



Anales de la Facultad de Medicina

ISSN: 1025-5583

anales@unmsm.edu.pe

Universidad Nacional Mayor de San Marcos

Perú

Bullón, Luz

La estadística en la investigación epidemiológica: El estudio de casos y controles

Anales de la Facultad de Medicina, vol. 63, núm. 2, 2002, pp. 46-56

Universidad Nacional Mayor de San Marcos

Lima, Perú

Disponible en: <http://www.redalyc.org/articulo.oa?id=37963207>

- ▶ Cómo citar el artículo
- ▶ Número completo
- ▶ Más información del artículo
- ▶ Página de la revista en [redalyc.org](http://redalyc.org)

 redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal  
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

## La estadística en la investigación epidemiológica: El estudio de casos y controles

LUZ BULLÓN

*Departamento de Medicina Preventiva y Salud Pública. Facultad de Medicina, UNMSM.*

### RESUMEN

Los problemas más frecuentes en la investigación en epidemiología son resueltos mediante el diseño adecuado de estudios. La estadística ha contribuido a la conceptualización, desarrollo y éxito de dichos estudios. Este trabajo presenta el estudio de casos y controles, su relación con el estudio de cohortes, una revisión de los aportes en los últimos 50 años, los cuales fueron de enorme importancia en salud pública y ciencias biomédicas. También, se discute la forma de medir la asociación entre factor de riesgo y enfermedad, utilizando el parámetro *razón de chances (odds ratio)* y el mismo desde el aporte de Mantel y Haenszel.

*Palabras clave: Estadística; estadísticas de salud; estadísticas sanitarias; epidemiología; investigación; investigación sobre servicios de salud.*

### STATISTICS IN EPIDEMIOLOGICAL RESEARCH: THE CASE-CONTROL STUDY SUMMARY

The most common problems in epidemiological research use appropriate study design. Statistics has contributed to conceptualisation, development and success of these studies. This paper presents the case-control study, its relationship with cohort study and its important contribution in the past 50 years in public health and biomedicine. It also discusses the way of measuring the association between risk factor and disease using the *odds ratio* and the Mantel and Haenszel's contribution.

*Key words: Statistics; health statistics; sanitary statistics; epidemiology; research; health services research.*

### Introducción

Los problemas más frecuentes en la investigación epidemiológica son el estudio de la asociación entre la exposición a un factor y el eventual desarrollo de una enfermedad, la comparación de opciones terapéuticas y el estudio de factores de pronóstico para pacientes sometidos a un tratamiento dado. Para resolver estos problemas se encuentra, frecuentemente, las tres formas de investigación siguientes: estudios de casos y controles, estudios de cohortes y ensayos clínicos aleatorizados (35).

---

#### Correspondencia:

Ing. Luz Bullón Camarena  
Facultad de Medicina - UNMSM.  
Av. Grau 755, Lima 1, Perú  
E-mail: [d100639@unmsm.edu.pe](mailto:d100639@unmsm.edu.pe)  
[correolucero@yahoo.es](mailto:correolucero@yahoo.es)

---

Estos estudios son comparativos, es decir, la conclusión final se obtiene a través de la comparación de dos grupos de individuos. En los estudios de casos y controles, la comparación se da entre un grupo de personas enfermas (los casos) y un grupo de personas no enfermas (los controles). En los estudios de cohortes, se compara un grupo expuesto al factor de riesgo en estudio con otro grupo no expuesto. Los estudios de casos y controles y de cohortes son estudios observacionales, es decir, el investigador no impone un tratamiento, utiliza los datos producidos por la historia clínica del paciente y son, por razones éticas o prácticas, las estrategias de investigación más viables. El ensayo clínico aleatorizado es un estudio experimental y es la forma más conclusiva de investigación donde, además de la presencia del grupo de control, la comparación de los grupos es garantizada por

la asignación aleatoria de los pacientes a los tratamientos.

La estadística ha contribuido enormemente a la conceptualización, desarrollo y éxito de dichos estudios, en particular, de los estudios de casos y controles y, en los últimos 50 años, esta contribución ha sido de las más importantes para la salud pública y las ciencias biomédicas. Por todo ello, este trabajo presenta el diseño de los estudios de casos y controles, estableciendo su relación con el estudio de cohortes, hace un breve recorrido histórico por las principales contribuciones del punto de vista estadístico de este diseño (6) y señala algunos desafíos de investigaciones en esta línea.

### El muestreo en los estudios de casos y controles

Los estudios de casos y controles se inician con la presencia o ausencia de un resultado o evento de interés y luego se investiga hacia atrás en el tiempo, para tratar de identificar causas o factores de riesgo posibles de dicho resultado.

Los casos en estos estudios se seleccionan de manera individual, sobre la base de alguna enfermedad o consecuencia de ella; los controles son personas comparables con los casos, pero sin la enfermedad o consecuencia. Las historias o sucesos previos de ambos, casos y controles, se los analiza, en un intento por identificar una característica o factor de riesgo presente en los casos, pero no en los controles.

La naturaleza de la indagación en el tiempo permite que, en ocasiones, sean denominados estudios retrospectivos. Son también longitudinales, debido a que la indagación cubre un cierto tiempo. La Figura 1 presenta un esquema que aclara esto último. Las áreas sombreadas representan sujetos expuestos al factor antecedente, las áreas claras a no expuestos. Los cuadrados representan sujetos con la enfermedad o consecuencia que interesa, los rombos son los sujetos sin ella.

