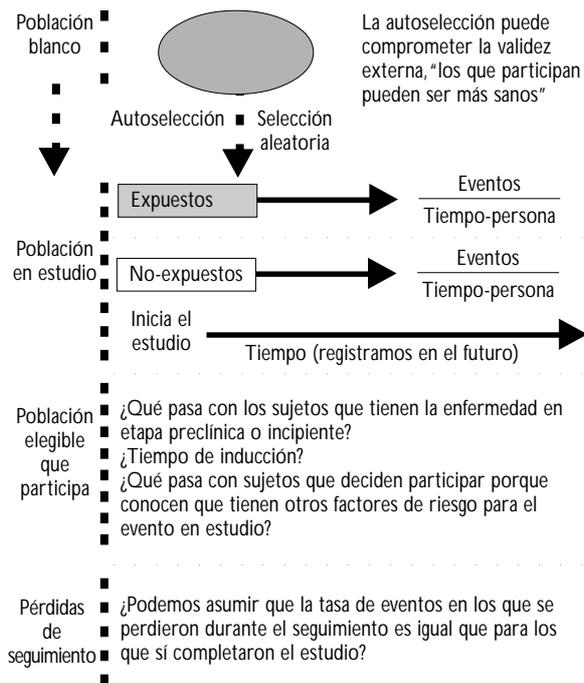


FIGURA 4. SESGOS DE SELECCIÓN EN ESTUDIOS DE COHORTE PROSPECTIVOS. POBLACIÓN ELEGIBLE QUE PARTICIPA

aleatorio, éste puede ocurrir tanto en la medición de la exposición como de la enfermedad o evento de interés.

Sesgos de selección

Los sesgos de selección en una cohorte tienen que ver tanto con la validez interna como con la validez externa o extrapolación de los resultados que se obtengan. Este tipo de sesgos está relacionado, evidentemente, con el procedimiento utilizado para conformar la cohorte o población en estudio: cuando la población en estudio se constituye con voluntarios, la representatividad que este grupo pueda tener de la población blanco (de la población a la cual se pretende generalizar los resultados) puede estar limitada por el hecho de que los voluntarios son diferentes en algunos aspectos de la población general. Por ejemplo, en el Estudio europeo sobre dieta y cáncer,⁷ la cohorte española se ha constituido a partir de donadores de sangre, es decir, hombres y mujeres altruistas, mayoritariamente jóvenes. Sin embargo, es pertinente mencionar que en el ejemplo anterior la validez interna de los resultados no se vería afectada. ¿Cómo se puede constituir una cohorte



FIGURA 5. SESGOS DE SELECCIÓN EN ESTUDIOS DE COHORTE RETROSPECTIVOS. POBLACIÓN ELEGIBLE QUE NO PARTICIPÓ

representativa de una determinada población? Un diseño experimentado con cierto éxito es el de seguimiento de muestras representativas de la población general. Un ejemplo de este tipo de diseño lo constituye el estudio de Sepúlveda y colaboradores^{8,9} que se basó en el seguimiento activo de los cerca de 3 241 niños de una muestra representativa de la Delegación Tlalpan, en la ciudad de México, que permitió la investigación prospectiva del papel del estado nutricional como un factor de riesgo para enfermedades diarreicas, en una cohorte representativa de un área geográfica de la ciudad de México.

Sin embargo, a pesar de contar con una cohorte inicial representativa, ¿cómo se puede asegurar que determinadas personas que retiren su participación del estudio, ya sea en la fase inicial de recolección de información o incluso después de cierto tiempo de seguimiento, no condicionen diferencias en los grupos estudiados que conduzcan a errores en los resultados? Las pérdidas en el seguimiento son, justamente, la principal causa de sesgo de selección, ya se trate de cohortes constituidas a partir de muestras representativas de la población, voluntarios o colectivos definidos

(como pueden ser los trabajadores de una determinada fábrica o sector industrial o grupos profesionales determinados como médicos, enfermeras o de profesionales de la educación). Las pérdidas en el seguimiento no invalidan *per se* el estudio, pero los investigadores deben utilizar procedimientos para minimizar su ocurrencia y, en caso de que se presenten, considerar si afectan o no los resultados observados en el estudio. Para ello, se intentará recoger información clave de los participantes que abandonen el estudio, en especial para investigar si el abandono tiene alguna relación con las exposiciones o con las enfermedades o eventos estudiados.

Un claro ejemplo de sesgo de selección que puede alterar considerablemente los resultados obtenidos es el llamado efecto del trabajador sano, observado frecuentemente en cohortes laborales cuando el grupo no-expuesto queda constituido por la población general. Por ejemplo, al estudiar la mortalidad por cáncer y por enfermedades cerebrovasculares de los trabajadores de una determinada empresa relacionada con productos radioactivos, y compararla con la mortalidad esperada de acuerdo con las tasas observadas en la población general, se observa que el riesgo relativo de muerte por esas causas es de 0.78 y 0.81, ambos con intervalos de confianza que no incluyen la unidad.¹⁰ Los resultados parecerían sugerir que el trabajo en esa empresa protege a sus empleados con una disminución de la frecuencia con la que se presentan las enfermedades estudiadas. ¿Cómo explicar estos resultados? La cohorte de trabajadores representa un grupo especial de personas de la población general: son personas en edad laboral y con capacidad física óptima, existe una autoselección de las personas que van a desarrollar tareas concretas en esa fábrica, de manera que personas con determinados trastornos crónicos se autoexcluyen (o son excluidos por la dirección de la empresa) o pasan a trabajos menos pesados. Así, al comparar la mortalidad en ese grupo de trabajadores sanos con el experimentado en la población general del país, la observada en la empresa debe ser inferior a la de aquél. Una manera de evitar este tipo de sesgo de selección es realizando comparaciones internas en el seno de la cohorte (por ejemplo, la mortalidad por cáncer en los trabajadores de una misma empresa pero expuestos a diferentes niveles de radiación) con lo que se puede eliminar en buena parte el sesgo del trabajador sano.¹¹ Los estudios de cohorte retrospectiva son más vulnerables a los sesgos de selección, esto se debe a que al inicio de este tipo de estudios el evento ya ocurrió en un buen número de los participantes y a que esto puede influir en la probabilidad de participación en el estudio. Este tipo de sesgo en estudios retrospectivos es

particularmente serio cuando los participantes conocen también su condición de exposición y cuando la presencia conjunta de estos eventos (exposición y enfermedad) motiva una participación diferencial en el estudio.

Otro tipo de sesgo de selección, acaso el más problemático y frecuente en los estudios de cohorte, es el debido a pérdidas en el seguimiento. Si las pérdidas están relacionadas con alguna característica de los participantes como puede ser la misma exposición o el desenlace estudiado, las pérdidas determinarán de alguna manera (es decir, estarán sesgando) los resultados obtenidos. Así, las pérdidas deben ser independientes de la condición de exposición, es decir, se deben presentar con la misma frecuencia en los grupos expuesto y no-expuesto. En este caso, como en el caso del efecto del trabajador sano, el sesgo de selección introducido por las pérdidas durante el seguimiento compromete la validez interna del estudio, es decir, los grupos expuesto y no-expuesto no son comparables por lo que se pierde la veracidad de los resultados. Uno de los motivos más frecuentes de pérdidas en el seguimiento obedece a la movilidad de los participantes de la cohorte, que dificulta los contactos repetidos que se utilizan tanto para realizar nuevas determinaciones de las exposiciones como para conocer el desarrollo de las condiciones estudiadas (incidencia de enfermedades o muerte). Para mantener y maximizar la participación, y minimizar las pérdidas en el seguimiento, es necesario implementar diferentes estrategias durante el desarrollo del estudio.¹²

Un ejemplo de estudio de cohorte en el que se ha invertido grandes esfuerzos en el seguimiento es el estudio de las enfermeras estadounidenses. Este estudio comenzó en 1976, con la inclusión de 121 700 enfermeras certificadas, en edades comprendidas entre 30 y 55 años, que contestaron un cuestionario postal sobre estilos de vida y condiciones médicas.¹³ Su objetivo primario era investigar la relación entre el consumo de anticonceptivos orales y el cáncer de mama, y tras más de 25 años de seguimiento ha generado múltiples resultados con relación a ese y a otros numerosos problemas de salud. ¿Cómo han conseguido los investigadores realizar un seguimiento continuado de las más de 100 000 participantes?

1. El diseño inicial se centró en enfermeras certificadas, de manera que aunque cambien de domicilio y de localidad de residencia, si siguen ejerciendo su profesión, seguirán activas y podrán ser localizadas.
2. En el cuestionario inicial se solicitó, además del nombre de la participante, su número de la segu-

- ridad social, la fecha de nacimiento y el nombre, la dirección y el número de teléfono de un contacto personal.
3. Además, gran parte del esfuerzo se concentra en el seguimiento mediante el cuestionario postal: se envían cuestionarios postales de seguimiento cada dos años, que se acompañan de una carta de presentación y un boletín con información actualizada de los progresos del estudio; se actualiza la información de los contactos personales cada cuatro años; se envía hasta cinco veces el cuestionario (la quinta vez se trata de una versión abreviada) en caso de no respuesta.
 4. Tras el quinto envío postal se realiza un seguimiento telefónico, se utiliza correo certificado o de mensajería privada, se consulta a los carteros locales, los colegios de enfermería y los contactos personales.
 5. Finalmente, una parte importante del seguimiento de la mortalidad se realiza mediante el uso del Registro Nacional de Defunciones estadounidense,¹⁴ que es un índice sistematizado de todas las muertes acaecidas en los Estados Unidos de América, desde 1979. De esta manera, mediante la utilización combinada de estas estrategias, desde 1990, la cohorte de las enfermeras cuenta con 90% de seguimiento.

Sesgos de información

La introducción de errores sistemáticos que comprometan la validez interna del estudio por el modo en que se obtuvo la información o los datos de los participantes se conoce como sesgos de información. El sesgo de información en los estudios de cohorte se presenta cuando la información se obtiene de manera diferente en los grupos estudiados; por ejemplo, cuando los participantes en el grupo expuesto son seguidos, monitorizados o vigilados de manera más cuidadosa que los participantes en el grupo no-expuesto se introduciría un sesgo de información. En este mismo sentido, en estudios clínicos de seguimiento es frecuente que algunos participantes presenten condiciones comórbidas que generen, incluso de manera no apreciable para los investigadores, una mayor vigilancia o control de esos pacientes con relación a otros sujetos del estudio, aumentando artificialmente las posibilidades de diagnóstico de la condición de interés. Por ejemplo, entre las participantes posmenopáusicas, en la cohorte de las enfermeras estadounidenses es más probable que las enfermeras que toman terapia sustitutiva hormonal se sometan a un seguimiento más riguroso que, acaso, incluya prácticas preventivas

adicionales como es la detección oportuna de cáncer de mama, lo que haría más probable la detección de tumores mamarios en este grupo de la cohorte, en comparación con las enfermeras posmenopáusicas que no siguen un tratamiento sustitutivo y llevan un seguimiento médico y ginecológico habitual.¹⁵ Así, al cuantificar la asociación entre la terapia hormonal sustitutiva y el riesgo de cáncer mamario se observaría un incremento de la incidencia de la enfermedad debido al sesgo de información (mayor detección de cáncer de mama) entre las enfermeras en tratamiento sustitutivo.

En ocasiones es el propio investigador quien evalúa de forma sesgada la presencia o no de la condición de interés, puesto que conoce las hipótesis bajo investigación o la historia de exposiciones de los participantes. Este tipo de error se conoce como sesgo del observador. Imaginemos que en el estudio mencionado anteriormente “estamos convencidos” de que la terapia sustitutiva aumenta el riesgo de cáncer mamario, por lo que podríamos incurrir en un sesgo del observador si se intensificara el seguimiento de las enfermeras que toman la terapia sustitutiva, a partir de la idea de que se encuentran en mayor riesgo de padecer el evento, o al revés, podríamos minimizar el seguimiento de las enfermeras que no consumen hormonas. El problema se deriva de realizar un esfuerzo diferente en el seguimiento de los expuestos y los no-expuestos, derivado de las hipótesis *a priori* –o en algunos casos de la propia organización logística– del estudio. ¿Cómo se puede evitar el sesgo de información? Garantizando que todas las mediciones realizadas tanto mediante cuestionarios como por muestras biológicas se realicen con el mismo grado de error (misma sensibilidad y especificidad) en el grupo expuesto y no-expuesto. Esto se puede lograr en algunos estudios manteniendo ciegos tanto a los participantes como a los observadores sobre la condición de exposición y la hipótesis de estudio.

