



## INTRODUCCIÓN BÁSICA A LA INVESTIGACIÓN EPIDEMIOLÓGICA

### A BASIC INTRODUCTION TO EPIDEMIOLOGIC RESEARCH

Paternina Caicedo Ángel<sup>1</sup>

Correspondencia: angel.paternina@gmail.com

Recibido para evaluación: octubre – 01 – 2010 – Aceptado para publicación: octubre – 30 – 2010

### RESUMEN

La epidemiología, ciencia que da un marco conceptual a la evidencia científica en la salud, ha tenido una rápida evolución a lo largo del siglo XX y comienzos del siglo XXI. El entendimiento de los diferentes tipos de diseño epidemiológico, ha sido uno de los principales avances en la medicina, que ha permitido pasar de causas esotéricas y/o divinas de la enfermedad, a una ciencia basada en hechos comprobables. El exceso de evidencia epidemiológica contradictoria hace imperante un conocimiento adecuado de los mecanismos para crear esta evidencia y la evaluación crítica de su importancia.

**Rev.cienc.biomed. 2010: 1(2): 246 - 253**

### PALABRAS CLAVES

Epidemiología. Estudios de casos y controles. Estudios de cohorte. Sesgos.

### SUMMARY

*Epidemiology is the science that gives a framework to health-related scientific evidence. It has had a rapid evolution in the XX century and beginning of the XXI century. The understanding of the different types of epidemiologic research has been one of the most important advances in medicine, and it has allowed the pass from divine and/or holistic cases of disease, to a science based in proved facts. The excess of current contradictory epidemiological evidence makes imperative an adequate knowledge of the mechanism to create this evidence and the critic evaluation of its importance.*

### KEYWORDS

*Epidemiology. Case - control studies. Cohort studies. Bias.*

<sup>1</sup> Médico. Estudiante de Maestría en Epidemiología Clínica. Facultad de Medicina. Universidad de Cartagena.

## INTRODUCCIÓN

La epidemiología estudia “la distribución y determinantes de eventos o estados relacionados con la salud en poblaciones específicas y su aplicación para el control de problemas de salud” (1).

A pesar de ser la base de la salud pública, del control, prevención y tratamiento clínico de enfermedades, médicos y profesionales de la salud suelen no darse cuenta de su importancia. La epidemiología y la salud pública, han sido los pilares de la medicina en la última mitad de siglo.

Desde la predicción del riesgo cardiovascular en un paciente, con todos los factores de riesgo que ello implica, hasta el tratamiento médico para el dolor más leve, deben pasar por los estudios epidemiológicos. Ha sido tan importante y preponderante el desarrollo conceptual, que ya no se concibe nuevo conocimiento en salud poblacional, sin que haya pasado por algún tipo de evaluación epidemiológica.

Por esto es importante que los profesionales de la salud conozcan diferentes métodos de valoración epidemiológica que permitan de forma crítica, evaluar estudios biomédicos y emplear los resultados en sus prácticas clínicas.

## CONSIDERACIONES ACERCA DE LA CAUSALIDAD

La causa de una enfermedad es “todo acontecimiento, condición o característica que juega un papel esencial en producir su ocurrencia” (2). En 1965, Sir Austin Bradford Hill (1897-1991) postuló unas consideraciones -mal llamadas criterios por algunos textos-, en las cuales se delinearon puntos a seguir para una apropiada definición de causalidad (3, 4, 5, 6, 7). Sin embargo, estas consideraciones no son necesarias o suficientes para definir asertivamente la presencia o no de causa (2, 3, 8, 9) Tabla Nº 1.

Kenneth Rothman (2) en su libro clásico de epidemiología, delinea el modelo de causalidad e inferencia casual que mejor se ajusta al conocimiento actual acerca de

enfermedades. Introduce el modelo de causa suficiente y causas componentes.