### El estudio de casos y controles frente al estudio de cohortes

Se presenta el concepto y el esquema correspondientes a un estudio de cohortes, de manera que se comprenda posteriores análisis y comparaciones. En el estudio de cohortes, se selecciona sujetos al principio y se determina si poseen el factor de riesgo o han sido expuestos. Se observa estas personas expuestas y no expuestas por cierto tiempo, para constatar el efecto de este factor. Debido a que los eventos de interés, como enfermedad, se manifiestan después de iniciado el estudio, en ocasiones se denomina prospectivo (Figura 2).

Los estudios de cohortes también sirven para identificar factores de riesgo posibles y, por tanto, siguen de manera natural a un estudio de casos y controles.

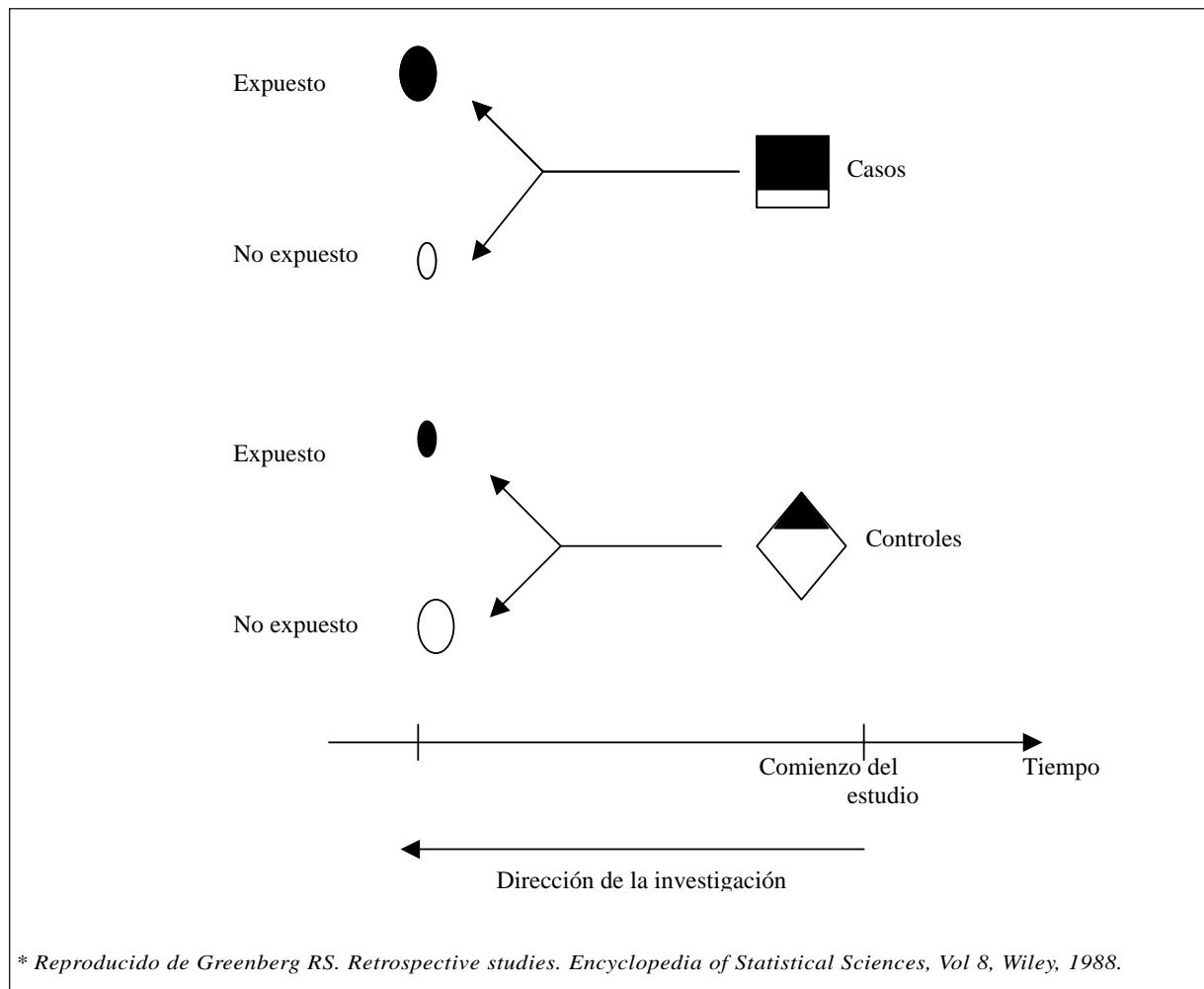
En estos diseños resulta adecuada la comparación de los grupos mediante las tasas de incidencia de la enfermedad en estudio (D, resultado observado). Desde que los tamaños de grupos expuesto y no expuesto han sido predeterminados, una medida intuitiva, natural del efecto de exposición al factor de riesgo, es la razón de tasas (probabilidades). Esta razón es ampliamente conocida como riesgo relativo (RR).

$$RR = \frac{P(D / \text{expuesto})}{P(D / \text{no expuesto})}$$

Las magnitudes de las probabilidades no tienen importancia (aún cuando los eventos sean raros o poco probables); sólo es importante el cociente de estas probabilidades.

