

Epidemiología aplicada a los Grados de Medicina, Nutrición y Odontología

**Ejercicios resueltos con software libre:
Open Epi y R**

Javier Santabárbara Serrano

*EPIDEMIOLOGÍA APLICADA A LOS GRADOS DE
MEDICINA, NUTRICIÓN Y ODONTOLOGÍA
Ejercicios resueltos con software libre: Open Epi y R*

*EPIDEMIOLOGÍA APLICADA A LOS GRADOS DE
MEDICINA, NUTRICIÓN Y ODONTOLOGÍA
Ejercicios resueltos con software libre: Open Epi y R
Javier Santabárbara Serrano*

Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra solo puede ser realizada con la autorización de sus titulares, salvo excepción prevista por la ley. Dirijase a CEDRO (Centro Español de Derechos Reprográficos, www.cedro.org) si necesita fotocopiar o escanear algún fragmento de esta obra.

- © Javier Santabábara Serrano
- © De la presente edición, Prensas de la Universidad de Zaragoza
(Vicerrectorado de Cultura y Patrimonio)
1.ª edición, 2025

Colección de Textos Docentes, n.º 319

Prensas Universitarias de Zaragoza. Edificio de Ciencias Geológicas, c/ Pedro Cerbuna, 12, 50009 Zaragoza, España. Tel.: 976 761 330
puz@unizar.es <http://puz.unizar.es>



Esta editorial es miembro de la UNE, lo que garantiza la difusión y comercialización de sus publicaciones a nivel nacional e internacional.

ISBN 978-84-1340-539-1

Impreso en España

Imprime: Servicio de Publicaciones. Universidad de Zaragoza

D.L.: Z 1067-2025

Introducción

El material incluido en esta obra forma parte de los apuntes elaborados para los alumnos de los Grados en Medicina, Nutrición y Odontología de la Universidad de Zaragoza. Pero también deseo que sea de utilidad para aquellos estudiantes del Master en Iniciación a la Investigación en Medicina de la Universidad de Zaragoza e interesados en Salud Pública y Epidemiología en general. Por último, señalar que dado el carácter eminentemente práctico de esta monografía son necesarios conocimientos básicos de Epidemiología y Bioestadística. Además, deseo hacer notar que no siempre los resultados calculados con la aplicación directa de la fórmula coinciden con exactitud con los reportados por los programas informáticos debido a los redondeos aplicados en los cálculos.

¿Por qué utilizamos Open Epi?

El escenario «ideal» para el análisis de datos epidemiológicos (a nivel de Grado) es aquel en el que la información se encuentra dispuesta en una tabla 2x2, como veremos a lo largo del manual, y en este sentido, *Open Epi* está diseñado para el manejo de datos tabulados (a diferencia de otros *softwares* de amplio uso como *IBM SPSS*, *Stata* o *SAS*, entre otros). Además, tiene el valor añadido de ser un *software* de uso gratuito.

Se puede trabajar directamente en *Open Epi* desde la web: <http://www.openepi.com>.

¿Por qué utilizamos R?

El paquete estadístico *R* es un entorno informático estadístico que incluye herramientas de análisis de datos y generación de gráficas. Es software libre y funciona bajo Windows, MacOS y Linux. Es una herramienta ideal para docencia, ya que permite que los alumnos la descarguen y la utilicen fuera del entorno universitario con toda libertad y sin cargo económico alguno.

La última versión del programa *R* se puede descargar desde la página web <https://cran.r-project.org/bin/windows/base/>, donde también se puede encontrar toda la información sobre el contenido del programa. El archivo que se descarga es necesario descomprimirlo. Recomendamos visitar la página de

software libre de la Universidad de Zaragoza <<http://oshuz.unizar.es/proyectos/r>> donde podréis descargaros el paquete, manuales, tutoriales, listas de correo...

Para aquellos que desean realizar el viaje iniciático al paquete estadístico *R* recomiendo la lectura de Bouso (2013) y, para los más osados, que se animen a programar en este magnífico *software* recomiendo Braun y Murdoch (2007).

Para la resolución de los primeros ejercicios del capítulo 0 (estudios transversales) he programado varias funciones *ad hoc*, y para el cálculo de las medidas de frecuencia, asociación e impacto potencial presentadas en el resto del capítulo 0 y los capítulos 1, 2 y 3 he utilizado la biblioteca *epiR*.