**TABLA 1. CONSIDERACIONES DE BRADFORD-HILL PARA ESTABLECER UNA RELACIÓN CAUSAL**

CONSIDERACIÓN	COMENTARIO
<b>Fuerza de una asociación:</b> asociaciones más fuertes son más probable causa que asociaciones débiles	Algunas asociaciones pueden tener factores de confusión que aumenten el efecto.
<b>Consistencia:</b> observación repetida de la asociación en poblaciones diferentes.	Algunas poblaciones pueden o no presentar el efecto, sin que deje de ser causal en otras.*
<b>Especificidad:</b> una causa conduce a un efecto único, ó una causa tiene un efecto único.	Un factor causal puede serlo para varias enfermedades, y una enfermedad puede tener diferentes causas.
<b>Temporalidad:</b> la causa debe preceder al efecto.	Es la consideración más fuerte, y la única que puede ser un criterio de causalidad.
<b>Gradiente biológico:</b> presencia de dosis-respuesta en la relación causa-efecto.	No siempre se da. Una infección se presenta comúnmente después de un umbral de entrada, después del cual no importa la dosis, la respuesta será parecida.
<b>Plausibilidad:</b> se refiere a la plausibilidad biológica de la hipótesis.	Algunas veces, no se tiene la información suficiente para definir plausibilidad.
<b>Coherencia:</b> una relación causa-efecto debe asemejarse a lo que se conoce de la enfermedad	Si se conoce poco de la enfermedad, este criterio no sirve.
<b>Evidencia experimental:</b> evidencia de experimentos en humanos y/o animales confirmando la asociación.	Esta información es pocas veces obtenible de poblaciones humanas.
<b>Analogía:</b> relación de la causa-efecto con otros causa-efecto.	Falta de imaginación del investigador es lo único que limitaría esta consideración.

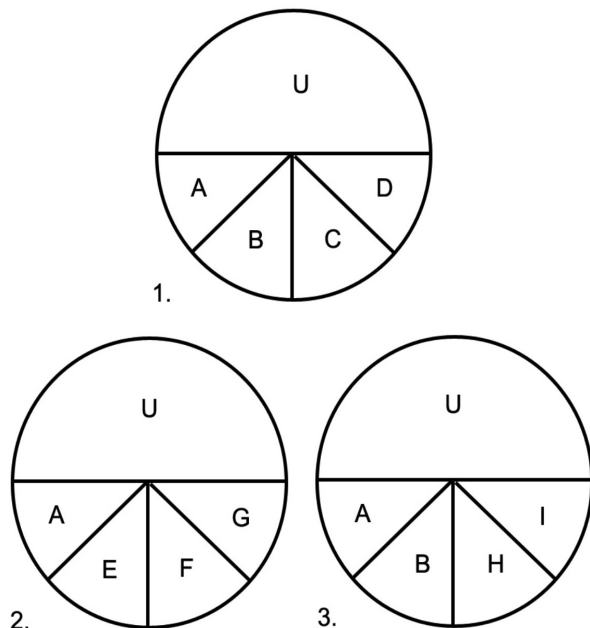
\* Por ejemplo, transfusiones con virus de inmunodeficiencia humana causan infección, sin embargo, esto no ocurre siempre (2)

En un modelo de enfermedad, una causa suficiente es un grupo de condiciones mínimas que son suficientes para producirla. Cada enfermedad puede tener varios modelos, y si

un componente causal se presenta en todos los modelos de enfermedad, se le llama causa necesaria. Por ejemplo, en el modelo de enfermedad de hepatitis B (Figura N° 1), el virus es una causa necesaria para la presentación de la enfermedad (círculo 1, 2 y 3, torta A), pero esta no se presenta sin la causa suficiente, es decir, sin presentar inyección por aguja contaminada (círculo 1, torta C), transfusiones contaminadas con el virus (círculo 2, torta E), o contacto sexual con una pareja contaminada (círculo 3, torta D). Estos factores no son comunes a todos los modelos, por lo que se denominan causa suficiente, y pueden ser por ejemplo, el uso de drogas intravenosas 'recreativas', promiscuidad y factores conocidos para el desarrollo de hepatitis B.