Mala-clasificación o sesgos de clasificación no diferencial

En los estudios de cohorte se deben tener en cuenta los sesgos de información debidos a la clasificación errónea (mala-clasificación) de los participantes respecto a la existencia o a la cuantificación de la exposición estudiada o a la ocurrencia de la enfermedad o de la condición de interés.¹⁶ ¿De qué depende este tipo de error? La principal fuente de sesgo deriva de los instrumentos utilizados y de su modo de aplicación (cuestionarios, técnicas analíticas, biomarcadores, etc.). Si disponemos de un cuestionario para determinar el consumo de alcohol semanal que produce estimaciones sesgadas in-

feriores a las reales (por ejemplo, porque no considera de forma independiente el consumo de alcohol de fin de semana del consumo del resto de la semana) incurriremos en una infravaloración del consumo real de alcohol. Esa subestimación ¿se produce del mismo modo en los grupos expuesto y no-expuesto, entre los sujetos que desarrollaron el evento y los que no? Esa es una cuestión clave que nos permitirá caracterizar la mala-clasificación en dos tipos: a) diferencial, cuando el error en la clasificación depende del valor de otras variables, y b) no-diferencial, cuando el error no depende de las otras variables. Volvamos al ejemplo anterior. Supongamos que el cuestionario sobre consumo de alcohol es administrado por entrevistadores en el domicilio de los participantes. Si el protocolo de administración no se aplica estrictamente o ha fallado el adiestramiento de los entrevistadores, podría suceder que algunos entrevistadores enfatizen e investiguen hasta la última gota de alcohol consumida por el participante, sobre todo cuando se entrevista a varones de mediana edad en zonas o barrios deprimidos de la ciudad en los que el consumo de alcohol parece una conducta habitual. Sin embargo, al entrevistar a mujeres, también de mediana edad, de zonas más favorecidas, el entrevistador no enfatiza la metodología de la entrevista, asumiendo que los participantes de esas características no tienen un consumo de alcohol importante. Nos encontramos, pues, ante un típico ejemplo de mala-clasificación de la exposición, en este caso diferencial en lo que se refiere a la variable de nivel socioeconómico. Cuando el error afecta por igual a todos los participantes, independientemente de su exposición verdadera (los entrevistadores aplican de manera similar el cuestionario, pero éste recoge consumos inferiores de alcohol sistemáticamente) nos hallamos ante mala-clasificación no diferencial o aleatoria. En este último caso, el sesgo introducido tiende a modificar las estimaciones del riesgo relativo hacia la hipótesis nula, es decir, hacia la subestimación de la verdadera asociación. Sin embargo, la dirección del sesgo en el caso de mala-clasificación diferencial depende del tipo de exposición investigada y es, en muchos casos, impredecible.

Lo comentado para el caso de la medida de la exposición es aplicable igualmente a la determinación de la condición o enfermedad estudiada. Veamos un último ejemplo. En una cohorte de adolescentes, constituida durante el primer año de la escuela secundaria del estado de Morelos, en México, se recogieron, en el momento de constitución de la cohorte, diferentes variables sociodemográficas y de hábito tabáquico, y variables sobre el rendimiento escolar. Cinco años después, se desea realizar una entrevista y examen físico a to-

dos los participantes, con la intención de investigar los efectos a corto y mediano plazo del tabaquismo. Debido a que los investigadores conocen las características basales de los participantes, y saben que el consumo de tabaco es más frecuente entre los adolescentes con menor nivel económico y menor rendimiento, en el momento de determinar la presencia de enfermedades (mediante un cuestionario y un examen médico sencillo) tienden a investigar con más detenimiento a los participantes con esas características, aumentando, así, de manera diferencial la probabilidad de diagnóstico de enfermedad en los expuestos.

Las posibles soluciones para minimizar los errores de medición aleatorios consisten en la validación de los instrumentos de medida utilizados, ya sean cuestionarios estructurados, pruebas psicométricas, instrumentos médicos (esfigmomanómetros, balanzas, y otros), técnicas de laboratorio, etcétera, junto con la implantación de protocolos de aplicación estrictos, previo entrenamiento y estandarización, sobre todo cuando los observadores son múltiples e, incluso, de diferentes centros y la realización de medidas repetidas en los mismos sujetos. Finalmente, la realización de una prueba piloto en condiciones reales y la implementación de controles de calidad continuados de la información recolectada pueden alertar sobre la aparición de este tipo de sesgos¹⁷ que difícilmente son controlables en las fases de análisis del estudio.

De lo comentado se deduce que los sesgos pueden minimizarse con un buen diseño, que en los estudios de cohorte se ha de incluir la planificación detallada de la constitución de la cohorte y de los mecanismos de seguimiento, además de los instrumentos de captura de la información. La utilización de métodos estadísticos nos permite controlar en el análisis algunos de los sesgos aleatorios que no pudieron ser prevenidos en el diseño, tal como el análisis estratificado y los métodos multivariados. Un estudio libre de sesgos nos garantizará su validez interna, así como su validez externa o extrapolación.

Conclusiones

La utilización de los estudios de cohorte ha aumentado considerablemente durante los últimos años. Como parte de la revolución informática ocurrida en la práctica de la epidemiología se han desarrollado avances considerables en los métodos que permiten el seguimiento eficiente y costo-efectivo de grandes y diversos grupos poblacionales, así como la aplicación de métodos estadísticos sofisticados que permiten prevenir, corregir y controlar diferentes sesgos y examinar las relaciones epidemiológicas en un contexto más

controlado. Estos avances han permitido el desarrollo de conocimiento derivado de estudios epidemiológicos de cohorte que han tenido un gran impacto en la práctica médica; un ejemplo destacado de esto lo constituyen los datos sobre el efecto del tabaquismo y el plomo sobre la salud humana. Sin duda, los estudios de cohorte han modificado la percepción de los estudios observacionales y ahora se consideran una herramienta importante para el avance del conocimiento médico. La continua mejoría de los registros médicos y epidemiológicos en México, la implementación de bancos de biomarcadores –como los que están siendo implementados en las Encuestas Nacionales de Salud–, así como la capacidad técnica e infraestructura institucional desarrolladas en países como el nuestro, sin duda contribuirán a que este tipo de diseño sea utilizado con mayor frecuencia en nuestro medio para el estudio de diversas exposiciones ambientales, infecciosas, nutricionales, así como exposiciones asociadas a diversos estilos de vida en diversos contextos geopolíticos y grupos de riesgo, lo que nos permitirá estudiar con mayor rigor metodológico el proceso salud enfermedad y el impacto de los diferentes programas de intervención.

Referencias

1. Breslow NE, Day NE. Statistical methods in cancer research. The design and analysis of cohort studies. Lyon, Francia: IARC Scientific Publications, 1987;vol. I.
2. Rothman KJ, Greenland S. Modern epidemiology. 2a. edición. Boston: Lippincot-Raven, 1998:5-100.
3. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958; 53: 457-481.
4. Nieto GJ. Los estudios de cohorte. En: Martínez N, Antó JM, Castellanos PL, Gili M, Marset P, Navarro V, ed. *Salud Pública*. Madrid: McGraw-Hill / Interamericana, 1999.
5. MacMahon B, Trichopoulos D. *Epidemiology. Principles and methods*. 2a. edición. Boston: Little Brown and Company, 1996:165-225.
6. Rothman KJ, Greenland S. *Modern epidemiology*. 2a. edición. Boston: Lippincot-Raven, 1998:120-186.
7. European prospective investigation into cancer and nutrition. EPIC Group of Spain. Relative validity and reproducibility of a diet history questionnaire in Spain. I. *Foods. Int J Epidemiol* 1997;26 (Suppl 1):S91-S99.
8. Sepúlveda J, Willett W, Muñoz A. Malnutrition and diarrhea. A longitudinal study among urban Mexican children. *Am J Epidemiol* 1988;127:365-376.
9. Sepúlveda J. Malnutrition and infectious diseases. A longitudinal study of interaction and risk factor. Cuernavaca, Morelos: Instituto Nacional de Salud Pública 1990;(Perspectivas en Salud Pública).
10. Checkoway H, Pearce N, Crawford-Brown DJ. *Research methods in occupational epidemiology*. Nueva York: Oxford University Press, 1989: 77-80.
11. Checkoway H, Eisen EA. Developments in occupational cohort studies. *Epidemiol Rev* 1998; 20(1):100-111.
12. Hunt JR, White E. Retainig and tracking cohort study members. *Epidemiol Rev* 1998;20(1):57-70.
13. Colditz GA. The nurses' health study: A cohort of US women followed since 1976. *J Am Med Women Assoc* 1995;50(2):40-44.
14. Howe GR. Use of computerized record linkage in cohort studies. *Epidemiol Rev* 1998;20(1):112-121.
15. Franceschi S, La Vecchia C. Colorectal cancer and hormone replacement therapy: An unexpected finding. *Eur J Cancer Prev* 1998;7(6): 427-438.
16. Mertens TE. Estimating the effects of misclassification. *Lancet* 1993;342: 418-421.
17. Whitney CW, Lind BK, Wahl PW. Quality assurance and quality control in longitudinal studies. *Epidemiol Rev* 1998;20(1):71-80.

Estudios clínicos experimentales

Juan José Calva-Mercado, M.C., M.Sc.⁽¹⁾

Los estudios clínicos experimentales, también conocidos como ensayos clínicos controlados, son estrategias diseñadas para evaluar la eficacia de un tratamiento en el ser humano mediante la comparación de la frecuencia de un determinado evento de interés clínico (o desenlace) en un grupo de enfermos tratados con la terapia en prueba con la de otro grupo de enfermos que reciben un tratamiento control. Ambos grupos de enfermos son reclutados y seguidos de la misma manera y observados durante un mismo periodo de tiempo. La esencia de los ensayos clínicos es que el propio investigador decide qué individuos serán sometidos al tratamiento en prueba (grupo experimental) y quiénes estarán en el grupo comparativo (grupo control o de contraste). Es justamente esta propiedad la que distingue a los ensayos clínicos experimentales de los estudios observacionales (de cohortes) ya que en éstos el propio paciente o su médico tratante decide quién se somete (y quién no) a la maniobra en evaluación; esta decisión obedece a múltiples razones, algunas estrechamente ligadas al pronóstico de la enfermedad. Debido a que los ensayos clínicos controlados son estudios diseñados con antelación, la asignación de la maniobra experimental por el investigador puede seguir diversos procedimientos; cuando es mediante un sorteo el estudio se conoce como un ensayo clínico aleatorizado, o controlado por sorteo.

Cabe aclarar la connotación de diversos términos empleados en la anterior definición y a lo largo del presente capítulo. El uso del término tratamiento (o tera-

pia) es con un sentido amplio; se refiere no sólo a un medicamento sino que también incluye otro tipo de intervenciones (o maniobras) tales como un procedimiento quirúrgico, una medida preventiva (o profiláctica), un programa educativo, un régimen dietético, etcétera. De igual manera, el término evento (o desenlace) se puede referir a toda una diversidad de resultados, tales como: mediciones bioquímicas, fisiológicas o microbiológicas, eventos clínicos (intensidad del dolor, aparición de infecciones oportunistas, desarrollo de un infarto al miocardio, recaída de una leucemia aguda, etc.), escalas de actividad de una enfermedad (como la del lupus eritematoso generalizado), mediciones de bienestar o funcionalidad (calificación de Karnofsky, escala de calidad de vida) o el tiempo de supervivencia. Por último, el grupo control se refiere al grupo de individuos que reciben una intervención que sirve de contraste para evaluar la utilidad relativa de la terapia experimental y que no necesariamente tiene que ser un placebo pues en ocasiones lo más adecuado (y ético) es que sea el tratamiento estándar, es decir, la mejor alternativa terapéutica vigente en el momento del diseño del experimento clínico.

¿Por qué y cuándo son necesarios?

La pregunta nos remite a reflexionar sobre dónde surge la idea de que un cierto tratamiento pueda modificar la historia natural de una enfermedad y acerca de la necesidad de contar con suficientes observaciones

(1) Subdirección de Investigación Clínica, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán. México, D.F., México.

Solicitud de sobretiros: Juan José Calva Mercado. Subdirección de Investigación Clínica, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán. Vasco de Quiroga 15, Delegación Tlalpan, 14 000 México, D.F., México.
Correo electrónico: jcalva@quetzal.innsz.mx

sistematizadas para conocer el verdadero efecto de la terapia en cuestión en los seres humanos, antes de prescribirlos de manera rutinaria a los enfermos.

En ocasiones, surge la idea de que un fármaco pudiera ser clínicamente útil al comprender su íntimo mecanismo de acción, así como la patogenia a nivel celular y molecular de una determinada enfermedad. De hecho, en ocasiones resulta muy tentador intentar predecir el efecto de un cierto medicamento con base únicamente en la información generada en el laboratorio, tanto *in vitro* como en modelos animales de experimentación. Sin embargo, por muy profundo que sea este conocimiento siempre será incompleto si no se tiene la experiencia en seres humanos intactos; de no ser así, se corre el riesgo de sorpresas desagradables. Por ejemplo, se sabe que el antimetabolito citarabina interfiere con la síntesis de la pirimidina y que es capaz de inhibir *in vitro* a varios virus con ADN, incluyendo al virus herpes varicela-zóster. A algún ingenuo clínico se le ocurriría que pudiera beneficiar con este medicamento a sus enfermos con herpes zóster generalizado; afortunadamente a un grupo de médicos escépticos se les ocurrió comparar la evolución de un grupo de estos enfermos, a quienes se les administró la citarabina, con la de otro grupo de enfermos semejantes, quienes sólo recibieron placebo, y demostraron no sólo lo ineficaz del antiviral sino una peor evolución en los enfermos que la habían recibido, explicable por sus efectos inmunosupresores.