Presentado de otra forma, si X denota el resultado de interés, la enfermedad (1, si y 0, no) y E denota la exposición (1 para expuestos y 0 para no expuestos), entonces:

$$RR = \frac{P(X = 1 / E = 1)}{P(X = 1 / E = 0)} \quad (1)$$



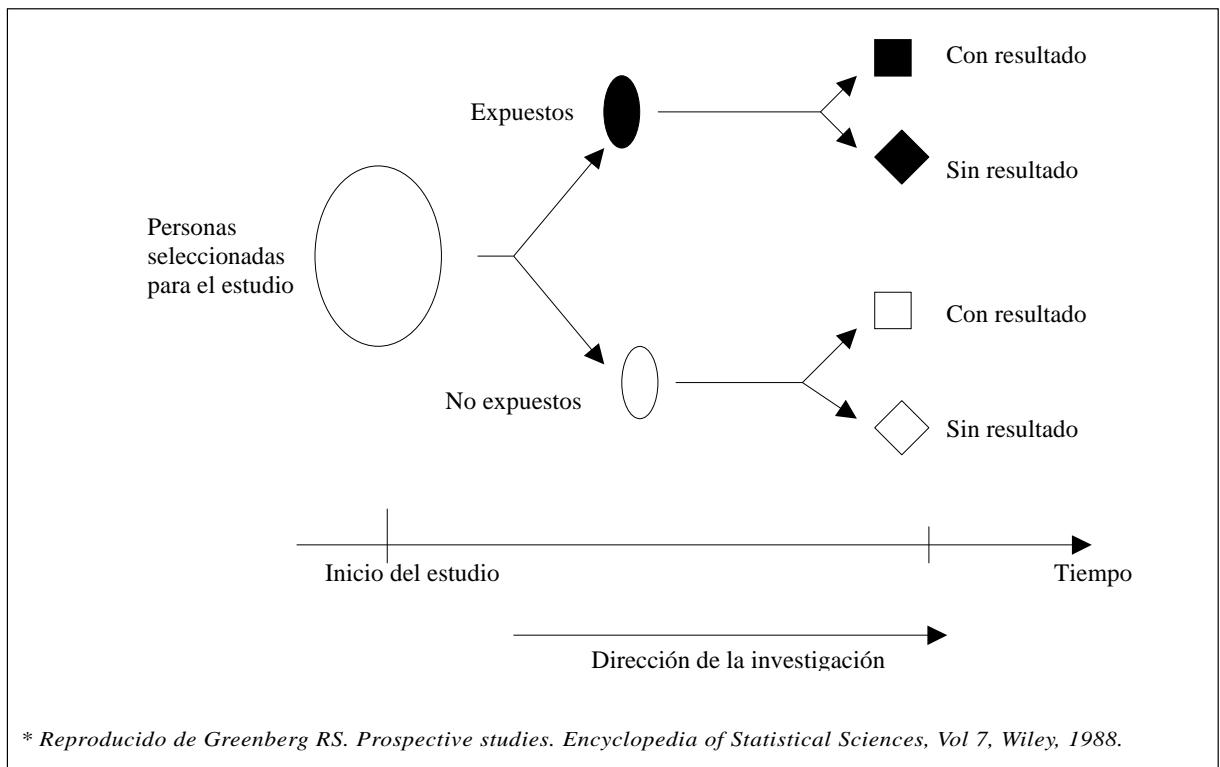
**Figura 1.-** Esquema de un estudio de casos y controles.

### Orígenes y antecedentes

El ejemplo más antiguo de un estudio de casos y controles es el de Guy, en 1843, que compara las ocupaciones de pacientes con enfermedades pulmonares, como tuberculosis, con aquellas de pacientes con otras enfermedades<sup>(23)</sup>. El método se vuelve muy popular en los años 20 para el estudio del cáncer, siendo éxitos notables las asociaciones descubiertas entre cáncer a los labios y el tabaquismo en pipa<sup>(2)</sup>, entre el

cáncer mamario e historia reproductiva de la mujer<sup>(24)</sup> y entre cáncer oral y hábito de fumar<sup>(26)</sup>. Debido a que estas enfermedades eran raras, no resultaba práctico ser estudiadas de otra forma; por ejemplo, por seguimiento de una población inicialmente sana.

La atención dirigida a este tipo de estudios se incrementó, y a la crítica de la metodología siguió la publicación, en 1950, de varios estudios de la asociación entre hábito de fumar y



\* Reproducido de Greenberg RS. Prospective studies. Encyclopedia of Statistical Sciences, Vol 7, Wiley, 1988.

**Figura 2.-** Esquema de un estudio de cohortes.

cáncer pulmonar (36). Bajo el liderazgo de Harold Dorn, los estadísticos del Instituto Nacional de Cáncer de los Estados Unidos de América se vieron estimulados por la controversia e investigaron sobre las ventajas y limitaciones de los estudios de casos y controles. Una idea que aún ahora prevalece es que muestras separadas de casos y controles no proporcionan información cuantitativa relevante acerca de los parámetros, de mayor interés y utilidad, como las tasas de enfermedad.