**Figura No. 1.**  
**CAUSA SUFICIENTE Y CAUSA**  
**NECESARIA EN**  
**INFECCIÓN POR HEPATITIS B**

MODELO TEORICO DE LA INFECCIÓN POR VIRUS  
DE LA HEPATITIS B



A= Virus de la hepatitis B. B= Uso de drogas intravenosas.  
C= Inyección por aguja contaminada. D= Tatuajes.  
E= Transfusión contaminada. F= Drepanocitosis. G=  
Viajes de zonas endémicas. H= Promiscuidad. I=  
Contacto sexual con pareja contaminada. U= Factores  
desconocidos.

El componente U (ver Figura 1, torta U), está presente en todos los modelos en donde no se conozcan todos los factores de riesgo para el desarrollo de una enfermedad, debido a las limitaciones del conocimiento humano, está presente en todos los modelos.

Este modelo de causa suficiente y causas componentes es el más usado en los estudios epidemiológicos actuales. Permite la inclusión de diferentes factores de riesgo (también llamados determinantes), en los mecanismos causales de la enfermedad.

En el ejemplo de la infección por el virus de la hepatitis B, claramente existe una causa necesaria (el virus), pero esto es cierto en sólo algunos casos, generalmente en enfermedades infecciosas y toxicológicas, que tienen un ente causal (causa necesaria) y factores de riesgo (causa suficiente). En el resto de enfermedades, los modelos están compuestos únicamente por causas suficientes, que inciden en el desarrollo de patologías.

## TIPOS DE DISEÑO EPIDEMIOLÓGICO

Los estudios descriptivos son la base de los programas de vigilancia epidemiológica, demográfica y en salud pública de cada país. Usualmente son realizados por instituciones que realizan una descripción de muertes, nacimientos, enfermedades, de la población en general de una región o país.

Los estudios de la epidemiología clínica son mayormente analíticos, es decir, traen un componente de análisis estadístico en sus diseños (Tabla N° 2). Estos a su vez pueden ser experimentales y no experimentales, diferenciándose en la manipulación de variables independientes en los estudios experimentales a través de la asignación aleatoria. La gran mayoría de estudios publicados son observacionales, por lo que en este manuscrito por cuestiones prácticas y de espacio, solo se analizan los diseños observacionales en epidemiología.

<b>TABLA 2. CONSIDERACIONES IMPORTANTES EN INVESTIGACIÓN EPIDEMIOLOGICA</b>	
<b>TIPOS DE VARIABLE EPIDEMIOLOGICA</b>	
<b>Variable</b>	<b>Categorica</b> <ul style="list-style-type: none"> <li>Nominal (dicotómico)</li> <li>Ordinal</li> </ul> <b>Cuantitativa:</b> <ul style="list-style-type: none"> <li>Intervalo</li> </ul>
<b>Variabes de estudio</b>	<b>Variable desenlace</b> - dependiente (o enfermedad) <b>Variable de exposición</b> - independiente (o factor de riesgo)
<b>Tipos de estudio</b>	
<b>Por tipo de análisis</b>	<b>Descriptivo</b> <b>Analítico</b>
<b>Por tipo de diseño</b>	<b>No experimental (observacional)</b> <ul style="list-style-type: none"> <li>Transversales (cross-sectional)</li> <li>Ecológicos</li> <li>Casos y controles</li> <li>Cohortes</li> <li>Estudios híbridos</li> </ul> <b>Experimental</b> <ul style="list-style-type: none"> <li>Ensayos clínicos controlados aleatorizados</li> <li>Cuasi-experimentos</li> </ul>
<b>Factores que afectan la validez</b>	
<b>Sesgo</b>	<b>Sesgo de selección</b> <b>Sesgo de información</b> <b>Confusión</b>
<b>Métodos para controlar la confusión</b>	
<b>En el diseño</b>	<b>Aleatorización</b> <b>Restricción</b> <b>Emparejamiento</b>
<b>En el análisis</b>	<b>Estratificación</b> <b>Modelamiento matemático</b>

## EL UNIVERSO Y LA MUESTRA

Los estudios observacionales trabajan con muestras para determinar la asociación entre la variable de exposición (factor de riesgo) o de desenlace (enfermedad). El universo es toda la población a la cual se va a estudiar, de la cual, por razones prácticas y con fundamento estadístico se extrae una muestra. Por muestra, se entiende a una selección de la población, que puede ser, dependiendo de sus características, no probabilística (por conveniencia), o probabilística (aleatoria, estratificada, o por conglomerados).