Otras dos fuentes de ideas sobre el posible valor de una cierta terapia suelen ser: a) las observaciones empíricas de clínicos perspicaces; así, por serendipia surgió la idea de que la amantadina pudiera ser de utilidad en los enfermos de Parkinson puesto que a quienes se les prescribía para prevenir la influenza mostraban mejoría de sus manifestaciones neurológicas o como el caso de la reducción de crisis de fiebre familiar del Mediterráneo con el uso de colchicina administrada con la intención de prevenir ataques de gota. En estos ejemplos, el valor de estos tratamientos no provino de un entendimiento de la patogenia de estas enfermedades, la que de hecho, aún no se conoce bien, y b) las observaciones de estudios poblacionales donde se llega a establecer una asociación directa (o inversamente) proporcional entre la frecuencia de una enfermedad y alguna condición ambiental; tal como ha sido el caso de una relativa menor frecuencia de enfermedad coronaria en poblaciones con mayor ingesta de alimentos ricos en antioxidantes, o de cáncer de colon y una dieta con alto contenido en fibra.

El punto esencial es que cualquiera que sea el origen de las hipótesis del posible beneficio de una terapia, éstas deben probarse, y demostrarse como ciertas,

mediante estudios clínicos; es decir, mediante la observación sistematizada y objetiva de su efecto en seres humanos que la reciben, y su comparación con lo que habitualmente sucede en un grupo de enfermos sin la terapia.

Hay circunstancias en las que, por una parte, se conoce muy bien la historia natural de una enfermedad y que es tan consistente que es posible, con razonable certeza, predecir su curso clínico (generalmente muy desfavorable) y, por la otra, que el beneficio de una cierta terapia en este tipo de enfermedades es tan dramático e incuestionable que resulta innecesario realizar un ensayo clínico para aceptar su uso generalizado. Es decir, es suficiente con comparar la evolución clínica de un grupo de casos tratados con la nueva terapia con lo que habitualmente pasa en los enfermos antes del acceso a ésta (controles históricos), tal y como ha sucedido con el beneficio de los antibióticos en el tratamiento de la neumonía bacteriana, de los antifímicos en la meningitis tuberculosa o de la cirugía abdominal en la apendicitis: no hay alguien en el mundo que haya exigido el haber realizado un ensayo clínico en el que un grupo de estos enfermos recibieran un placebo. Sin embargo, esta es una circunstancia excepcional; lo común es que diferentes individuos con la "misma" enfermedad tengan cursos disímiles (en ocasiones, en sentidos totalmente opuestos) e impredecibles, de tal suerte que esta incertidumbre pronóstica impide deslindar qué tanto la mejoría o las complicaciones observadas en un grupo de enfermos que reciben una nueva terapia es atribuible a ésta, o bien, es parte de su historia natural o consecuencia de otros determinantes ajenos a la intervención en evaluación. Por ejemplo, un grupo de enfermos puede tener una mejor evolución clínica a pesar de recibir un tratamiento totalmente inútil, porque son pacientes en etapas tempranas de su enfermedad, reciben otros medicamentos concomitantes o mejor atención médica, están bajo el efecto placebo de la terapia novedosa o porque comparten el entusiasmo de los investigadores. Es en estos casos, que constituyen más la regla que la excepción, cuando es indispensable (ética y científicamente) realizar (y exigir) el diseño, la conducción y la publicación formal de un estudio clínico experimental; el ensayo clínico controlado por sorteo es el diseño metodológico más confiable para distinguir si el beneficio atribuible a un tratamiento es real o sólo un espejismo.

En la historia de la medicina abundan los bochornosos ejemplos de tratamientos (novedosos en su época) a los que se le atribuyeron espectaculares beneficios (por lo que multitudes las recibieron) y que no fue hasta que un ensayo clínico aleatorizado los

puso en su justo lugar: en el archivo de terapias inútiles e, incluso, dañinas. Tal es el caso de la congelación gástrica en el tratamiento de la úlcera péptica, de la ligadura de la arteria mamaria interna en la prevención de la angina de pecho recurrente, de los esteroides en los pacientes con sepsis grave, de la plasmaféresis en los pacientes con polimiositis, entre muchos otros ejemplos.

También hay ejemplos de intervenciones terapéuticas o profilácticas ampliamente arraigadas en la práctica médica cotidiana actual pero cuyo beneficio real ha sido, en mayor o menor grado, cuestionado; tal como sucede con la vacuna antituberculosa con el bacilo de Calmette-Guerin, el uso de antibióticos en la otitis media aguda no complicada en los niños o el internamiento de enfermos en una unidad coronaria. Sin embargo, estos son ejemplos en los que resulta casi imposible (por razones éticas o logísticas) llevar a cabo en la actualidad un estudio clínico experimental, por lo que parece que estamos condenados a no resolver el dilema de su verdadero costo-beneficio. De ahí que hay quien recomienda que se realicen los ensayos clínicos controlados de manera temprana en la fase de evaluación de nuevas terapias, antes de que sean adoptadas de manera irreflexiva por la comunidad médica. Sin embargo, también debe tomarse en cuenta que lo común es que los ensayos clínicos sean estudios costosos y logísticamente complejos por lo que deben plantearse teniendo información completa de su farmacología, de su toxicidad y preliminar de su posible eficacia.

¿Qué determina la validez de sus resultados?

Esta pregunta tiene que ver con la confianza que lleguemos a tener de que los resultados del ensayo clínico revelen con exactitud la dirección y magnitud de lo que les pasa a los sujetos bajo estudio y que esto sea realmente atribuible al tratamiento en prueba. Es necesario que tengamos los elementos para evaluar si los resultados representan una estimativa no sesgada del efecto de la intervención o, si por el contrario, son conclusiones falsas determinadas por algún error sistemático. A continuación se describen los principales componentes, porque no son los únicos, que determinan la validez interna de un ensayo clínico.

La asignación del tratamiento en prueba por medio de un sorteo

La decisión de a cuál grupo pertenecerá cada uno de los individuos participantes en el estudio debe ser aje-

na a cualquier predisposición, o prejuicio, del investigador y la mejor manera de lograrlo es mediante el empleo de un proceso aleatorio: es únicamente el azar quien se encarga de asignar a los sujetos (figura 1).

En la década de los setenta y principios de los ochenta era común que a los pacientes con insuficiencia arterial cerebral se les practicara una cirugía de revascularización extra-intracranéa mediante la colocación de un puente en el que se unía una rama de la arteria carótida externa (la temporal superficial) con una rama cortical de la arteria carótida interna (la cerebral media). Los cirujanos que la practicaban entonces lo hacían con mucho entusiasmo convencidos de que su procedimiento realmente prevenía la oclusión de las arterias cerebrales en sus nobles pacientes. Esta creencia se basaba en la observación repetida de que grupos de pacientes así operados con menor frecuencia sufrían eventos isquémicos cerebrales en comparación con los enfermos que no eran sometidos a la cirugía. El detalle importante es que estas eran observaciones provenientes de la práctica clínica cotidiana, es decir, no se habían obtenido de un estudio planeado en donde un grupo de investigadores clínicos hubieran decidido, mediante un procedimiento semejante

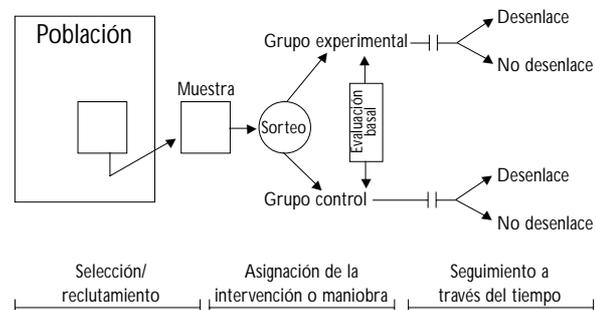


FIGURA 1. DIAGRAMA DE LA ESTRUCTURA DE UN ENSAYO CLÍNICO CONTROLADO POR SORTEO (ALEATORIZADO).

LA POBLACIÓN DE INTERÉS SE DEFINE CON BASE EN LOS CRITERIOS DE INCLUSIÓN/EXCLUSIÓN Y LA MUESTRA ESTUDIADA GENERALMENTE SE SELECCIONA Y RECLUTA POR SU ACCESIBILIDAD. LA ASIGNACIÓN DE LOS PARTICIPANTES AL GRUPO EXPERIMENTAL (QUIEN RECIBE LA INTERVENCIÓN BAJO EVALUACIÓN) O AL GRUPO CONTROL (QUIEN RECIBE LA INTERVENCIÓN DE CONTRASTE: EL PLACEBO O LA MANIOBRA HABITUAL) LA REALIZA EL PROPIO INVESTIGADOR MEDIANTE UN PROCEDIMIENTO ALEATORIO (SORTEO). EL EFECTO DE LA INTERVENCIÓN EN ESTUDIO SE MIDE AL COMPARAR LA INCIDENCIA DEL DESENLACE DE INTERÉS EN EL GRUPO EXPERIMENTAL CON LA DEL GRUPO CONTROL

al de un “volado”, quiénes era operados y quiénes no. En el quehacer clínico cotidiano no es nada raro que el cirujano ofrezca sus servicios (o que el paciente se deje operar) preferentemente a pacientes que están relativamente en mejores condiciones que otros; es decir, que desde antes de la cirugía gozan de un inherente mejor pronóstico, en cuanto al riesgo de oclusión arterial, de complicaciones de la cirugía e, incluso, de supervivencia; en cambio (y por razones totalmente entendibles), no suelen hacerlo con pacientes más deteriorados, en quienes hay menos posibilidades de que su intervención quirúrgica resulte exitosa. Esta práctica puede hacer entonces que, como los pacientes con menos posibilidades de una buena evolución son incluidos preferentemente en el grupo (cohorte) control (o de contraste), un procedimiento ineficaz parezca (de manera distorsionada) como todo lo contrario. Es justamente mediante el sorteo (también conocido como asignación aleatoria de la maniobra, que sólo puede hacerse en un estudio experimental) que se pretende que todos los pacientes tengan la misma oportunidad de recibir el tratamiento en prueba; de tal suerte que las diferencias en el desenlace clínico entre los sometidos (o no) al tratamiento sean explicables únicamente por este hecho y no por otros determinantes ajenos a la maniobra en estudio. Resulta interesante que fue mediante un enorme estudio experimental en humanos, aleatorizado, multicéntrico y publicado en 1987, que se logró demostrar tanto que a los pacientes así operados en realidad les va peor que a los tratados únicamente con medicamentos en el periodo posoperatorio inmediato, como que la evolución clínica a largo plazo era idéntica entre los sometidos o no a la revascularización quirúrgica.

La evolución clínica de los enfermos obedece a múltiples causas, el tratamiento sólo es una de ellas; de tal forma que es común que la gravedad de la enfermedad, la presencia de comorbilidad y toda una gama de determinantes pronósticos (unos que podemos identificar y otros muchos que desconocemos) hagan que tengamos una falsa impresión de las bondades (perjuicios) de los tratamientos que ofrecemos. Es por ello que las decisiones de los médicos, basadas en información proveniente de estudios observacionales (o de experimentales no aleatorizados) estén propensos a caer en este tipo de errores. De hecho, se ha documentado que los estudios en los que el tratamiento se ha asignado por cualquier otro método que no haya sido un sorteo tienden a mostrar efectos de mayor beneficio (y frecuentemente falsos) que lo observado en los ensayos clínicos aleatorizados. La ventaja de la aleatorización es que logra, si el tamaño de la muestra es suficientemente grande, que los deter-

minantes de la evolución de los enfermos (tanto los conocidos como los desconocidos) estén distribuidos de manera equilibrada entre el grupo de enfermos que reciben la maniobra bajo estudio (grupo experimental) y los que no la reciben (grupo control).

Seguimiento completo de los individuos en estudio

Este punto tiene que ver tanto con que se logre un seguimiento completo de todos los sujetos participantes en el estudio durante el tiempo programado como con que se incluyan en el análisis a todos los individuos respetando su pertenencia al grupo (experimental o control) al que fueron originalmente asignados.

Todo paciente que ingresa al ensayo debe ser tomado en cuenta en el análisis y conclusiones; de lo contrario, si un número importante de ellos se reporta como “sin seguimiento suficiente” la validez del estudio se verá seriamente cuestionada. Mientras más individuos se pierden mayor será la posibilidad de sesgo en el estudio debido a que los pacientes que no completan su seguimiento pueden tener un pronóstico diferente al de quienes sí permanecen hasta el final del estudio. Las pérdidas en el seguimiento pueden obedecer a la aparición de eventos adversos al medicamento en estudio, a una mala respuesta al mismo, a muerte, etcétera o, al contrario, porque tienen una evolución particularmente benigna y su bienestar hace que no regresen a sus evaluaciones. Así, al ocurrir esto se puede tener una impresión distorsionada de las bondades (o de su ausencia) del tratamiento bajo prueba.