Esta concepción errada fue corregida por Jerome Cornfield (11), quien demostró que la *razón de chances* (odds ratio, OR) de la exposición para los casos respecto a los controles es igual a la *razón de chances* de la enfermedad para los expuestos respecto a los no expuestos, y que esta última se aproxima de la razón de

tasas de enfermedad (RR), siempre que la enfermedad sea rara.

La *chance* a favor de un evento E, que ocurre con probabilidad  $p$ , es de  $\frac{p}{1-p}$ . Por ejemplo,

si la probabilidad de E es  $p = 1/2$ , la *chance* de E es  $\frac{1/2}{1/2} = 1$ , es decir de 1 a 1, mientras que ésta

será de 2 a 1 si  $p = 2/3$ .

De esta forma, la *razón de chances* de la enfermedad (D) para los expuestos respecto a los no expuestos es definida como la *chance* a favor de la enfermedad entre individuos expuestos dividida por la *chance* de enfermedad entre los no expuestos:

$$OR = \frac{P(D / \text{expuesto}) / [1 - P(D / \text{expuesto})]}{P(D / \text{no expuesto}) / [1 - P(D / \text{no expuesto})]}$$

Análogamente, para la *razón de chances* de la exposición E, para los casos respecto a los controles

$$OR = \frac{P(E / \text{caso}) / [1 - P(E / \text{caso})]}{P(E / \text{control}) / [1 - P(E / \text{control})]}$$

Formalmente, si D denota la enfermedad (1 para casos y 0 para controles) y X denota la exposición (1 para expuestos y 0 para no expuestos), entonces:

$$\begin{aligned} & \frac{P(X = 1 / D = 1) P(X = 0 / D = 0)}{P(X = 1 / D = 0) P(X = 0 / D = 1)} \\ &= \frac{P(D = 1 / X = 1) P(D = 0 / X = 0)}{P(D = 1 / X = 0) P(D = 0 / X = 1)} \\ &\approx \frac{P(D = 1 / X = 1)}{P(D = 1 / X = 0)} \end{aligned} \quad (2)$$

La *razón de chances* de la exposición, ahora ampliamente conocida como el riesgo relativo, de esta forma es conveniente para aproximar la razón de tasas de la enfermedad (1).

También, Cornfield demostró que si la tasa de enfermedad en la población general es conocida, entonces ésta puede ser combinada con los datos de casos y controles para proporcionar estimadores separados de tasas de enfermedad para expuestos y no expuestos. Él estuvo consciente de la necesidad de que los grupos de casos y controles sean "representativos de los grupos correspondientes en la población general" (11), para que los cálculos sean válidos.

Las tasas de enfermedad a los que Cornfield hace referencia son las tasas de prevalencia, es decir, proporciones simples de individuos con la enfermedad, tal como sería obtenido en una muestra de un estudio transversal. Sin embar-

go, para estudios de etiología de una enfermedad, es preferible estudiar tasas de incidencia y estimar la probabilidad (o razón de probabilidades) de desarrollo de la enfermedad durante un periodo de tiempo determinado para individuos quienes son libres de enfermedad en el inicio. En caso contrario, se corre el riesgo de confundir los efectos de la exposición sobre la tasa de fatalidad de los casos. Ahora se sabe, que aún sin el supuesto de enfermedad rara, la razón de chances de exposición es estimable en un estudio de casos y controles, aproximándolo por la razón de tasas instantáneas de incidencia, en vista que los controles son muestrados proporcionalmente con los casos incidentales, a través del periodo de estudio.

La demostración de Cornfield no aquietó todas las críticas, estando entre éstas las de Fisher (17). Fisher colocó a la luz la cuestión de la asociación frente a la causalidad, que no esclarece la interpretación de cualquier estudio observacional. Es famosa su hipótesis en la cual sugiere que la tal asociación puede ser explicada por el efecto de un factor de confusión (nuevo factor asociado tanto a la enfermedad como al factor en estudio). Cornfield y otros autores (13) respondieron ampliamente a estas y otras críticas en artículos de lectura valiosa, respecto a inferencia causal. El argumento central de la respuesta fue un simple cálculo, mostrando que, para que un factor de confusión explique un riesgo relativo de una magnitud dada, tal como  $\psi$ , este factor tiene que ser  $\psi$  veces más prevalente entre los expuestos que entre los no expuestos. Más adelante, este cálculo se formalizó dentro del concepto de razón de riesgo de confusión, el cual mide la posible extensión de la confusión sobre el riesgo relativo observado (29,33). Estos aportes y desarrollos dieron posteriormente al riesgo relativo solidez como un parámetro significativo y de gran importancia en todo estudio epidemiológico.

A finales de la década, fueron comprendidos bastante bien los requisitos deseables para los estudios de casos y controles, de manera que éstos produzcan resultados confiables y con

implicancias en las acciones posibles de salud pública. Los puntos principales, resumidos por Dorn (14), fueron los siguientes. El estudio debería a) ser conducido en una población definida claramente, b) incluir todos los casos incidentales ocurridos durante un periodo de tiempo determinado, c) utilizar medidas objetivas de la exposición a los factores de riesgo, d) usar múltiples grupos de control, e) ser repetido en diferentes poblaciones, y f) ser verificado por un estudio de cohorte.