Las muestras no probabilísticas son muy y mal usadas en los diseños epidemiológicos observacionales. Un ejemplo claro son las muestras extraídas de encuestas a transeúntes en una calle, que si bien pueden ser 'aleatorias', no son representativas de una población. Muchos estudios epidemiológicos declaran muestras probabilísticas cuando en realidad usan muestras por conveniencia.

Existen varios tipos de muestreo probabilístico. En el muestreo aleatorio simple, todos los sujetos deben tener la misma probabilidad de ser elegidos. Es muy común en escenarios cerrados, como hospitales, cárceles y escuelas. Es uno de los muestreos más usados en epidemiología por su fácil obtención, a pesar de esto, es susceptible de sesgos de selección si no se eligen bien los sujetos. Por ejemplo, si una patología (como en infarto agudo al miocardio) es más frecuente y/o severo a una hora determinada del día, no se pueden elegir solo pacientes de unas horas determinadas (por ejemplo, durante el 'día').

El muestreo estratificado es un método complejo de selección de pacientes, en el cual se dividen los sujetos por una determinada característica, a la que se le asigna una probabilidad; dentro de la cual se toma una muestra aleatoria simple. El muestreo por conglomerados se puede dividir en dos: monoetápico, y estratificado polietápico. El primero se realiza idealmente en poblaciones dispersas, por conjuntos geográficos. El polietápico se realiza cuando la población a cubrir es muy grande y ampliamente distribuida.

Ambos requieren de técnicas estadísticas especiales, disponibles en algunos paquetes estadísticos informáticos (por ejemplo: Epi-Info, SPSS, Stata, SAS).

## DISEÑOS OBSERVACIONALES

Los estudios observacionales pueden llegar a ser muy eficientes para descubrir o lanzar hipótesis determinantes o factores de riesgo para una enfermedad, sin llegar a dilucidar en la inmensa mayoría de los casos, la causa de las mismas. Al mismo tiempo, sus alcances y metodologías pueden ser limitantes de

tal forma que los invalidan a la luz de los conocimientos epidemiológicos y estadísticos actuales.

Dependiendo del tipo de diseño observacional a utilizar, los estudios observacionales tienen ventajas y desventajas que los hacen útiles para exposiciones poco frecuentes (estudios de cohorte), enfermedades raras (estudios de casos y controles), o para ganar eficiencia en la generación de hipótesis (estudios transversales).

Los estudios que mayor nivel de evidencia proveen son los estudios de cohortes, le siguen los de casos y controles, y finalmente, abajo en los niveles de evidencia epidemiológica se encuentran los transversales.

**Estudios trasversales o cross-sectional:** estos determinan la prevalencia de una enfermedad dada en un momento singular del tiempo, con casos nuevos y viejos. En este tipo de estudio, la variable de exposición no está sujeta al control del investigador, por lo que es difícil determinar su grado de involucramiento en el desenlace.

Son muy útiles para generar nuevas hipótesis acerca de un componente causal, sin llegar a confirmarlo. Dadas sus características, este tipo de estudio es muy eficiente, y es uno de los más comunes en epidemiología. Pero esa eficiencia lleva un costo: la temporalidad, necesaria para confirmar la causalidad. No se puede determinar y no se sabe que fue primero, si la variable de exposición o de desenlace. Es por esto que en estos estudios, se omite la definición de causalidad, y se prefiere la de asociación, para dejar claro que lo que se encontró fue una relación estadística entre una variable de exposición y otra de desenlace.

El análisis de este tipo de estudios se puede realizar mediante test estadísticos que miran la asociación más allá del azar o estimadores epidemiológicos. Teniendo en cuenta que para evaluar asociaciones que evalúen factores confusores se necesitan modelos multivariados que miren varias variables independientes.