Tal como sucede en la práctica médica cotidiana, los pacientes en los estudios experimentales dejan de tomar sus medicamentos, ya sea por olvido o por efectos indeseables de los mismos. Excluir a estos individuos –que no se adhirieron de manera completa a los lineamientos del protocolo– del análisis puede igualmente generar sesgos. Igual que en el punto anterior, es frecuente que esta falta de apego esté estrechamente vinculada con un pronóstico diferente al de los enfermos cumplidores, tal como se ha llegado a constatar en algunos ensayos clínicos donde incluso los pacientes a quienes se les asignó recibir el placebo tuvieron una mejor evolución que los que mostraron un buen apego a su ingesta en comparación con los incumplidos. Al excluir a éstos del análisis uno corre el riesgo de sólo evaluar a pacientes con un mejor pronóstico, alterando la condición de equidad entre grupos dada por la asignación por sorteo de la maniobra.

Es justamente este principio de incluir en el análisis a todos los pacientes, tal como fueron asignados de

acuerdo con el sorteo, el que define el análisis de “intención a tratar”. Esta estrategia metodológica conserva el valor de la aleatorización, es decir, el de lograr que los pacientes con diferentes pronósticos de la enfermedad queden igualmente distribuidos en los grupos (brazos) del ensayo.

Evitar que las expectativas de los pacientes y de sus evaluadores influyan en la medición de los desenlaces

Igual que los investigadores, los enfermos que saben que están recibiendo un nuevo tratamiento en experimentación generalmente tienen una idea prejuiciada sobre su eficacia. Estas expectativas, ya sea optimistas o desfavorables, pueden llegar a distorsionar la medición de los resultados particularmente cuando su evaluación es mediante indicadores propicios a ser influidos por la subjetividad del informante o del propio evaluador (“datos blandos”). Esto puede ocurrir cuando se miden síntomas (como el dolor), sentimientos de bienestar (como la calidad de vida), o signos clínicos sin definiciones suficientemente objetivas y poco reproducibles. En estas condiciones, y cuando los investigadores conocen quién de sus pacientes del estudio recibe el tratamiento experimental y quién no, se pueden sesgar los resultados al hacer una búsqueda diferencial de los desenlaces o al dar interpretaciones diferentes a los hallazgos.

La mejor manera de evitar este tipo de errores es mediante el enmascaramiento, que no es otra cosa que tratar de que ni el enfermo ni el investigador que lo evalúa (doble cegamiento) sepan si aquél se encuentra recibiendo la terapia en prueba o la intervención de contraste. Esto habitualmente se logra administrando un placebo (o la terapia habitual) con apariencia, sabor y textura indistinguibles de la terapia experimental. De no ser posible esto, y si el tipo de desenlaces a medir lo amerita, habrá que diseñar una estrategia en la que el investigador que los evalúa desconozca a qué grupo de estudio pertenecen sus pacientes.

Brindar igual atención médica, fuera de la terapia en estudio, a los grupos del ensayo

Si los cuidados médicos, incluyendo otras terapias diferentes a la que se estudia, se dan de manera diferente en el grupo experimental que en el control se corre el riesgo de comprometer seriamente los resultados del ensayo al cometer el sesgo denominado de “cointervención”; es decir, será difícil distinguir qué tanto de la diferencia (o no diferencia) observada entre los grupos es efecto de la terapia en estudio vs. consecuencia de una atención médica diferencial. Una

manera de evitar este tipo de error sistemático puede ser mediante el diseño de un ensayo doble ciego.

¿Qué determina que sus resultados sean aplicables a otros grupos de enfermos?

Lo común es que una de las principales motivaciones que lleva a un investigador a realizar un ensayo clínico es el deseo de que la terapia experimental eventualmente sea utilizada por otros médicos, en otros sitios, como una medida útil y que se brinde para el beneficio real del mayor número de enfermos. Esto dependerá fundamentalmente de tres aspectos.

Primero, de qué tanto se parezcan los enfermos que estudia a los enfermos en quienes posteriormente se quiera extrapolar la experiencia obtenida en el ensayo clínico. Esto depende del rigor en la selección de los sujetos a incluirse en el experimento; mientras más criterios de inclusión sea necesario cumplir menos posibilidades de que los resultados sean aplicables a otras poblaciones de enfermos. De hecho, el investigador suele enfrentarse al dilema de que por querer estudiar a una población relativamente homogénea de enfermos que le permita incrementar la eficiencia de su estudio (es decir, demostrar con claridad un efecto benéfico de la terapia experimental con el menor número de enfermos y costos posibles) establece criterios de inclusión al estudio demasiado estrictos que lo alejan de las características del quehacer médico cotidiano: de la “vida real”.

Segundo, del desenlace que se haya elegido para valorar la eficacia de la terapia experimental. Habitualmente y por razones de eficiencia, el investigador suele elegir mediciones fisiológicas, bioquímicas, microbiológicas (o de otra estirpe) más fáciles de medir, en corto tiempo y con menos pacientes como sustitutos de desenlaces clínica o socialmente más importantes (como puede ser la calidad de vida o la supervivencia) bajo el supuesto de que aquéllas pueden predecir éstos. Así, no es raro que en los estudios experimentales se evoque el beneficio de una nueva terapia porque ésta mejora las pruebas de función respiratoria, o porque disminuye la carga viral en el plasma, o los niveles séricos de colesterol (por mencionar sólo algunos ejemplos), asumiendo que al lograrlo los enfermos vivirán más tiempo y mejor. Es necesario ser cauteloso en la selección de los resultados a evaluar y que sean realmente importantes para los individuos y para la sociedad; de lo contrario, se corre el riesgo de promulgar la bondad de una terapia nueva porque mejora una medición intermedia, pero que a la larga se demuestre su efecto nocivo en desenlaces clínicamente más importantes o incluso en la supervivencia de los

sujetos. Ejemplos de ello es lo sucedido con el uso de hipolipemiantes (como el clofibrato y ciertas estatinas) en la reducción del colesterol sérico o de ciertos antiarrítmicos (como la encainida) posterior a un infarto al miocardio.

Tercero, de que la intervención que se evalúa sea única y precisa, como lo es un medicamento. En este caso su reproducción por otros investigadores (u otros médicos) es fácil. Por el contrario, hay tratamientos que implican varios elementos cambiantes o que demandan un grado de habilidad muy particular de quien realiza la maniobra (como sería el caso del manejo médico de un cierta entidad clínica en un determinado ambiente, una intervención quirúrgica, un programa educativo, etc.) cuyos efectos no necesariamente serán los mismos cuando son realizados por otros médicos o investigadores.

¿Cómo se analizan sus resultados?

Lo más común es que los ensayos clínicos controlados midan la incidencia de algún evento en los grupos de individuos seguidos en un determinado lapso y que este evento se exprese de manera dicotómica (es decir, la presencia o no del desenlace: infarto al miocardio, recurrencia de una neoplasia, muerte, etc.) como la proporción de sujetos que llegan a presentarlo. Pongamos como ejemplo un estudio en el que 20% (0.20) de los enfermos en el grupo control fallecieron en contraste con sólo 15% (0.15) de los que recibieron el tratamiento en evaluación. En el cuadro I se resume la forma como se puede presentar el efecto de éste, a saber:

1. La diferencia absoluta (o la reducción del riesgo absoluto -RRA-), que se obtiene al sustraer la proporción de individuos que fallecieron en el grupo experimental (Y) de la proporción de individuos que lo hicieron en el grupo control (X): $X - Y = 0.20 - 0.15 = 0.05$ (5%).
2. El riesgo relativo (RR), es decir, el riesgo de morir en los pacientes sometidos a la terapia experi-

mental con relación al de los pacientes en el grupo control: $Y/X = 0.15/0.20 = 0.75$.

3. El complemento del riesgo relativo (o la reducción del riesgo relativo -RRR-) que se expresa como un porcentaje: $[1 - (Y/X)] \times 100 = [1 - 0.75] \times 100 = 25\%$. Esta cifra significa que el nuevo tratamiento reduce el riesgo de morir en 25% con relación a lo que ocurre en los pacientes del grupo control; mientras mayor sea la RRR mayor es la eficacia del tratamiento.
4. El número necesario de pacientes a tratar (NNT) indica si el beneficio ofrecido por la nueva terapia retribuye el esfuerzo y costo en su adquisición o implantación. Por ejemplo, una reducción de 25% en el riesgo de morir puede parecer impresionante, pero su impacto en el paciente o en la práctica clínica puede, sin embargo, ser mínimo. La utilidad de un tratamiento está no sólo en función de la reducción relativa del riesgo sino también del riesgo del desenlace adverso que se quiere prevenir (en nuestro ejemplo, la muerte); de tal forma que mientras menor sea este riesgo mayor será el número necesario de enfermos a tratar con la nueva terapia para prevenir una muerte (es decir, menor su impacto). En el cuadro II se ilustran dos circunstancias: una en la que el riesgo de muerte en una población de enfermos, en un determinado periodo de tiempo, es sólo de 1%; en contraste con otra población con un riesgo mayor (de 10%). En el primer caso, la nueva terapia reduciría el riesgo de fallecer en 25%, es decir, una reducción del riesgo absoluto de 0.0025 (o 25 muertes en 10 000 pacientes tratados). El número necesario de pacientes a tratar se obtiene al calcular la inversa de esta reducción del riesgo absoluto ($1/0.0025 = 400$); así, sería necesario tratar 400 pacientes durante un tiempo determinado para salvar una sola vida. En cambio, en el segundo caso, una reducción relativa de 25% de muerte en una población en mayor riesgo de morir (de 10%) lleva a una reducción del riesgo absoluto de 0.025 (o 25 muertes en 1 000

Cuadro I
MEDICIONES DEL EFECTO DE UN TRATAMIENTO EN EVALUACIÓN*

Reducción del riesgo absoluto (diferencia de riesgos) (RRA)	$X - Y$	$0.20 - 0.15 = 0.05$ o 5%
Riesgo relativo (RR)	Y/X	$0.15/0.20 = 0.75$
Reducción del riesgo relativo (RRR)	$1 - (Y/X)$ $X \times 100$ o $(X - Y)/X \times 100$	$1 - 0.75 \times 100 = 25\%$ $0.05/0.20 \times 100 = 25\%$
Número de pacientes necesario a tratar para prevenir un evento (NNT)	$1/(X - Y)$	$1/(0.20 - 0.15) = 20$

* X= riesgo del evento en los pacientes sin el tratamiento (grupo control) $20/100 = 0.20$ o 20%
Y= riesgo del evento en los pacientes con el tratamiento (grupo experimental) $15/100 = 0.15$ o 15%

Cuadro II
EJEMPLO DEL EFECTO DEL PRONÓSTICO DE LA ENFERMEDAD EN EL NÚMERO
NECESARIO DE PACIENTES A TRATAR

Riesgo de muerte en los pacientes en el grupo control (riesgo basal): X	1% o 0.01	10% o 0.10
Riesgo relativo en los pacientes en el grupo experimental: Y/X	75% o 0.75	75% o 0.75
Reducción del riesgo relativo $1 - (Y/X) \times 100$ o $(X - Y)/X \times 100$	25%	25%
Riesgo de muerte en los pacientes en el grupo experimental: Y	$0.01 \times 0.75 = 0.0075$	$0.10 \times 0.75 = 0.075$
Reducción del riesgo absoluto $X - Y$	$0.01 - 0.0075 = 0.0025$	$0.10 - 0.075 = 0.025$
Número de pacientes a tratar para evitar una muerte $1 / (X - Y)$	$1 / 0.0025 = 400$	$1 / 0.025 = 40$

pacientes tratados) de tal suerte que se tendría que tratar a sólo 40 individuos para salvar una vida ($1 / 0.025 = 40$). Este ejemplo señala un elemento clave en la decisión de implantar una nueva terapia: considerar la magnitud del riesgo del desenlace adverso en los pacientes no tratados con ella. Para una misma reducción del riesgo relativo mientras mayor sea la probabilidad de padecer un evento indeseable si no se trata, mayor será el beneficio con la nueva terapia y menor el número de pacientes que tendremos que tratar para prevenir un evento. Cabe mencionar que en esta decisión debe considerarse también el costo, la factibilidad y grado de seguridad de la nueva terapia en cuestión. Si, por ejemplo, ésta conlleva un riesgo de 10% de un cierto efecto adverso por el tipo de medicamento, en la población de enfermos con un bajo riesgo de muerte, se tendrían 40 individuos sufriendo el efecto indeseable de la droga por cada vida salvada contra sólo cuatro si se le da a la población de enfermos con un mayor riesgo de muerte. Será, finalmente, el costo de la terapia, así como la gravedad y tipo de consecuencias del efecto adverso del medicamento, lo que nos haga decidir si 4, 40, 400, 4 000 o 40 000 pacientes necesarios a tratar es una cantidad importante o no.

Otro aspecto importante a evaluar en la medición de los resultados de un ensayo clínico es qué tan precisa fue la estimativa del efecto del tratamiento. La verdadera reducción del riesgo es algo que nunca llegaremos a conocer; lo más que podemos alcanzar es llegar a estimarla y el mejor estimado es el valor observado en el estudio (el llamado "estimado puntual"). Mediante el cálculo estadístico del intervalo de confianza (IC) uno puede establecer una zona de valores, alrededor de este estimado puntual, donde pudiera encontrarse el verdadero valor poblacional. Lo habitual, aunque arbitrario, es que se use el IC al 95%; es decir, se establece el intervalo que incluye al verdade-

ro valor de la reducción del riesgo relativo en 95% de las veces. Será raro (con una probabilidad de sólo 5%) que éste se encuentre más allá de los límites de este intervalo; lo que va de acuerdo con lo que convencionalmente se establece como el nivel de significancia estadística (o valor de p).