### El aporte de Mantel y Haenszel

Cole (10) escribió “Los estudios de casos y controles llevados a cabo por epidemiólogos en los últimos 20 años se han soportado sobre los hombros de gigantes, ... y la epidemiología no debe perder de vista a la raíz más importante de nuestra disciplina, debemos recordar que todos aquellos hombres son y fueron estadísticos”.

Los estadísticos a quienes Cole se refiere son Cornfield y Dorn y sus colegas Mantel y Haenszel. Éstos últimos (27), publicaron en la revista del Instituto Nacional del Cáncer un artículo que marcó un hito. Mantel y Haenszel (27), entre muchos aportes, aclaran la relación entre los estudios de casos y controles (retrospectivos) y el de cohortes (prospectivos), señalando la diferencia conceptual entre ambos tipos de estudio y anticipa el surgimiento y desarrollo de los estudios de casos y controles anidados (sección 6).

También introdujeron dos procedimientos estadísticos para el control del factor de confusión, mediante estratificación de los datos en una serie de tablas 2×2. El primero de ellos fue un ajuste de la prueba de Chi-cuadrado previamente presentada por Cochran (9). El segundo fue el estimador sumario del riesgo relativo, destinado a resumir las *razones de chances* (OR) individuales, ponderándolos por precisión e importancia.

El estimador sumario del riesgo relativo, en adelante denominado estimador de Mantel-Haenszel (MH), es definido por la fórmula simple que señalaremos a seguir utilizando la notación establecida en las Tablas 1 y 2.

**Tabla 1.-** Datos dispuestos en una tabla única 2×2.

Exposición	Frecuencias observadas		
	Casos D = 1	Controles D = 0	Total
X = 1	A	B	T
X = 0	C	D	N - t
Total	N <sub>1</sub>	N <sub>0</sub>	N

Los datos son estratificados en k estratos, de acuerdo a una o más variables de confusión. Se forman tablas 2×2 para cada estrato i = 1, 2, ..., k, relacionando exposición y enfermedad.

**Tabla 2.-** Notación para la serie de tablas 2×2.

Exposición	Frecuencias observadas			Frecuencias esperadas		
	Casos D = 1	Controles D = 0	Total	Casos D = 1	Controles D = 0	Total
X = 1	A <sub>i</sub>	B <sub>i</sub>	t <sub>i</sub>	N <sub>1i</sub> p <sub>1i</sub>	N <sub>0i</sub> p <sub>0i</sub>	
X = 0	C <sub>i</sub>	D <sub>i</sub>	N <sub>1i</sub> - t <sub>i</sub>	N <sub>1i</sub> q <sub>1i</sub>	N <sub>0i</sub> q <sub>0i</sub>	
Total	N <sub>1i</sub>	N <sub>0i</sub>	N <sub>i</sub>	N <sub>1i</sub>	N <sub>0i</sub>	

Razón de chances estimada del estrato i,  $\Psi_i = (p_{1i} q_{0i}) / (q_{1i} p_{0i})$

Luego el estimador de MH es

$$\psi_{MH} = \frac{\sum_i R_i}{\sum_i S_i} \quad (3)$$

$$\text{donde } R_i = \frac{A_i D_i}{N_i} \quad \text{y} \quad S_i = \frac{B_i C_i}{N_i}$$

Ambos aportes, el estimador M-H y la prueba de hipótesis de Mantel- Haenszel, se tornaron procedimientos eventualmente adoptados para uso rutinario de epidemiólogos, con el beneficio en el análisis de sus datos en un arreglo tabular y la posibilidad de hacer comparaciones de riesgos relativos individuales y sumarios, los cuales los alertaban de una posible heterogeneidad.

El impacto Mantel y Haenszel <sup>(27)</sup> fue enorme. Un estudio del año 1994 da cierta cuenta de este impacto; el artículo ha tenido 4000 citaciones en la literatura científica, siendo además uno de los 200 artículos más citados desde 1945. Continúa siendo mencionado a razón de 250 veces por año, más hoy en día que durante las décadas 70 y 80 (dato del *Institute for Scientific Information*, comunicación personal, Filadelfia).

Mantel y Haenszel, sin embargo, no presentaron una fórmula analítica para su estimador e hicieron referencia al trabajo de Cornfield <sup>(12)</sup> para el cálculo de intervalos de estimación. Parte del problema ocasionado por la falta de dicha variancia fue el hecho de que el estimador haya tenido la misma utilidad en dos contextos asintóticos diferentes: 1) un número pequeño de tablas con un número grande de frecuencias, y 2) un número grande de tablas con frecuencias pequeñas, como indistintamente se puede encontrar en estudios publicados. Llevó 25 años encontrar dicha fórmula. (siguiente sección)

### Verosimilitud e inferencia en la estimación de OR

El aspecto del trabajo de Fisher de mayor impacto sobre las aproximaciones modernas al análisis de datos de estudios de casos y con-

troles es el desarrollo de inferencia a partir de la verosimilitud, basada en modelos explícitos de probabilidad <sup>(15)</sup>. Este aspecto es enfatizado en textos de estadística aplicada a la epidemiología, con capítulos referidos a verosimilitud. El mismo Fisher introduce la idea de la verosimilitud para la razón de chances en tablas 2×2, en su artículo clásico sobre la lógica inductiva Fisher <sup>(16)</sup>.