0. Estudios transversales

0.1. Conceptos básicos

Los *estudios transversales* son estudios de tipo *observacional*, descriptivos o analíticos, en los que la información recogida se refiere a un momento puntual en el tiempo. Todas las variables se miden a la vez.

También se denominan *estudios de prevalencia* y de *asociación cruzada*. De prevalencia porque su finalidad principal es la estimación de la prevalencia (proporción) de una enfermedad o de una característica dada de la población. De asociación cruzada porque también se pueden utilizar para estudiar la relación entre dos o más variables de una población en un momento dado, pero sin asumir necesariamente la existencia de una *relación causa-efecto*, dado que ambas variables se miden de forma simultánea.

Dependiendo del tipo de variables que se estén estudiando, los resultados pueden expresarse de diversas maneras. Así, si se trata de una variable de tipo cualitativo, los resultados se presentan generalmente en porcentajes (proporciones o *prevalencias*) y si son de tipo cuantitativo, a través de medias, desviaciones típicas... como ya vimos en la asignatura de Bioestadística.

El principal objetivo de los estudios transversales es describir la población, pero también sirven para estimar parámetros poblacionales y contrastar hipótesis sobre los valores que estos pueden tomar mediante el estudio de muestras. Por tanto, este diseño lleva implícito que la muestra en la que se realizan las estimaciones y los contrastes de hipótesis sea representativa de la población de la que ha sido extraída.

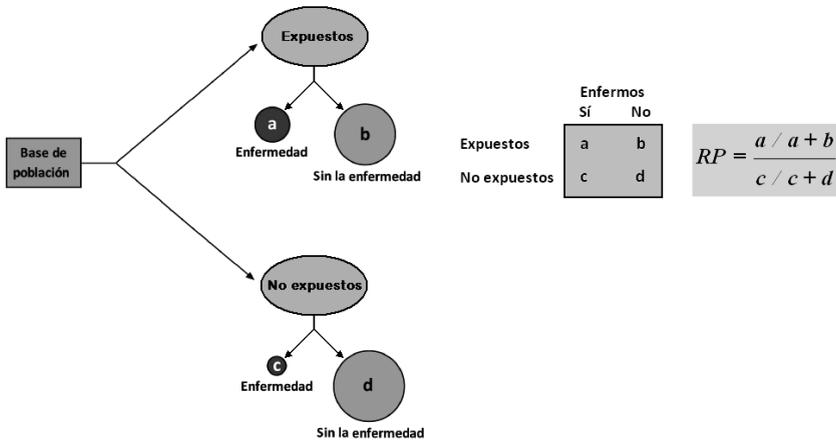


FIGURA 0.1. Esquema de un estudio transversal

0.2. Características

El diseño de los estudios epidemiológicos para la investigación médica se sustenta en la relación que existe entre posibles determinantes y una condición de salud, que en el caso de la medicina es usualmente una enfermedad. Si la exposición es decidida por el investigador se trata de un experimento. Si por el contrario, el investigador solo observa el efecto de una variable sobre un resultado de salud, entonces el diseño es de tipo *observacional*. El *estudio transversal* es un estudio *observacional* que mide tanto la exposición como el resultado *en un punto determinado en el tiempo*. Este diseño se lleva a cabo usualmente para estimar la *prevalencia* de una enfermedad dentro de una población específica y es común que uno de sus propósitos centrales sea proporcionar información para efectuar intervenciones de Salud Pública. Hay dos tipos de estudios transversales, los de tipo *descriptivo* que caracterizan la prevalencia de una enfermedad en la población de interés, y los *analíticos* que examinan la relación entre la exposición y la enfermedad, aunque sin poder establecer juicios definitivos de causalidad debido a la ambigüedad temporal que ocurre al recoger simultáneamente información acerca de los factores de riesgo y el resultado de interés.

En cuanto a la *temporalidad*, se trata de un estudio *retrospectivo*, lo que quiere decir que ya ha ocurrido el evento bajo estudio, esto es, que está presente la enfermedad y es posible identificarla en la muestra.

0.3. Ventajas y desventajas

Es preciso tener en cuenta las ventajas y limitaciones/desventajas de este tipo de estudio para llevar a cabo una buena investigación.