Examinemos un ejemplo. Si en un ensayo clínico se aleatorizaron 100 pacientes al grupo experimental y 100 pacientes al grupo control y se llegan a observar 15 muertes en el primer grupo y 20 en el segundo, el cálculo del estimado puntual de la reducción del riesgo relativo sería de 25%: $X = 20 / 100$ o 0.20, $Y = 15 / 100$ o 0.15, y $[1 - (Y/X)] \times 100 = [1 - 0.75] \times 100 = 25\%$. Por lo anteriormente dicho, pudiera ser que el verdadero valor de la RRR fuera significativamente menor o mayor que esta cifra de 25%, la que se obtuvo de una diferencia de tan sólo cinco muertes entre los dos grupos, y hasta pensarse que el tratamiento no fuera eficaz (una RRR de 0%) e incluso perjudicial (una RRR con un valor negativo). De hecho así es: estos resultados son consistentes tanto con una RRR de menos 38% (es decir, que a los pacientes que reciben el nuevo tratamiento tuvieran un riesgo de morir 38% mayor que los pacientes en el grupo control) y una RRR de casi 59% (es decir, que los pacientes en el grupo experimental tuvieran un riesgo menor de morir de casi 60%). En otras palabras, el intervalo de confianza al 95% (IC 95%) de este estimado de la RRR es de menos 38% a 59%, y el estudio en realidad no nos ayuda a decidir si el nuevo tratamiento es útil. Ahora, supongamos que en lugar de 100, se hubieran sorteado 1 000 pacientes por grupo y que se hubiera observado la misma proporción de desenlaces, es decir, 150 muertes en el grupo experimental ($Y = 150 / 1\ 000 = 0.15$) y 200 muertes en el grupo control ($X = 200 / 1\ 000 = 0.20$). Nuevamente, el estimado puntual de la RRR sería 25%: $[1 - (Y/X)] \times 100 = [1 - (0.15 / 0.20)] \times 100 = 25\%$. Se puede predecir que, en este ensayo con un número bastante mayor de individuos estudiados, el verdadero valor de la RRR esté más cerca del valor de 25%;

efectivamente, el IC 95% de la RRR en estos datos va de 9 a 41%.

Lo que estos ejemplos nos muestran es que mientras mayor sea el número de participantes en el ensayo clínico mayor será el número de eventos observados y mayor la certidumbre de que el valor verdadero de la RRR (o de cualquier otra medida de eficacia) está cercano al valor obtenido en el estudio. En el segundo ejemplo anterior, el valor posible más bajo de la RRR fue de 9% y el más alto, de 41%. El estimado puntual (en este caso 25%) es probablemente el que más se acerca al valor real (poblacional) de la RRR. Mientras más lejanos los valores del estimado puntual menos probable es que sean consistentes con el valor observado y aquéllos más allá de los límites del IC son valores con muy pocas posibilidades de representar la verdadera RRR, dado el estimado puntual (la RRR observada).

Siendo que mientras más grande es el tamaño de la muestra estudiada más estrecho es el IC, ¿cuál es un número suficiente de sujetos a estudiar? Cuando en un ensayo clínico se concluye que los resultados fueron positivos (es decir, que muestra que el tratamiento fue eficaz) es porque el valor inferior del IC del estimado de la RRR (el más bajo y aún consistente con los resultados del estudio) es “clínicamente importante”, es decir, lo suficientemente grande como para que el tratamiento sea prescrito a los enfermos. En esta circunstancia puede decirse que el tamaño de la muestra estudiada fue suficiente. Si, en cambio, se considera que el límite inferior de este IC no es “clínicamente importante”, entonces el estudio no puede ser considerado como definitivo, a pesar de que la diferencia entre tratamientos sea estadísticamente significativa (es decir, que excluya que la RRR sea de cero).

El IC también ayuda a interpretar los estudios con resultados “negativos”, en los que los autores concluyen que el tratamiento experimental no mostró ser mejor que el del grupo control: si el límite superior del IC muestra que la RRR pudo haber sido “clínicamente significativa” quiere decir entonces que el estudio no logró excluir un efecto importante de la nueva terapia. En el primer ejemplo anterior, el límite superior del IC fue una RRR de 59%; de tal forma que el beneficio del tratamiento pudo haber sido sustancial y se concluiría que aunque los investigadores no lograron demostrar que la terapia experimental fuera mejor que el placebo tampoco pudieron descartar un efecto importante de ella. Esto es lo que se conoce como un estudio con poca precisión (poder o sensibilidad) por haber estudiado un tamaño de muestra insuficiente.

Cabe señalar la asociación entre el IC y el valor de significancia estadística, o valor de p . Si éste es

igual o mayor a 0.05 (5%) quiere decir que el límite inferior del IC 95% de la RRR es el valor nulo (cero), o un valor negativo, y el RR, de uno, o menor; es decir, que no se logra descartar la hipótesis nula de no diferencia entre el tratamiento experimental y el placebo (o terapia en el grupo control). Conforme el valor de p disminuye (menor a 5%) el límite inferior del IC 95% de la RRR es mayor a cero y se dice que la diferencia entre tratamientos mostró ser estadísticamente significativa; es decir que, si la hipótesis nula es cierta, es muy pequeña la posibilidad de ver una diferencia de esta magnitud o mayor. La prueba estadística para el cálculo del valor de p dependerá del tipo de variable, de cómo esté expresado el evento desenlace (dicotómica, ordinal, continua) y su descripción no es motivo de este capítulo.

¿Es ético hacerlos?

Realizar un ensayo clínico controlado puede generar inquietudes de orden ético. Algunos pacientes se pueden llegar a incomodar por el hecho de ser interrogados o examinados para propósitos diferentes a los de su estricta atención médica, o bien, al saber que investigadores que no son sus médicos llegan a tener acceso a la privacidad de su expediente clínico. El diseño experimental significa que los tratamientos son asignados por un investigador, no por su médico tratante ni elegidos por el propio paciente. Cuando la asignación es mediante un sorteo o cuando hay un doble enmascaramiento (es decir, que ni el paciente ni sus médicos que lo atienden saben a qué medicamento se está sometiendo) los aspectos éticos se vuelven más complejos e incluso más importantes que los científicos.

Para que un estudio clínico experimental sea éticamente justificable debe cumplir la premisa de que, al momento del inicio del estudio, no haya evidencia de que alguno de los tratamientos ofrecidos en cada brazo (o grupo del ensayo, incluyendo el control) sea superior al(os) otro(s). No es raro que al plantearse la conducción de un estudio clínico experimental surja la inquietud de que a los participantes no se les esté ofreciendo la mejor opción terapéutica conocida; es decir, que se les está privando de una terapia experimental novedosa y superior a lo ya conocido o, por el contrario, que se les esté exponiendo a un nuevo tratamiento de dudosa utilidad y seguridad. En efecto, el estudio no sería ético si se tiene la suficiente información de que uno de los tratamientos es más eficaz (o dañino); pero si en verdad se desconoce si el tratamiento en evaluación es el óptimo, no habría justificación para dicha preocupación; de hecho, podría

argumentarse la falta de ética al estar prescribiendo abierta y rutinariamente un tratamiento cuya utilidad (y seguridad) relativa se desconocen porque se adoptó de manera prematura, antes de su evaluación formal. En este punto puede existir la dificultad de definir a partir de cuándo la evidencia es "suficiente" o "convinciente" para justificar o no la realización de un estudio experimental en humanos; desde luego que la perspectiva de un clínico (que procura el máximo beneficio de un individuo en particular) será diferente que la del investigador clínico (quien busca la "verdad" con rigor científico). Por otra parte, es habitual que en el proceso de desarrollo de un nuevo medicamento se le exija a quien lo manufactura que presente información de su eficacia obtenida en pequeños estudios no controlados (la llamada fase II) antes de encaminarse a un ensayo clínico controlado (fase III).

Aunque estos dilemas éticos pueden llegar a ser de difícil solución, en muchos centros de investigación del mundo el consentimiento informado y los comités institucionales de investigación en humanos han venido a constituir importantes salvaguardas de la ética. El consentimiento informado es requerido por la mayoría de las agencias financiadoras de proyectos de investigación experimental en seres humanos así como por las revistas científicas y habitualmente es obligatorio que sea firmado por el participante. El consentimiento informado implica que al individuo se le da la oportunidad de preguntar y enterarse de la naturaleza y detalles del estudio y se hace de su conocimiento que goza de entera libertad para decidir si participa en el estudio clínico así como de que no habrá ningún tipo de represalia (o cambio en su atención médica) si decide no hacerlo, o si abandona el estudio antes de lo programado. Sin embargo, a pesar de estos requisitos, y dada la creciente complejidad de los protocolos de los estudios experimentales, no es rara la circunstancia en la que el sujeto participante no llegue a estar completa y claramente informado. Los comités institucionales de investigación en humanos son cuerpos colegiados que suelen estar constituidos por investigadores, médicos clínicos, enfermeras, administradores, abogados y representantes de la comunidad y su tarea consiste en revisar escrupulosamente el protocolo de investigación, constatando que el planteamiento de la pregunta de investigación y el diseño sean adecuados (después de todo no es ético realizar un estudio con importantes errores metodológicos) y asegurando la preservación de la integridad y seguridad de los participantes en el estudio y que la carta de consentimiento informado esté correctamente elaborada. El comité puede también solicitar que los datos

del estudio sean analizados periódicamente por un comité externo y ante la eventualidad de encontrar una clara diferencia en la eficacia entre los tratamientos, que se interrumpa anticipadamente el estudio y se publiquen los hallazgos.

Finalmente, cabe comentar sobre la justificación de cuándo dar un placebo a los enfermos en el grupo control. Esto ocurre en dos circunstancias: una, cuando no existe tratamiento alguno que convincentemente haya mostrado ser eficaz en modificar favorablemente el curso de la enfermedad; es decir, que en una buena práctica médica en condiciones habituales un médico no prescribiría medicamento específico alguno (tal es el caso de la evaluación del zanamivir, un inhibidor de la neuraminidasa, en el tratamiento de la influenza o del uso de la warfarina en la prevención de fenómenos embólicos en la fibrilación auricular sin valvulopatía). La segunda circunstancia es cuando sí existe un tratamiento estándar efectivo pero se desea evaluar si la adición de otro medicamento ofrece una ventaja adicional. En este caso a ambos grupos de enfermos se les administra la terapia estándar, pero al experimental se le adiciona el nuevo medicamento en estudio y, en cambio, al grupo control un placebo indistinguible del nuevo medicamento (tal es el caso de la evaluación de si los ácidos grasos 3-omega, como tratamiento complementario, mejoran la funcionalidad de los enfermos con artritis reumatoide activa que además reciben medicación antiinflamatoria).

En cambio hay circunstancias en las que sería una grave falta ética dar únicamente un placebo y privar de un tratamiento activo ya ampliamente establecido y aceptado como útil a los individuos en el grupo control. Difícilmente alguien que desea probar que una cierta terapia es eficaz para mejorar el pronóstico de los pacientes sintomáticos por infección por el VIH la compararía con un grupo de enfermos en quienes se omitiera el tratamiento antirretroviral estándar en la actualidad: dos análogos nucleósidos inhibidores de la transcriptasa reversa más un inhibidor de proteinasas.

¿Cuáles son sus ventajas y desventajas?

Respecto a los estudios observacionales de cohortes, la principal ventaja que ofrecen los ensayos clínicos controlados (especialmente los que son mediante un sorteo) es que asignan la maniobra experimental (el tratamiento en prueba) independientemente de determinantes pronósticos o de la selección de la muestra; de tal suerte que disminuye el riesgo de sesgos de susceptibilidad diferencial (confusores) o de un muestreo

distorsionado. Además, el diseño experimental facilita el enmascaramiento de los individuos participantes y de sus evaluadores (doble cegamiento) ya que en un estudio observacional los tratamientos generalmente son decididos y prescritos abiertamente por los médicos tratantes.

Por otra parte, la posibilidad real de realizar los estudios experimentales puede verse seriamente comprometida por razones de índole práctica (logística) de costos y de ética. Además, en un ensayo clínico se puede llegar a una situación tal de control de variables (homogeneidad de la muestra estudiada, regulación del apego al tratamiento, atención y seguimiento del enfermo, etc.) que haga que sus resultados se alejen de la realidad del quehacer clínico cotidiano, restringiendo seriamente su extrapolación al resto de los enfermos; es decir, que en la compulsión por diseñar estudios metodológicamente impecables no se logre responder a las preguntas originales.

A pesar de esto último, los ensayos clínicos aleatorizados permanecen como el estándar de oro (o pa-

radigma) en la evaluación de la utilidad de nuevos tratamientos. A pesar de no ser infalibles y de, en ocasiones, dejar resquicios de incertidumbre, los estudios experimentales han contribuido enormemente a la evolución hacia terapias y medidas profilácticas más eficaces y seguras.