La característica más importante de su análisis es el uso de la distribución condicional de datos provenientes del producto de dos binomiales con los totales marginales fijos, esto debido a que el tamaño de la muestra de casos así como de controles son establecidos de antemano. Así, el número de expuestos resultante en cada grupo es una variable aleatoria que tiene distribución binomial con parámetro  $p_i$  para los casos y  $p_0$  para los controles.

La distribución resultante del producto de binomiales de tamaños fijos es conocida como una extensión de la distribución hipergeométrica <sup>(7)</sup>, aquella donde el parámetro es la razón de chances  $\psi_i$  desconocida. Continuando con la notación de la Tabla 2 se tiene

$$P(A_i = a_i | t_i; \psi_i) \propto \binom{N_{li}}{a_i} \binom{N_{0i}}{t_i - a_i} \psi_i^{a_i} \quad (4)$$

Cornfield (1956) approximó los límites de confianza exactos para  $\psi_i$  basados en (4) asumiendo la distribución normal para  $A_i$  con media  $\hat{A}_i$  el valor fijado determinado por los totales marginales y el valor desconocido  $\psi_i$  como la solución de la ecuación cuadrática

$$\frac{\hat{A}_i (N_{0i} - t_i + \hat{A}_i)}{(N_{li} - \hat{A}_i)(t_i - \hat{A}_i)} = \psi_i \quad (5)$$

y variancia asintótica

$$\text{var}_i(\hat{A}_i) = \left( \frac{1}{\hat{A}_i} + \frac{1}{N_{0i} - \hat{A}_i} + \frac{1}{t_i - \hat{A}_i} + \frac{1}{N_{0i} - t_i + \hat{A}_i} \right)^{-1} \quad (6)$$

El procedimiento fue extendido de la *razón de chances* común a la *razón de chances* para una serie de tablas, tanto para la estimación puntual como para el cálculo de intervalos de confianza (21). Límites de confianza exactos para el valor del estimador sumario, los cuales involucran la convolución de las distribuciones en (4), están disponibles en *softwares* comerciales.

La prueba de Mantel-Haenszel se basa en la diferencia entre las frecuencias observada

$O = \sum_i A_i$  y esperada correspondientes y en la variancia de  $A_i$ , que por la distribución hipergeométrica es

$$v_i = \frac{(A_i + B_i)(C_i + D_i)(A_i + C_i)(B_i + D_i)}{N_i^2(N_i - 1)}$$

La variancia de  $O$  es  $V = \sum_i v_i$ . Luego la hipótesis nula de no asociación tiene estadística de prueba

$$\chi_{MH}^2 = \frac{(|O - E| - 1/2)^2}{V}$$

cuya distribución es Chi-cuadrado, con un grado de libertad bajo el supuesto de  $H_0$  verdadera.

Se encuentra muchos trabajos a partir de la distribución y la verosimilitud condicional de Fisher. Particularmente, uno que involucra modelos de regresión de riesgos relativos de la forma  $\log \psi_i = z_i \beta$  (3), donde  $\beta$  es obtenido de las ecuaciones de máxima verosimilitud condicional.

### Algunas cuestiones informativas

#### 7.1 Estudios de casos y controles pareados y anidados

El principio de los estudios de casos y controles -que ambos sean comparables (simila-

res) excepto en el factor de riesgo en estudio-, llevó a que en la práctica se busque caso a caso, un individuo (control) par del caso. El entusiasmo inicial derivado del logro del pareamiento tendió a desaparecer, debido a que se ignoró los problemas aparentes surgidos del muestreo no independiente. En consecuencia, el pareamiento como técnica de selección de los controles, disminuyó en preferencia entre los epidemiólogos, debido a la complejidad que trajo consigo un análisis estadístico riguroso de los datos.

Procedimientos especiales para estos diseños pareados con exposición binaria se presentan en Miettinen (28), Pike and Morrow (30), pero un tratamiento general aún falta. Finalmente, se debe agregar que un análisis no pareado de datos pareados generalmente proporciona estimaciones conservadoras del riesgo relativo (4).

La idea de los estudios de casos y controles anidados surgió con Mantel y Haenszel, y la sugerencia de que podría llevarse a cabo un diseño de casos y controles aún con el muestreo de individuos considerados en una investigación bajo el diseño de cohortes, y con el fin de colectar datos de individuos adicionales. Con esta idea, establecieron la diferencia *conceptual* entre ambos tipos de estudio, siendo ésta que los estudios de casos y controles involucran el muestreo de una cohorte en lugar de la enumeración completa de ésta.

Prentice (31) introduce el diseño de casos y controles como un método alternativo de muestreo de una cohorte definida. De esta forma, una subcohorte aleatoria es muestreada de la cohorte entera en el periodo de seguimiento establecido; luego, la información de la exposición es procesada para los individuos de la subcohorte. Estos diseños ofrecen importantes ventajas sobre los diseños comunes, en situaciones donde enfermedades múltiples van a ser evaluadas debido a que individuos no afectados de la subcohorte pueden servir como controles, para los casos de cada tipo de enfermedad.