Muchos test estadísticos existen para comparar la asociación entre las variables de

exposición y desenlace. Cada uno se utiliza dependiendo del tipo de variable y su tipo de distribución de probabilidad. El principal indicador de la asociación más allá del azar es el valor de P, que por consenso, se establece que un valor menor de 0,1, 0,05, ó 0,01, indica una relación más allá del azar. Es decir una asociación propiamente dicha.

Uno de los estimadores epidemiológicos más utilizados en estudios es Odds Ratio (OR) o razón de momios. Es un número que indica el riesgo proporcional de presentar un desenlace, con relación a una exposición. Los valores de 0 a 1 indican un factor protector con relación al desenlace (a un valor de 0,4, se reduce la probabilidad del desenlace un 60%), y los valores mayores de 1 indican un factor de riesgo con relación al desenlace (a un valor de 1,4, se aumenta la probabilidad del desenlace un 40%). En teoría, el OR puede alcanzar el infinito, pero raras veces alcanza un valor de 100, a menos que existan sesgos o variables de confusión que lo afecten.

Existe una discusión epidemiológica acerca de la capacidad analítica de los estudios transversales. Sin embargo, a la luz del conocimiento actual, si la temporalidad de una exposición está bien documentada en un estudio transversal, no existen impedimentos que limiten su capacidad de realizar análisis que puedan generar hipótesis para luego ser comprobadas por estudios más complejos. Es así, que se han diseñado y validado diferentes modelos matemáticos para este tipo de estudios.

**Estudios de casos y controles:** el uso y entendimiento de los estudios de casos y controles ha sido uno de los más importantes avances de la epidemiología moderna (2). Su diseño es complejo y su entendimiento no es fácil. En vez de partir desde la exposición, se comienza desde el desenlace, llamándose 'caso' el paciente con desenlace, y 'control' el paciente sin el desenlace. A cada grupo se le busca evidencia previa de la exposición. Esto ha hecho que se le llamen retrospectivos, cuando en realidad pueden ser prospectivos o ambispectivos (retrospectivo y prospectivo al mismo tiempo), dependiendo de cómo se realicen. Debido a la naturaleza de este tipo de estudio, se adecua bien a enfermedades o desenlaces infrecuentes.

Un estudio de casos y controles parte del desenlace (outcome), buscándose la exposición hacia atrás. Esto lleva a un diseño eficiente y práctico. Esta ganancia de eficiencia lleva inconvenientes, que algunas veces son difíciles de eludir, el principal es que el muestreo debe ser independiente de la exposición. Es decir, la proporción de expuestos no debe verse sesgada en los casos y/o en los controles.

La selección de controles es uno de los aspectos críticos de este tipo de estudio. Para evitar sesgos, como primera medida, los controles deben seleccionarse de la misma base poblacional que los casos. Segundo, el tiempo-persona de cada control debe ser constante a través de todos los niveles de exposición. Este segundo punto se puede contrarrestar seleccionando los casos al mismo tiempo que los controles (el mismo día o momento determinado de tiempo). Por último, los factores de confusión deben ser contrarrestados en los controles, usualmente mediante análisis estadísticos.

Una de las estrategias para eludir la confusión es el pareamiento. Se realiza seleccionado los controles con base en una característica común en los casos, como la edad, el sexo, o el estrato socioeconómico. Una desventaja es que en el escenario que se realice el pareamiento de una variable, se pierde información sobre esta en los controles, pues su medición va a corresponder a la información de los casos. Además está el problema del sobrepareamiento, en el cual un estudio está emparejado por muchas variables; puede disminuir la validez interna y/o eficiencia del estudio.

Estudios de cohorte: los estudios de cohorte bien diseñados son uno de los métodos pivote en la determinación de causalidad. Una cohorte se define como un grupo de personas con una característica definida hacia un desenlace, en este caso la característica es la exposición. Este control de la exposición hace que uno de los criterios de causalidad sea fielmente investigado, el de la temporalidad. A pesar de esto, la propiedad observacional del estudio, y la falta de asignación aleatoria de variables de confusión, hace que no se pueda establecer con certeza la causalidad. Por este motivo, se han diseñado diferentes modelos

matemáticos y métodos epidemiológicos para controlar la confusión.