Además, el entendimiento de los principios científicos que rigen a los ensayos clínicos controlados le ha ayudado al médico a adquirir la destreza para discriminar con rapidez los estudios clínicos cuyas conclusiones son poco sustentables, por ser estudios de mal diseño y realización, de aquellos realmente válidos y aplicables a su propia práctica. En la medida en que los médicos sólo demos credibilidad a los trabajos publicados con solidez científica (es decir, que nos transformemos de lectores pasivos y puramente receptivos en evaluadores críticos) elevaremos la calidad de nuestras decisiones y el grado de beneficio que reciben nuestros enfermos.

Sesgos en estudios epidemiológicos

Mauricio Hernández-Avila, Ph.D.,⁽¹⁾ Francisco Garrido, M.C., M. en C.,⁽²⁾
Eduardo Salazar-Martínez, Dr. en C.⁽¹⁾

Un objetivo en todo estudio epidemiológico, en el que se desea conocer la frecuencia con la que ocurre un evento o estimar la asociación entre un factor de riesgo y una enfermedad, debe ser el de medir y calcular con la mayor precisión y exactitud posibles dichas determinaciones. En otras palabras, la validez del conocimiento derivado de cualquier estudio epidemiológico dependerá, en gran medida, de la ausencia de error y de la capacidad de estimar o predecir el parámetro verdadero en la población blanco. En el contexto de la epidemiología, la validez se refiere a la ausencia de sesgo o error. A lo largo del presente trabajo se hace referencia a dos tipos de validez: a) la validez interna, que se refiere principalmente a los errores cometidos durante el proceso de selección de la población de estudio, durante las mediciones que se realizan en dicha población o a errores ocasionados por la falta de comparabilidad de los grupos estudiados, y b) la validez externa, que se refiere a la capacidad del estudio de generalizar los resultados observados en la población en estudio hacia la población blanco. Es importante notar que la validez externa depende de que exista la interna, es decir, es necesario cumplir con los requisitos necesarios de validez interna para poder extrapolar los resultados, por esta razón en los diferentes estudios epidemiológicos se privilegian acciones que maximizan la validez interna, aun comprometiendo, en cierta medida, la validez externa.

Todo estudio epidemiológico está sujeto a un cierto margen de error, por lo que será muy importante conocer cuáles son sus fuentes principales y los diferentes procedimientos que pueden ser utilizados para minimizar su impacto en los resultados. Los errores se pueden clasificar en dos grandes tipos: los errores no aleatorios o sistemáticos y los errores aleatorios (no sistemáticos); ambos tipos de error, de no controlarse adecuadamente, pueden comprometer la validez del estudio. El error aleatorio (no sistemático) ocurre cuando las mediciones repetidas, ya sean en un mismo sujeto o en diferentes miembros de la población en estudio, varían de manera no predecible, mientras que el error sistemático (no aleatorio) ocurre cuando estas medidas varían de manera predecible y, por lo tanto, se tiende a sobre o subestimar el valor verdadero en mediciones repetidas. La analogía que se utiliza para describir ambos conceptos es la práctica de "tiro al blanco" donde el punto medio del objetivo es el valor verdadero en la población blanco y los "disparos" son las diferentes mediciones que se realizan en la población en estudio para estimar dicho valor verdadero. Un buen tirador cuya arma no está bien calibrada apuntará al blanco equivocado, podrá ser muy preciso (todos los disparos dan en el mismo lugar), pero ninguno de ellos da en el blanco correcto. Esto corresponde al error sistemático. Por otra parte, un tirador con mano temblorosa, pero con un arma bien calibrada, estará apuntando al blanco correcto

(1) Centro de Investigación en Salud Poblacional, Instituto Nacional de Salud Pública (INSP), Cuernavaca, Morelos, México.

(2) Centro de Investigación en Sistemas de Salud, INSP, Cuernavaca, Morelos, México.

aun cuando sus disparos no den en el punto medio del blanco seleccionado. Se puede aplicar la analogía descrita en el contexto de un estudio epidemiológico cuyo principal objetivo es determinar las concentraciones de plomo en sangre en la población general. Suponga que en este estudio participan cuatro laboratorios y un laboratorio de referencia en el cual se repite un cierto porcentaje de las determinaciones de plomo. Los resultados se muestran en la figura 1. En el panel A se muestra un laboratorio cuyas determinaciones son válidas y precisas respecto de las mediciones que realiza el laboratorio de referencia (situación ideal) ya que el valor estimado corresponde muy cercanamente al valor real; en el panel B se muestran resultados de un laboratorio cuyas mediciones son válidas pero con precisión intermedia; en el laboratorio del panel C la validez es cuestionable, pero la precisión es aceptable; mientras que en el panel D las mediciones son imprecisas y no válidas. Mientras que en los dos primeros escenarios el parámetro estimado en la población en estudio es el mismo, en ambos laboratorios se estima una media poblacional de 12.2 µg/dl, que

corresponde a la media verdadera de 12.2 µg/dl, en los paneles C y D se estimaría una media de 15.4 µg/dl, y 27.7 µg/dl, respectivamente. La utilización de laboratorios con resultados como los que se muestran en los paneles C y D para determinar las concentraciones de plomo en sangre en la población en estudio arrojaría resultados claramente erróneos al sobrestimar el valor real, lo que nos llevaría a concluir equivocadamente que existen concentraciones de plomo en sangre más elevadas que las reales.

En los estudios epidemiológicos analíticos en los que se pone a prueba una hipótesis comparando dos o más grupos de estudio, los errores también pueden ser tanto aleatorios o no sistemáticos como sistemáticos. El error sistemático o sesgo se ha definido como cualquier error diferencial –en relación con los grupos que se comparan– en que se puede incurrir durante el diseño, conducción o análisis del estudio y que invariablemente resulta en una conclusión errónea, ya sea proporcionando una estimación más baja o más alta del valor real de la asociación que existe en la población blanco.

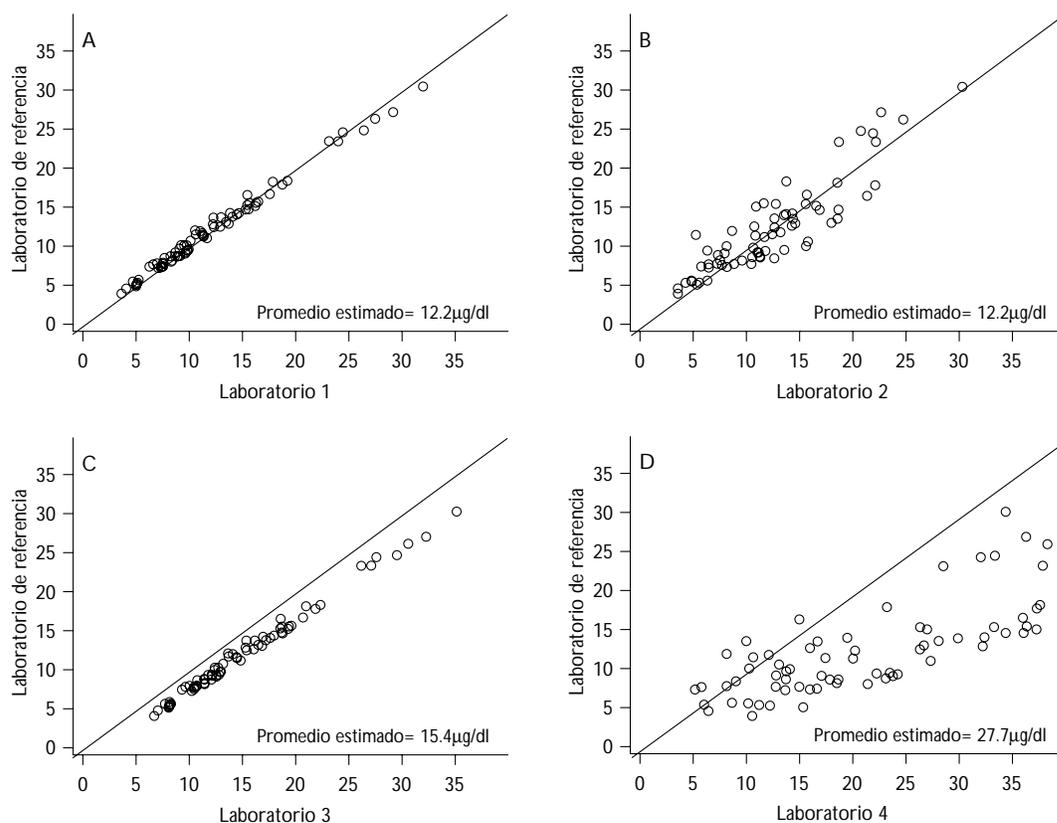


FIGURA 1. VALORES DE PLOMO EN SANGRE REPORTADOS EN CUATRO LABORATORIOS CON DIFERENTES CONDICIONES DE PRECISIÓN Y VALIDEZ

Dependiendo de la etapa del estudio en que se originan, los sesgos que interfieren con la validez interna de un estudio se han clasificado en tres grandes grupos: a) los sesgos de selección, que se refieren a los errores que se introducen durante la selección o el seguimiento de la población en estudio; b) los sesgos de información, que son errores en los que se incurre durante los procesos de medición en la población en estudio, y c) los sesgos de confusión, que se originan por la imposibilidad de asignar la exposición de manera aleatoria en los estudios observacionales y que básicamente se originan por una no comparabilidad de los grupos en estudio. Todo diseño epidemiológico no experimental, en mayor o menor medida, es susceptible de este tipo de sesgos por lo que es un imperativo para los investigadores planear adecuadamente cada una de las etapas de un estudio con el propósito de evitar o disminuir al máximo la posibilidad de incurrir en dichos errores. Sin duda, la etapa más crítica de un estudio corresponde a la del diseño ya que resulta casi imposible corregir *a posteriori* los sesgos introducidos durante esta etapa.

Sesgos de selección

Son errores sistemáticos que se introducen durante la selección o el seguimiento de la población en estudio y que propician una conclusión equivocada sobre la hipótesis en evaluación. Los errores de selección pueden ser originados por el mismo investigador o ser el resultado de relaciones complejas en la población en estudio que pueden no ser evidentes para el investigador y pasar desapercibidas. En este contexto, una posible fuente de sesgo de selección puede ser cualquier factor que influya sobre la posibilidad de los sujetos seleccionados de participar o permanecer en el estudio y que, además, esté relacionado con la exposición o con el evento en estudio.

Los sesgos de selección pueden ocurrir en cualquier estudio epidemiológico, sin embargo, ocurren con mayor frecuencia en estudios retrospectivos y, en particular, en estudios transversales o de encuesta. En los estudios de cohorte prospectivos los sesgos de selección ocurren raramente ya que el reclutamiento y selección de la población en estudio se da antes de que ocurra el evento en estudio, así que se puede suponer que la selección de los participantes se realiza de manera independiente del evento y, en general, la participación en el estudio no puede ser influida por el evento, ya que éste aún no ha ocurrido. En contraste, la permanencia de los participantes en el estudio sí puede ser determinada por el evento, cuando esto

ocurre, y es de diferente magnitud para los grupos expuesto y no expuesto, existirá la posibilidad de que los resultados se vean distorsionados por esta permanencia diferencial. Por esta razón, se recomienda maximizar las tasas de permanencia y seguimiento en los estudios de cohorte.

En los estudios retrospectivos los sesgos de selección pueden ocurrir cuando los participantes potenciales o los investigadores conocen la condición de exposición y/o de enfermedad, y este conocimiento influye diferencialmente la participación en el estudio. Suponga que se llevará a cabo un estudio de una cohorte retrospectiva y que el resultado de este estudio servirá para fijar una posible compensación económica si se logra comprobar que los sujetos que recibieron cierta exposición están en mayor riesgo de desarrollar la enfermedad en estudio. Si esta información fuera del conocimiento de los posibles participantes podría condicionar que los sujetos que se saben expuestos y que desarrollaron la enfermedad en estudio participaran más frecuentemente que aquellos que desarrollaron la enfermedad y que se identifican como no expuestos. La participación diferencial antes mencionada se reflejaría en un déficit de eventos en el grupo de no expuestos, ya que los sujetos no expuestos que desarrollaron la enfermedad no tendrían la misma motivación para participar en el estudio y participarían en menor proporción. Esta participación diferencial condicionaría una subestimación de la frecuencia del evento en el grupo no expuesto y una sobrestimación de la diferencia real.

Los estudios de casos y controles son particularmente susceptibles a este tipo de sesgo ya que en la mayoría de sus aplicaciones se trata de estudios retrospectivos. En el contexto de este tipo de estudios, cualquier factor que influya sobre la probabilidad de selección, ya sea como caso o control, y que a su vez esté relacionado con la exposición en estudio, será una posible fuente de sesgo de selección. Este tipo de sesgo se ha propuesto como una de las explicaciones a la asociación reportada entre el consumo de café y el cáncer de páncreas por McMahon y colaboradores.¹ En este estudio los casos de cáncer de páncreas se seleccionaron de diferentes hospitales, mientras que los controles fueron seleccionados del mismo grupo médico de donde se había originado el diagnóstico de los casos. Al elegir este mecanismo de selección de controles los investigadores incluyeron como controles a diferentes sujetos con patología gastrointestinal, dado que la mayoría de los casos de cáncer de páncreas se habían notificado por el servicio de gastroenterología. Así, los controles representan un sector

de la población con un consumo menos frecuente de café, lo que pudo haber condicionado una subestimación de la prevalencia de exposición en el grupo control y, por lo tanto, una asociación espuria entre el consumo de café y el cáncer de páncreas.