La elección entre ambos diseños de casos y controles es menos clara cuando hay una sola enfermedad en estudio. Langholz and Thomas (25) mostraron que el diseño de casos y controles anidado puede ser más eficiente que el común cuando hay censuras aleatorias moderadas dentro de la cohorte.

## 7.2 Propuestas y desafíos en el aspecto del muestreo

Está claro que los estudios de casos y controles son usados para el estudio de enfermedades raras, debido a que ellos son más eficientes cuando son comparados con los estudios de cohorte del tamaño necesario para proporcionar el mismo número de casos. El método concentra recursos donde hay la mayor cantidad de información, es decir, los casos. Últimamente hay análisis conducentes a una posible mejora en la eficiencia, mediante esquemas de muestreo más complejos para maximizar la variación en las exposiciones de los casos y controles.

La propuesta de White (37) tiene que ver con el diseño de muestreo de dos etapas para estudiar los efectos de una exposición rara, sobre

una enfermedad rara, que inicia con un muestreo de cohortes en la primera etapa. Breslow y Cain (5) han considerado un diseño de casos y controles bietápico. La Tabla 3 ilustra la idea básica con datos ficticios de un estudio bietápico de cáncer pulmonar y el factor empleo en una planta industrial. En la etapa inicial 1500 casos y 500 controles son seleccionados al azar de la comunidad y clasificados de acuerdo al empleo en la planta industrial. Es esencial la información respecto al hábito de fumar, para el ajuste de la estimación del riesgo relativo, pero los investigadores pueden llevar a cabo una entrevista sólo a aproximadamente la mitad de los 1000 sujetos. En vista de lo raro de la exposición, los investigadores deciden entrevistar a todos los casos y controles expuestos en la etapa II, sin considerar la información ya disponible de ellos. Esto significa, sin embargo, que la relación entre exposición y enfermedad es distorsionada por aquellos quienes tienen los datos completos (Tabla 3).

Lo anterior supone, de manera general, que  $N_1$  casos y  $N_0$  controles son clasificados en J estratos sobre la base de una, posiblemente gruesa, medida de exposición inicial.

**Tabla 3.-** Datos ficticios de un estudio de casos y controles bietápico.

Etapa I ( $N_{ij}$ )			
Exposición	Casos	Controles	
Empleados de planta industrial	40	20	
Empleados otros	460	480	
Total	500	500	

Etapa II ( $n_{ijk}$ )				
Exposición	Casos (i =1)		Controles(i =0)	
	Planta Indust. (j =1)	Otros (j =2)	Planta Indust. (j =1)	Otros (j =2)
Fumadores (k =1)	30	37	10	42
No fumadores (k =2)	10	133	10	128
Total ( $n_{ij}$ )	40	170	20	170

El desafío estadístico es combinar la información disponible de ambas etapas de muestreo en una forma más eficiente, de manera que se estime y ajuste el riesgo relativo. Propuestas de solución presentan Schill y colegas (34), así como Breslow y Cain (5).

Los diseños bietápicos tienen muchas potenciales aplicaciones. Un ejemplo es el estudio de casos y controles con submuestreo de validación, para la corrección de errores de medición (8).

### Observaciones finales

Los estadísticos han contribuido enormemente en el desarrollo de los modernos estudios de casos y controles. Sus aportes, en términos de muestreo de un estudio de cohorte real o ficticio, aclaran la estrecha relación entre estas dos principales metodologías de la epidemiología analítica. Ellos han logrado que se entienda mejor el papel del pareamiento en el diseño de este tipo de estudio, así como el análisis de los datos que de éstos provienen. La regresión logística y otras involucrando riesgos relativos, basados en conceptos de verosimilitud, proporcionan a los epidemiólogos modernas herramientas estadísticas para evaluación y validación de los modelos, así como de detección de datos atípicos (22).

Sin embargo, a despecho de estas técnicas avanzadas, persisten los problemas fundamentales de obtención de inferencias causales, a partir de datos observacionales. Fisher se mantuvo escéptico ante la afirmación de que el cigarrillo causa cáncer pulmonar, aún frente a la comunidad médica que miraba una evidencia abrumadora. Su punto de vista extremista se entiende mejor si se recuerda que él tuvo de ambos: el genetista sabedor y consciente de la influencia de la herencia sobre las enfermedades y el estadístico, quien perfeccionó la aleatorización como el método de obtención de conclusiones de causalidad en el contexto experimental.

Las limitaciones de la metodología de estudios de casos y controles son también conocidas (1). Ellos señalan como limitaciones: 1) sesgo de selección, frecuentemente causado por la alta tasa de no participación, así como controles no representativos de la población en riesgo; 2) errores de medición de exposición; una gran mayoría de trabajos publicados tiene como instrumento de medición de la exposición una fuente primaria o el cuestionario, sin una forma objetiva de medición; 3) sesgo debido a factores de confusión, que no son tomados en cuenta en el estudio.

Las limitaciones señaladas, así como muchas otras que no son mencionadas, sin duda serán materia de trabajos de investigación futuros de la estadística.