Las *ventajas* son las siguientes:

1. En comparación con el resto de diseños, que veremos en los siguientes capítulos, son estudios de bajo costo y fáciles de realizar.
2. Requieren poco tiempo de ejecución puesto que no hay seguimiento de los individuos.
3. La información se puede obtener a través de cuestionarios o es factible utilizar datos de fuentes secundarias.
4. Generan información de alto valor administrativo para la gestión y planificación de salud porque identifican las necesidades de salud y a los grupos vulnerables.
5. Proporcionan estimadores de la prevalencia (de la enfermedad y/o de la exposición a los factores de riesgo estudiados).
6. Permiten estudiar varias enfermedades o exposiciones, lo que los hace útiles para generar hipótesis etiológicas que pueden ser evaluadas con otros diseños.
7. Se puede tener un buen control de la selección de los sujetos de estudio.

Entre las *desventajas*, se pueden listar:

1. Por lo general no se establece relación causal, pues no se sabe si se presentó primero la causa o el efecto, ya que ambos se miden de manera simultánea; a esto se le conoce como ambigüedad temporal.
2. Solo es posible hacer conjeturas tentativas sobre factores presentes al momento del estudio.
3. Cuando se identifica alguna asociación estadísticamente significativa y tiene sustento teórico a esta variable, se le denomina *factor asociado* y no factor de riesgo como en otros diseños.
4. La información de la exposición es vulnerable a errores de medición, en especial si es recogida de modo retrospectivo.
5. Tal vez haya sobrerrepresentación de enfermos con larga supervivencia o con mejor curso clínico, por tanto, no son representativos de todos los casos.
6. No es eficaz para estudiar enfermedades raras, letales o de corta supervivencia, salvo que se estudien muestras grandes (encuestas nacionales).
7. Las medidas de impacto permiten determinar prioridades, no metas.

0.4. Sesgos

Al realizar estudios transversales se deben tener en cuenta los sesgos que puedan existir.

Sesgo de selección. Ocurre cuando las personas bajo estudio son incluidas por participación voluntaria y no en forma aleatoria.

Sesgo de duración. Es la sobrerrepresentación de los casos cuando la enfermedad es de larga duración y tiende a infraestimar casos cuando la enfermedad es de corta duración.

Sesgo de información. Los datos son poco verídicos o están incompletos; también se genera cuando la persona trata de complacer al entrevistador dándole la respuesta que cree será aprobada.

Sesgo de memoria. Cuando se hacen preguntas sobre exposiciones o eventos pasados, las personas que han sufrido una experiencia traumática que marcó su vida tal vez recuerden las exposiciones más que quienes no tuvieron dicha experiencia.

0.5. Análisis

0.5.1. Medidas de frecuencia

Las medidas de frecuencia sirven para describir los fenómenos de salud y enfermedad (contribuyendo al análisis de la situación de salud de una población), y, por otro lado, permiten, al asociarlas, formular y contrastar hipótesis.

Los cálculos necesarios para obtenerlas en un estudio de tipo transversal se ofrecen en la siguiente tabla (ver Figura 0.1):

	Enfermedad	Sin enfermedad	TOTAL	Prevalencia de la enfermedad
Expuesto	a	b	$a + b$	$\frac{a}{a + b}$
No expuesto	c	d	$c + d$	$\frac{c}{c + d}$
TOTAL	$a + c$	$b + d$		
Prevalencia de la exposición	$\frac{a}{a+c}$	$\frac{b}{b+d}$		

TABLA 0.1. Análisis de un estudio transversal.

Donde en cada celda se encuentran el número de individuos que:

- a : Están expuestos al factor y están enfermos.
- b : Están expuestos al factor de estudio pero no están enfermos.
- c : No están expuestos, pero sí están enfermos.
- d : No están expuestos y no están enfermos.

Y en los marginales tendremos:

- $a+b$: Total de individuos expuestos.
- $c+d$: Total de individuos no expuestos.
- $a+c$: Total de individuos enfermos.
- $b+d$: Total de individuos no enfermos.
- $a+b+c+d$: Total de la muestra.