La detección diferencial es un tipo particular de sesgo de selección, se origina cuando la prueba diagnóstica para detectar el evento se realiza con mayor frecuencia en el grupo expuesto. Un ejemplo de este sesgo se presentó en un estudio de casos y controles en el que se observó una fuerte asociación entre el uso de estrógenos de remplazo y el cáncer de endometrio.² Los casos en este estudio no representaban adecuadamente los casos originados en la población base, ya que las mujeres con síntomas sugestivos de cáncer de endometrio, y que eran usuarias de estrógenos, eran hospitalizadas para diagnóstico con mayor frecuencia que las mujeres con los mismos síntomas pero que no eran usuarias de estrógenos. Esta conducta de los médicos generó una sobreselección de casos expuestos a estrógenos, lo que ocasionó una asociación falsa positiva entre el uso de estrógenos y el riesgo de cáncer de endometrio en el estudio.

La no respuesta, por parte de los participantes en un estudio, puede también introducir sesgo de selección, siempre y cuando esté relacionada con la exposición o el evento en estudio, es decir, que la tasa de participación sea diferente para expuestos y no expuestos en los estudios de cohorte, o para casos y controles en el contexto de un estudio de casos y controles. Sin embargo, no siempre la falta de respuesta se asocia con un sesgo de este tipo. Por ejemplo, si en un estudio de casos y controles 10% de los casos elegidos y 30% de los controles rehusan participar no se incurrirá en sesgo de selección, a menos que la proporción de expuestos en los casos y en los controles que no aceptaron participar sea diferente a la del grupo respectivo. La no respuesta es un problema que tiene mayor relevancia para la validez externa.

Los estudios transversales, y de casos y controles que se basan en casos existentes (prevalentes), presentan importantes limitaciones relacionadas con los sesgos de selección, en particular cuando la enfermedad en estudio tiene una alta letalidad cercana al diagnóstico inicial ya que los casos existentes tienden a sobrerrepresentar a los sujetos con cursos más benignos de la enfermedad. Si el factor en estudio se asocia con la letalidad, la medida de efecto derivada de un estudio de prevalencia será sesgada, dado que los casos en estudio corresponden a sobrevivientes de la enfermedad, por lo que representarán desproporcionadamente a los casos no expuestos.

Sesgos de información

El sesgo de información se refiere a los errores que se introducen durante la medición de la exposición, de los eventos u otras covariables en la población en estudio, que se presentan de manera diferencial entre los grupos que se comparan, y que ocasionan una conclusión errónea respecto de la hipótesis que se investiga. Una posible fuente de sesgo de medición puede ser cualquier factor que influya de manera diferencial sobre la calidad de las mediciones que se realizan en los grupos expuesto y no expuesto en el contexto de los estudios de cohorte o entre los casos y controles en el contexto de los estudios de casos y controles.

Es importante mencionar que aunque prácticamente no existen procedimientos libres de error de medición, no todos los errores de medición son fuente de sesgo de información. Es conveniente recordar que los errores de medición pueden ser no diferenciales (aleatorios) cuando el grado de error del instrumento o técnica empleada es el mismo para los grupos que se comparan y diferenciales (no aleatorios) cuando el grado de error es diferente para los grupos estudiados, el sesgo de información se refiere particularmente a este último tipo. Para comprender mejor la diferencia entre ambos errores de medición se analiza el siguiente ejemplo. En un estudio hipotético de casos y controles para evaluar la asociación entre tabaquismo e infarto agudo del miocardio, los casos se identificaron en las salas de urgencias, del Instituto Mexicano del Seguro Social al momento de su ingreso y los controles fueron seleccionados al azar entre los derechohabientes que son vecinos de cada caso. Suponga que, en un primer escenario, la exposición al tabaco se evalúa determinando la presencia de un marcador biológico de exposición que se mide en sangre. En este primer escenario, puede existir cierto grado de error en las determinaciones del biomarcador en sangre, sin embargo es posible suponer que el error es similar para los dos grupos, por lo que se considera no diferencial o aleatorio. En contraste, si la exposición al tabaco se hubiera evaluado mediante un cuestionario, la calidad de la información dependería, en parte, de la memoria de los participantes. Si los casos, dado que sufrieron el evento, tuvieran un estímulo mayor para recordar o participar entonces la calidad de la información sería mejor en este grupo que la que se podría obtener en el grupo control, situación que introduciría un error diferencial o no aleatorio. En general, el impacto de este último tipo de error es difícil de predecir ya que puede subestimar o sobrestimar la asociación real, a diferencia del error

aleatorio que en general tiende a subestimar las asociaciones reales.

En el cuadro I se ofrece una clasificación general de las diferentes fuentes de error de medición. Es importante hacer notar que durante el diseño de un estudio se deben establecer todas aquellas características que se habrán de medir para responder una pregunta de investigación. Una de las dificultades con las que se topa el investigador en esta etapa se deriva de la brecha o diferencia que se establece entre una entidad conceptual de una variable y su definición empírica que es, finalmente, la forma en que será medida. Esta diferencia estriba en la imposibilidad, la mayor parte de las ocasiones, de medir directamente la característica de interés. Por ejemplo, la exposición involuntaria al humo del tabaco podría ser definida teóricamente como “la exposición de las células del tejido pulmonar a agentes carcinogénicos que resulta de la inhalación del humo del tabaco proveniente de las emisiones de fumadores activos”, en la práctica esta definición teórica se tendría que medir mediante la utilización de algún indicador cercano de dicha condición, tal como “tiempo que pasa el sujeto de estudio junto a personas fumadoras”. Estas definiciones empíricas son utilizadas y plasmadas por los epidemiólogos y por investigadores de otras disciplinas, en los instrumentos de medición que se elaboran *ex profeso* para un estudio y que pueden ser medidas ya sea por un cuestionario o un biomarcador. En este último ejemplo, la exposición involuntaria se podría también medir mediante la determinación de cotinina (un metabolito de la nicotina) en la sangre de los sujetos en estudio. Mientras que las estimaciones de exposición realizadas mediante un cuestionario estarán sujetas a diferentes tipos de error como lo serían la memoria de los participantes y la comprensión de las

preguntas, en el caso del biomarcador, serán fuentes de error los procedimientos utilizados para obtener la muestra, el transporte y conservación de la misma, así como los diferentes procedimientos de laboratorio que se necesiten utilizar para su medición.

Para ilustrar el impacto que pueden tener los diferentes tipos de error de información en la estimación de las medidas de efecto, es necesario mencionar brevemente los conceptos de sensibilidad y especificidad de los instrumentos de medición que se utilizan. La sensibilidad de un cuestionario o prueba diagnóstica mide la habilidad del instrumento para clasificar correctamente a los sujetos que tienen verdaderamente la condición en estudio. Así, un cuestionario para identificar fumadores involuntarios que tiene una sensibilidad de 90%, podrá identificar correctamente a 90% de los sujetos que están expuestos al humo del tabaco, mientras que 10% serán clasificados erróneamente como no fumadores involuntarios. En contraste, la especificidad indica la habilidad del instrumento para clasificar correctamente a los sujetos que no tienen la condición de estudio. Para este ejemplo, una especificidad de 99% indicaría que se clasificaría adecuadamente a 99% de los sujetos y que restaría un 1% que sería clasificado como expuesto involuntariamente al tabaco, cuando en realidad no lo está. La sensibilidad y especificidad de nuestros instrumentos puede ser evaluada en estudios de validación cuando se cuenta con un estándar de oro, el cuadro II indica cómo se podrían estimar estos parámetros.

Los sujetos de estudio se pueden clasificar erróneamente con relación a la exposición o al evento, como se mencionó, el error puede ser de dos tipos: no diferencial y diferencial.

Los estudios epidemiológicos incurren en un error de clasificación no diferencial de la exposición cuando ésta se realiza por igual en los grupos que se

Cuadro I
CLASIFICACIÓN DEL ERROR DE MEDICIÓN SEGÚN SU ORIGEN

- a) El observador
- b) Sistema de medición
- c) Los sujetos de estudio
 - Memoria
 - Entrenamiento
 - Fatiga
- d) El instrumento
- e) Errores en las variables proxy
- f) El procesamiento de datos
 - Errores de codificación
 - Formulación errónea de modelos estadísticos
- g) Errores que dependen del tiempo

Cuadro II
ESTIMACIÓN DE SENSIBILIDAD Y ESPECIFICIDAD PARA UNA VARIABLE DICOTÓMICA

Clasificación de la exposición o de la enfermedad por el instrumento de medición	Clasificación verdadera de la exposición o de la enfermedad	
	Sí	No
Sí	a	b
No	c	d

Sensibilidad = $\frac{a}{a+c}$; Especificidad = $\frac{d}{b+d}$

comparan. Es decir, la condición de enfermedad de los sujetos de estudio no influye en la clasificación de los mismos respecto de la exposición. En esta situación, siempre y cuando el error no sea de gran magnitud, la medida de efecto estimada con error aleatorio tiende a subestimar la asociación real en la población blanco, por lo que se podría suponer que el valor estimado en la población en estudio, bajo el supuesto de que el error es aleatorio, tiende o se aproxima al valor de no asociación. La dificultad para medir correctamente la exposición involuntaria al humo del tabaco puede servir para ilustrar este tipo de error. Dado que no es factible medir directamente la cantidad de humo que inhala un no fumador se podría utilizar un monitor personal que registre en forma acumulada la concentración de nicotina ambiental a la que el sujeto estuvo expuesto durante el tiempo relevante para el evento en estudio. Supongamos un estudio longitudinal para examinar la asociación entre enfermedad respiratoria y exposición involuntaria al humo del tabaco. En este caso, el error de clasificación –no diferencial– en la medición de la exposición estaría dado por la habilidad del sujeto de llevar consigo el monitor, del tipo de ropa que utilice y de la conservación y trato que dé al instrumento de medición. En este caso se podría suponer que el grado de error será el mismo para los sujetos con y sin exposición al humo de tabaco y que el impacto de dicho error será el de subestimar cualquier asociación real que exista en la población blanco.

Otro ejemplo que permite ilustrar el contexto en el que ocurre el error de medición aleatorio o no diferencial es cuando la exposición en estudio no considera el periodo de riesgo relevante en la ocurrencia de una enfermedad. Si se estudia una enfermedad que tiene un periodo de latencia prolongado, como es el caso de la gran mayoría de las enfermedades crónicas, y el factor de exposición que se cree asociado con dicha enfermedad se mide en una sola ocasión, sin tomar en consideración la variabilidad de dicho factor durante el tiempo, se incurrirá en un error de clasificación no diferencial. De manera más explícita, se puede decir que si se desea medir la asociación entre el consumo de colesterol y enfermedad coronaria para lo cual medimos la ingesta de colesterol en un solo día, no se puede asegurar que se haya medido correctamente la exposición dado que la ingesta varía considerablemente a través del tiempo en un mismo sujeto y, por lo tanto, una sola medición no reflejaría adecuadamente la ingesta de colesterol durante el periodo de riesgo relevante.

A continuación se ofrece un ejemplo hipotético para ilustrar el efecto del error aleatorio o no diferen-

cial. En el cuadro III se resumen los datos de un estudio de seguimiento por 12 años en hombres de 45-64 años ($n=7\,221$) en el que se relacionó el ejercicio físico con el riesgo de enfermedad coronaria. Los expuestos fueron quienes no hacían ejercicio físico o lo hacían moderadamente. Asumamos que la medida de efecto en este primer escenario se estima sin error aleatorio y corresponde a una razón de incidencia acumulada de 1.41.

Supongamos ahora que la determinación del evento se realiza con una prueba diagnóstica que tiene una sensibilidad de 0.80 y especificidad de 1.0 y que la prueba se aplica de igual manera en expuestos y en no expuestos. Los resultados se presentan en el cuadro III-A, donde se puede observar que existe un subestimación del riesgo real, la razón de incidencia acumulada disminuye de 1.41 a 1.38. Sin embargo, si se supone una especificidad de 98%, entonces el error es de mayor impacto ya que la razón de incidencia estimada bajo este supuesto sería de 1.28.