### BIBLIOGRAFÍA

1. Austin H, Hill HA, Flanders WD. Limitations in the applications of case-control methodology. *Epidemiol Rev* 1994; 16: 65-76.
2. Broders AC. Squamous-cell epithelioma of the lip. *JAMA* 1920; 74: 656-64.
3. Breslow N. Regression analysis of the log odds-ratio: A method for retrospective studies. *Biometrics* 1976; 32: 409-16.
4. Breslow NE, Day NE. *Statistical Methods in Cancer Researches*, vol. 1. *The Analysis of Case-Control Studies*. IARC Lyon:Scientific Publications N° 32 International Agency for Research on Cancer. 1980.
5. Breslow NE, Cain KC. Logistic regression for two stage case-control data. *Biometrika* 1988; 75: 11-20.
6. Breslow NE. Statistics in epidemiology: The case-control study. *J Am Statistic Ass* 1996; 91: 14-28.
7. Bullón L. Análisis de Tablas de Contingencia  $2 \times 2$  en la Investigación Clínica y Epidemiológica. *Anales Científicos - UNALM*. 1998.
8. Carroll RJ, Gail MH, Lubin JH. Case-control studies with errors in covariates. *J Am Statistic Ass* 1993; 88: 185-99.
9. Cochran WG. Some methods for strengthening the common chi-square tests. *Biometrics* 1954; 10: 417-51.
10. Cole P. The evolving case-control study. *J Chron Dis* 1979; 32: 15-27.
11. Cornfield J. A method for estimating comparative rate from clinical data; Applications to cancer of the lung, breast and cervix. *J Ntl Cancer Inst* 1951; 11: 1269-75.
12. Cornfield J. A statistical problem arising from retrospective studies. *Proceedings of the Third Berkeley Symposium on*

- Mathematical Statistics and Probability. University of California Press. 1956: 135-48.
13. **Cornfield J, Haenszel W, Hammond EC, Lilienfield AM, Shimkin MB, Winder EL.** Smoking and lung cancer: Recent evidence and a discussion of some questions. *J Ntl Cancer Inst* 1959; 22: 173-203.
  14. **Dorn HF.** Some problems arising in prospective and retrospective studies of the etiology of disease. *New Engl J Med* 1959; 261: 571-9.
  15. **Fisher RA.** On the mathematical foundations of theoretical statistics. *Philosoph Transact Royal Soc London* 1922; 222: 309-68.
  16. **Fisher RA.** The logic of inductive inference. *J Royal Statis Soc* 1935; 98: 39-54.
  17. **Fisher RA.** Dangers of cigarette smoking. *BMJ* 1957a; July 6: 43.
  18. **Fisher RA.** Dangers of cigarette smoking. *BMJ* 1957b; August 3: 297-8.
  19. **Fisher RA.** Lung cancer and cigarettes. *Nature* 1958a; 182: 108.
  20. **Fisher RA.** Cancer and smoking. *Nature* 1958b; 182: 596.
  21. **Gart JM.** Point and interval estimation of the common odds ratio in the combination of  $2 \times 2$  tables with fixed marginals. *Biometrika* 1970; 57: 471-5.
  22. **Hosmer DW, Lemeshow S.** Applied Logistic Regression, New York:John Wiley. 1989.
  23. **Lilienfield AM, Lilienfield DE.** A century of case-control studies: Progress? *J Chron Dis* 1979; 32: 5-13.
  24. **Lane-Claypon JE.** A further report on breast cancer. *Reports on Public Health and Medical Subjects*, 32, London:Ministry of Health. 1926.
  25. **Langholz B, Thomas DC.** Efficiency of cohort sampling designs: some surprising results. *Biometrics* 1991; 47: 1563-71.
  26. **Lombard HL, Doering CR.** Cancer studies in Massachusetts. Habits, characteristics and environment of individuals with and without cancer. *New Engl J Med* 1928; 198: 481-7.
  27. **Mantel N, Haenszel W.** Statistical aspects of the analysis of data from retrospective studies of disease. *J Nat Cancer Inst* 1959; 22: 719-48.
  28. **Miettinen OS.** Estimation of relative risk from individually matched series. *Biometrics* 1970; 26: 75-86.
  29. **Miettinen OS.** Components of the crude risk ratio. *Am J Epidemiol* 1972; 96: 168-72.
  30. **Pike MC, Morrow RH.** Statistical analysis of patients-control studies in epidemiology: Factor under investigation an all-or-non variable. *Brit J Prevent Social Med* 1970; 24: 42-4.
  31. **Prentice RL.** A case-cohort design for epidemiologic cohort studies and diseases prevention trials. *Biometrika* 1986; 73: 1-11.
  32. **Rothman KJ.** Modern Epidemiology. Boston:Little Brown. 1986.
  33. **Schlesselman JJ.** Assessing effects of confounding variables. *Am J Epidemiol* 1978; 108: 3-8.
  34. **Schill W, Jöckel KH, Drescher K, Timm J.** Logistic analysis in case-control studies under validation sampling. *Biometrika* 1993; 80: 339-52.
  35. **Soares JF, Colosimo EA.** Métodos Estatísticos na Pesquisa Clínica, ICEX- UFMG, Ribeirão Preto-SP. 1995.
  36. **Surgeon General.** Smoking and Health Report of the Advisory Committee to the Surgeon General of the Public Health Service. Washington DC: Department of Health, Education and Welfare. 1964.
  37. **White JE.** A two stage design for the study of the relationship between a rare exposure and a rare disease. *Am J Epidemiol* 1982; 115: 119-28.