Los estudios de cohorte pueden ser prospectivos o retrospectivos. Teniendo en cuenta que en los estudios retrospectivos se debe tener una información válida y completa de todas las variables de interés. Esto es poco frecuente en países en desarrollo, y por eso en el contexto latinoamericano, generalmente estos estudios están reservados a poblaciones especiales, mayormente hospitalarias o de poca duración de seguimiento (pacientes maternas, por ejemplo). En países europeos y asiáticos desarrollados, con cobertura de salud universal, los datos de todos los pacientes son guardados en grandes bases de datos que permiten el análisis poblacional retrospectivo de los mismos.

Pueden clasificarse como fijos o dinámicos. Fijos cuando todos los individuos entran en el mismo momento del tiempo (el mismo día, por ejemplo), y dinámicos, cuando entran y salen a medida que pasa el tiempo de observación del estudio.

Un esquema simple de cómo se categorizan los pacientes y cómo se realiza el análisis en un estudio de cohortes se encuentra en la Figura 2. En los estudios de cohorte, como se miran solo los desenlaces nuevos, se estima la incidencia. El indicador epidemiológico ya no es el OR, sino el riesgo relativo (RR), que va a estar dado por la incidencia del desenlace en el grupo de expuestos sobre la incidencia del desenlace en el grupo de no expuestos.

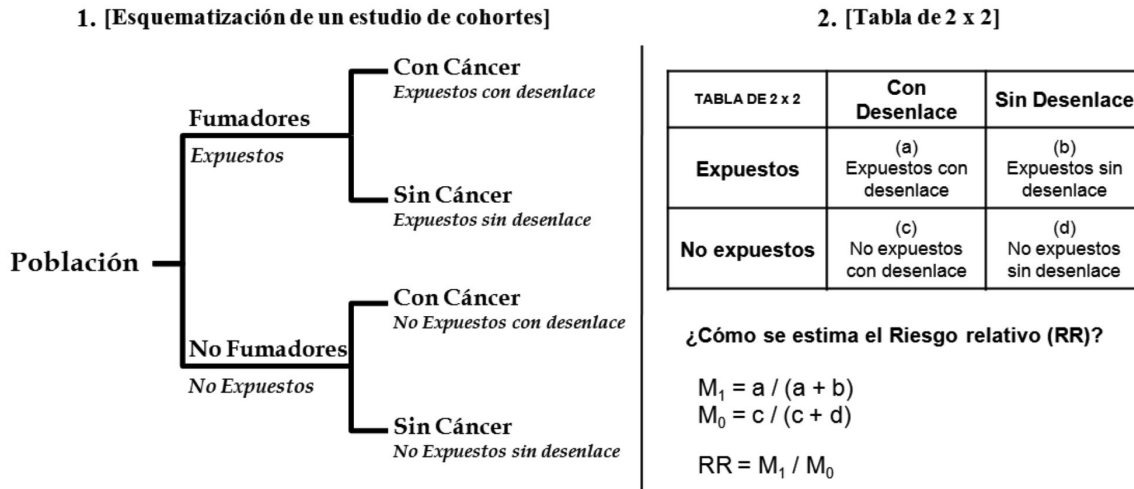
De esta manera (ver Figura 2; parte 2):

$$M_1 = a / (a + b)$$

$$M_0 = c / (c + d)$$

Donde  $M_1$  es la incidencia en el grupo de expuestos, y  $M_0$  la incidencia en el grupo de no expuestos. El riesgo relativo (RR), es la razón de las dos incidencias ( $M_1/M_0$ ) (ver Figura 2; parte 2).

EL RR es la medida de asociación entre la exposición y el desenlace que se utiliza en los estudios de cohortes, y se interpreta de la misma forma que el OR. Existen otras medidas de asociación que se utilizan en este tipo de estudio, pero analizarlas no está dentro de los objetivos de esta revisión.