Cuadro III
RELACIÓN ENTRE EJERCICIO FÍSICO Y RIESGO
DE ENFERMEDAD CORONARIA,
ESTIMADA SIN ERROR ALEATORIO

Evento coronario	Realiza ejercicio físico		Total
	No	Sí	
Sí	299	107	406
No	4 503	2 312	6 815
Total	4 802	2 419	7 221

Razón de incidencia acumulada= $299/4\,802 + 107/2\,419 = 1.41$

Cuadro III-A
RELACIÓN ENTRE EJERCICIO FÍSICO Y RIESGO
DE ENFERMEDAD CORONARIA, ESTIMADA
CON UNA PRUEBA DIAGNÓSTICA CUYA SENSIBILIDAD
ES DE 0.80 Y ESPECIFICIDAD DE 1.0

Enfermedad	Realiza ejercicio físico		Total
	No	Sí	
Sí	239	87	406
No	4 563	2 332	6 815
Total	4 802	2 419	7 221

Razón de incidencia acumulada= $239/4\,802 + 87/2\,419 = 1.38$

239 se deriva de 299 casos de los cuales únicamente 80% (299×0.80) fueron clasificados correctamente como eventos
4 563 corresponde a 4 503 sujetos sin eventos y a 60 sujetos con evento que fueron clasificados erróneamente como sanos

Como se mencionó, cuando el error de medición se realiza de manera diferencial entre los grupos que se comparan, entonces se introduce un sesgo de información que, a diferencia del error aleatorio que condiciona una subestimación, puede introducir error en cualquier dirección, condicionando ya sea una sobre o una subestimación de la asociación real. Existen varias situaciones que se pueden asociar con este sesgo. En la etapa de diseño del estudio se puede favorecer la introducción del sesgo cuando se decide recoger la información sobre la exposición después que se han identificado los casos de la enfermedad de interés y la presencia de la enfermedad modifica el reporte de la exposición (exagerándola o minimizándola). En el caso del estudio de exposición involuntaria al humo del tabaco antes referido, si los investigadores utilizaran un cuestionario como instrumento para la evaluación de tipo retrospectivo, no es difícil suponer que las madres cuyos hijos tuvieron problemas respiratorios tratarán de recordar con mayor detalle y exactitud la exposición al humo del tabaco comparadas con aquellas madres cuyos hijos permanecieron sanos durante el periodo de estudio; o que las madres fumadoras cuyos hijos tuvieron enfermedad respiratoria informarían una menor exposición como un mecanismo compensatorio de culpa. En ambas situaciones es claro que se puede incurrir en sesgos de información diferencial.

El sesgo de clasificación diferencial de la exposición ocurre también con frecuencia durante la fase de campo. El sesgo del entrevistador, un caso de sesgo de información, se puede introducir cuando el personal de campo conoce la condición de enfermo o de no enfermo del entrevistado y si, además, tiene conocimiento de las hipótesis u objetivos del estudio. Ante este hecho, el entrevistador puede hacer más énfasis en aquellas preguntas que miden la exposición o bien inducir las respuestas de los entrevistados mediante lenguaje corporal o la aportación de información adicional sobre dichas preguntas. En este sentido es recomendable realizar un buen entrenamiento de los entrevistadores y mantenerlos ciegos tanto en lo que respecta a la hipótesis del estudio como a la condición de caso-control o de expuesto-no expuesto de los participantes del estudio.

El uso de informantes sustitutos (proxy) puede también dar lugar a un sesgo de clasificación diferencial de la exposición. Como se mencionó, con cierta frecuencia los estudios epidemiológicos deben recurrir a informantes que no son los sujetos de estudio ya sea por alguna incapacidad de éstos o bien porque ya fallecieron. Dependiendo del tipo de relación que tiene o tuvo el informante sustituto con el

sujeto de estudio, las respuestas que éstos proporcionan sobre la exposición de interés pueden oscilar entre la negación o la exageración. Es probable que si la exposición se refiere a una característica considerada negativa y el informante sustituto es la esposa(o) del sujeto de estudio, las respuestas que se proporcionen se pueden modificar con base en la relación existente entre ambos.

Para ilustrar el impacto del error diferencial, supongamos ahora que la relación descrita en el ejemplo anterior entre ejercicio y enfermedad coronaria se evalúa mediante un estudio de casos y controles. En este estudio se identificarían 406 casos y para cada caso se seleccionarían dos controles, los resultados hipotéticos se presentan en el cuadro IV. La razón de momios (RM) estimada del estudio es de 1.40.

Ahora, supongamos que los casos, dado que sufrieron el evento coronario, hacen un mejor esfuerzo por recordar y reportar el patrón de ejercicio que realizaban y que la sensibilidad con que informan la exposición es de 95%, mientras que los controles reportan con una sensibilidad de 80%. Los resultados se presentan en el cuadro IV-A. En este escenario, el sesgo introducido por la menor calidad de reporte en los

Cuadro IV
RELACIÓN ENTRE EJERCICIO FÍSICO Y RIESGO DE ENFERMEDAD CORONARIA, ESTIMADA SIN ERROR DIFERENCIAL

Exposición	Hace ejercicio		Total
	No	Sí	
Casos	299	107	406
Controles	540	272	812
Total	839	379	1 218

Razón de momios= $299 \times 272 / 540 \times 107 = 1.40$

Cuadro IV-A
RELACIÓN ENTRE EJERCICIO FÍSICO Y RIESGO DE ENFERMEDAD CORONARIA, ESTIMADA A PARTIR DE UN INFORME DE LA EXPOSICIÓN CON SENSIBILIDAD DE 95% EN LOS CASOS Y DE 80% EN LOS CONTROLES

	Ejercicios	
	No	Sí
Casos	304	102
Controles	594	218

Sensibilidad en casos= 0.95 en controles= 0.80
Especificidad en casos= 1.0 en controles= 1.0

Razón de momios= $304 \times 218 / 594 \times 102 = 1.09$

casos condiciona una subestimación de la razón de momios.

Supongamos ahora que los casos tienden a subestimar el reporte de ejercicio y que la sensibilidad en este grupo es de 70% mientras que en el grupo control el cuestionario utilizado clasifica adecuadamente la exposición en 90% de los sujetos. Bajo los supuestos en este escenario, el sesgo diferencial de reporte condicionaría una sobrestimación del efecto de no hacer ejercicio (cuadro IV-B). Como se puede apreciar, el efecto puede condicionar desviaciones hacia cualquier dirección, por esta razón es necesario evitar en la medida de lo posible introducir errores diferenciales.

Sesgos de confusión

Todos los resultados derivados de estudios observacionales están potencialmente influenciados por este tipo de sesgo. El sesgo de confusión puede resultar en una sobre o subestimación de la asociación real. Existe sesgo de confusión cuando observamos una asociación no causal entre la exposición y el evento en estudio o cuando no observamos una asociación real entre la exposición y el evento en estudio por la acción de una tercera variable que no es controlada. Esta(s) variable(s) se denomina(n) factor(es) de confusión o confusor(es). Los resultados de un estudio estarán confundidos cuando los resultados obtenidos en la población en estudio apoyan una conclusión falsa o espuria sobre la hipótesis en evaluación, debido a la influencia de otras variables, que no fueron controladas adecuadamente ya sea durante la fase de diseño o de análisis. En este contexto, son fuente posible de sesgo de confusión cualquier variable asociada con la exposición que, además, esté causalmente asociada con el

evento en estudio y que se encuentre distribuida de manera diferencial entre los grupos que se comparan, ya sea entre expuestos y no expuestos en el contexto de los estudios de cohorte o entre casos y controles en el ámbito de los estudios de casos y controles.

En los estudios observacionales el sesgo de confusión se puede entender como un problema de comparabilidad cuyo origen está ligado a la imposibilidad de realizar una asignación aleatoria de la exposición en los sujetos de estudio. El objetivo de la asignación al azar de los tratamientos (de la exposición) en los estudios experimentales es lograr la formación de grupos homogéneos en lo que se refiere a todas las características que puedan influir en el riesgo de desarrollar el evento (edad, sexo, masa corporal u otras características que no se puedan medir), lo que se busca lograr es que los grupos sean similares en todo excepto en la exposición que se busca evaluar.

Analicemos el siguiente ejemplo, cuando Doll y Hill publicaron sus primeros resultados derivados del seguimiento de cerca de 40 000 médicos y reportaron que el cáncer de pulmón era considerablemente más frecuente en los fumadores,^{3,4} las críticas no se hicieron esperar. Se argumentó que la asociación observada por estos investigadores entre el tabaquismo y el cáncer de pulmón se debía muy probablemente a la acción de una tercera variable, a una susceptibilidad genética que predisponía tanto al cáncer de pulmón como al gusto por el tabaco, con este argumento se descartaba el posible efecto carcinogénico del humo de tabaco. La crítica anterior implica que el gusto por fumar no se da al azar sino que coexiste un factor genético y que este factor es la verdadera causa de cáncer de pulmón. Para poder separar el efecto del cigarrillo del de la susceptibilidad genética tendríamos que realizar un experimento, se podrían seleccionar 40 000 individuos y asignarlos aleatoriamente a dos tratamientos experimentales: fumadores (expuesto) y no fumadores (no expuesto). Al asignarlos al azar se puede suponer que la distribución del factor genético que causa cáncer de pulmón en los grupos formados mediante la aleatorización será la misma, es decir, si la aleatorización se realizó de manera correcta –en promedio– el porcentaje de sujetos genéticamente susceptibles al cáncer de pulmón sería la misma en ambos grupos, por lo tanto, si observáramos diferencias en la ocurrencia de cáncer de pulmón entre los fumadores y no fumadores ésta no podría ser atribuida al factor genético ya que éste afectaría de igual manera a los dos grupos. Sin embargo, es claro que este tipo de experimento sería imposible de realizar.

A diferencia de los estudios experimentales, en que los sujetos en estudio son asignados al azar a los

Cuadro IV-B
RELACIÓN ENTRE EJERCICIO FÍSICO Y RIESGO DE ENFERMEDAD CORONARIA, ESTIMADA A PARTIR DE UN INFORME DE LA EXPOSICIÓN CON SENSIBILIDAD DE 70% EN LOS CASOS Y DE 90% EN LOS CONTROLES

	Ejercicios	
	No	Si
Casos	331	75
Controles	563	249

Sensibilidad en casos= 0.70 en controles= 0.90

Especificidad en casos= 1.0 en controles= 1.0

Razón de momios= 1.95

grupos experimentales y, por lo tanto, los posibles confusores quedan igualmente distribuidos entre los grupos en contraste, en los estudios observacionales los sujetos reciben la exposición por muy diferentes motivos que dependen importantemente de patrones culturales y socioeconómicos. Entonces ¿qué es una variable confusora o confusor? Se trata de variables que están asociadas con el evento en estudio de manera causal, es decir, son factores de riesgo para el evento en estudio, además esta asociación debe ser por un mecanismo causal diferente al que actúa la exposición en estudio y debe ser independiente de esta misma. Esto último quiere decir que el confusor debe ser un factor de riesgo tanto para el grupo expuesto como para el no expuesto. Los sesgos de confusión en estudios de cohorte ocurren principalmente cuando los grupos expuesto y no expuesto no son comparables. Por ejemplo, si fuera de interés evaluar en un estudio de cohorte el efecto del consumo de café sobre el riesgo de infarto del miocardio y se establecieran dos grupos de contraste, uno conformado por bebedores de café y otro por no bebedores de café, se tendría que considerar el tabaquismo como un posible confusor ya que está asociado con el consumo de café y es un factor de riesgo para infarto de miocardio. En este estudio hipotético ignorar las diferencias en el consumo de tabaco podría hacernos concluir que existe una falsa asociación positiva entre el consumo de café y el riesgo de infarto del miocardio. En el contexto de los estudios de casos y controles ocurre cuando hay diferencias entre casos y controles en relación con una tercera variable, lo que puede ocurrir simplemente por la existencia de confusión en la población fuente o ser generada por el mismo proceso de selección.

¿Qué podemos hacer para prevenir la ocurrencia de sesgos de confusión? En el cuadro V se listan los diferentes procedimientos tanto en la fase de diseño como en la de análisis que nos permiten prevenir o corregir el efecto del sesgo de confusión, como se puede deducir de los diferentes procedimientos utilizados, el

Cuadro V
**ESTRATEGIAS PARA EL CONTROL DE LA CONFUSIÓN EN
 LOS ESTUDIOS EPIDEMIOLÓGICOS**

Fase	Estrategia	Efecto
Diseño	Aleatorización	Permite que las variables se distribuyan similarmente en los grupos de estudio haciéndolos comparables en todo, excepto en la variable de exposición
	Restricción	Limita la participación en el estudio a sujetos que son similares respecto a la variable de confusión
	Pareamiento	Iguala en el proceso de selección a los grupos de comparación con relación a los factores de confusión
Análisis	Estandarización	Permite comparar los grupos de estudio si la distribución del confusor fuera la misma en ambos
	Estratificación	Estima la medida de efecto en subgrupos que son similares con relación a los factores de confusión
	Modelos multivariados	Estima el efecto de la exposición, manteniendo constantes los valores del factor confusor

común denominador de todos es hacer los grupos de contraste lo más comparable posible con relación a variables externas, el objetivo es lograr que la única diferencia entre los grupos sea la característica en estudio.

Referencias

1. MacMahon B, Yen S, Trichopoulos D *et al.* Coffee and cancer of the pancreas. *N Engl J Med* 1981; 304:630-633.
2. Feinstein HR, Horowitz RI. A critique of the statistical evidence associating estrogens with endometrial cancer. *Cancer Res* 1978;38: 4001-4005.
3. Doll R, Hill AB. Mortality in relation to smoking: Ten years' observations of British doctors. *BMJ* 1964;1:1399-1410.
4. Doll R, Hill AB. Mortality in Relation to Smoking: Ten years' observations of British doctors (Concluded). *BMJ* 1964;1:1460-1467.