**FIGURA 2. ESQUEMATIZACIÓN EPIDEMIOLÓGICA DE UN ESTUDIO DE COHORTES**

## LA VALIDEZ DE UN ESTUDIO DE INVESTIGACIÓN

La validez de un estudio depende de forma general de dos elementos, la validez interna ¿Qué tan bien hecho está un estudio?, y la validez externa ¿Qué tan generalizable es en otros escenarios?.

Para que un estudio epidemiológico sea considerado como válido, se necesita un análisis de los sesgos y factores de confusión que amenazan la validez interna del estudio. Para que exista la generalizabilidad de los resultados, es necesario que exista ausencia de sesgos. Existen dos tipos principales de sesgos que amenazan la validez interna de un estudio: de selección y de información.

Un sesgo de selección está dado por un error en la asignación de los sujetos en el desenlace o en la exposición. Ocurre cuando por ejemplo se toma la población hospitalaria de un estudio, y por razones ajenas al investigador, las probabilidades de exposición son diferentes tanto en los casos como en los controles.

Los estudios de casos y controles son particularmente susceptibles a estos tipos de sesgo, dado que frecuentemente se hacen en escenarios hospitalarios, donde se desconoce muchas veces la frecuencia de enfermedad en la cohorte poblacional.

Los sesgos de información están relacionados

con la falta de una adecuada recolección de información, que hace que un grupo de pacientes se categorice equivocadamente dentro de las posibles opciones de por ejemplo, una tabla de contingencia (tabla de 2 x 2) (Figura 2; parte 2). Esto efectivamente sesga los resultados, consecuentemente quitándole su validez externa e interna.

## CONCLUSIONES

Los estudios observacionales, a pesar de sus limitaciones, proveen importante evidencia en la ciencia epidemiológica. Es así como algunas veces son indispensables en la determinación de factores de riesgo de las enfermedades. Sus limitaciones están dadas principalmente por la presencia de factores de confusión o sesgos que impidan la adecuada estimación de la importancia del factor de exposición en la enfermedad.

Se presentó en esta revisión una visión global básica acerca de causalidad y estudios observacionales en epidemiología. Para profundizar conceptos acerca de este tema, se recomiendan los libros de Epidemiología de Rothman (2008) (2) y Fletcher y col. (1996) (2), en los cuales se encontrarán las bases teóricas de este campo científico.

**CONFLICTOS DE INTERESES:** ninguno que declarar.

**FINANCIACIÓN:** recursos propios del autor.

## REFERENCIAS BIBLIOGRÁFICAS

1. Last J. Dictionary of Epidemiology, Third edition. New York: Oxford U. Press; 1995.
2. Rothman K, Greenland S. Modern epidemiology, 3rd Edition. Filadelfia: Lippincott-Raven Publishers; 2008.
3. Hofler M. The Bradford Hill considerations on causality: a counterfactual perspective. Emerg Themes Epidemiol. 2005 Nov 3;2:11.
4. Hill A. The environment and disease: Association or causation? Proceed Roy Soc Medicine - London 1965;58:295-300.
5. Goodman K, Phillips C. The Hill criteria of causation. Encyclopedia of Statistics in Behavioral Sciences. London: Wiley; 2005.
6. Naschitz JE, Kovaleva J, Shaviv N, Rennert G, Yeshurun D. Vascular disorders preceding diagnosis of cancer: distinguishing the causal relationship based on Bradford-Hill guidelines. Angiology. 2003 Jan;54(1):11-7.
7. Holt RI, Peveler RC. Antipsychotic drugs and diabetes--an application of the Austin Bradford Hill criteria. Diabetologia. 2006 Jul;49(7):1467-76.
8. Fletcher R, Fletcher S, Wagner E. Clinical epidemiology--The Essentials. Baltimore: Williams & Wilkins; 1996.
9. Kleinbaum D, Kupper L, Morgenstern H. Epidemiologic research: Principles and quantitative methods Belmont, CA: Lifetime Learning Publications; 1982.



### GRUPO DE INVESTIGACIÓN: ALTA TENSIÓN

LÍNEAS: Enfermedad Renal Crónica. Enfermedades Infecciosas. Hipertensión Arterial. Medicina Crítica  
Email: [jocodada@yahoo.es](mailto:jocodada@yahoo.es)