En primer lugar, definimos la *prevalencia de la enfermedad* (P) como el cociente entre la totalidad de enfermos en la muestra expuestos ($a + c$) y el tamaño total de la muestra ($a + b + c + d$).

$$P = \frac{a + c}{a + b + c + d} \tag{0.1}$$

Del mismo modo, podemos calcular, por ejemplo, la *prevalencia de exposición* $p_e = \frac{a+b}{a+b+c+d}$, de utilidad en el cálculo de medidas de impacto potencial, o específicamente, la *prevalencia de la exposición entre los casos de enfermedad* al factor $p_{ce} = \frac{a}{a+c}$.

Importante interés merecen la prevalencia de enfermedad en expuestos y no expuestos. Así, comenzamos con un grupo expuesto y un grupo no expuesto. De la totalidad de los individuos expuestos ($a + b$) la enfermedad está presente en a de ellos pero no en b . Luego la *prevalencia de la enfermedad en los expuestos* es $P_e = \frac{a}{a+b}$.

De manera similar, de la totalidad de los individuos no expuestos ($c + d$) del estudio, la enfermedad está presente en c de ellos pero no en d . Luego la *prevalencia de la enfermedad entre los no expuestos* es $P_0 = \frac{c}{c+d}$.

Estas medidas se resumen en la Tabla 0.2:

Medidas	Denominación	Fórmula	Unidades	Interpretación
Frecuencia	Prevalencia de enfermedad en expuestos	$P_e = \frac{a}{a+b}$	Casos de enfermedad por 100 expuestos	Riesgo (probabilidad) que tiene un expuesto de padecer la enfermedad.
	Prevalencia de enfermedad en no expuestos	$P_0 = \frac{c}{c+d}$	Casos de enfermedad por 100 no expuestos	Riesgo (probabilidad) que tiene un no expuesto de padecer la enfermedad.

TABLA 0.2. Esquema medidas de frecuencia en estudios transversales

Adicionalmente, podemos calcular la *prevalencia de exposición entre los enfermos* (p_{ce}) y *entre los no enfermos* (p_{0e}), que nos serán de utilidad para el cálculo de las *medidas de impacto potencial* en los capítulos dedicados a los estudios de casos y controles y de cohortes (ver capítulos 1 y 2):

$$p_{ce} = \frac{a}{a+c} \quad \text{y} \quad p_{0e} = \frac{b}{b+d} \quad (0.2)$$

0.5.2. Medidas de asociación

Las medidas de asociación, como la *razón de prevalencias (RP)* (u *odds ratio (OR)*), comparan una medida de frecuencia de enfermedad (incidencia o prevalencia) en expuestos respecto a no expuestos con el fin de estimar la magnitud de la asociación entre la exposición y el efecto (desenlace, resultado, enfermedad). En principio, y como dictan los criterios de Bradford Hill, cuanto mayor sea la

magnitud de la asociación más probable es que se trate de una relación causal; aunque, como ha descrito Rothman (1982) en su *modelo determinista modificado de causalidad*, es posible que este criterio no se cumpla siempre (puede existir sesgo de confusión, por ejemplo).

Las medidas de asociación en estudios transversales se basan en comparar prevalencias: la prevalencia de la enfermedad en las personas que se expusieron al factor estudiado (P_e) con la prevalencia de la enfermedad en las personas que no se expusieron al factor estudiado (P_0).

Así, la medida más utilizada para cuantificar la asociación entre la exposición y la presencia de enfermedad (generalmente *aguda*) es la *razón de prevalencias (RP)* y se define como el cociente entre la prevalencia de la enfermedad entre expuestos y no expuestos al factor en estudio:

$$RP = \frac{P_e}{P_0} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} \quad (0.3)$$

La *RP* es una *aproximación* al *riesgo relativo (RR)* obtenido en un estudio de cohortes (ver capítulo 2) siempre que la enfermedad no influya en el hecho de estar en el estudio y tampoco en el hecho de estar expuesto. Por tanto, se interpreta de la siguiente manera:

- $RP = 1$, se dice que *no* existe asociación entre la enfermedad y la exposición al factor.
- $RP > 1$, se dice que existe una *asociación positiva* entre la enfermedad y la exposición al factor (de *riesgo*). El valor se interpretará como el número de veces que aumenta el riesgo de enfermedad entre los que están expuestos respecto a los que no están expuestos.
- $RP < 1$, se dice que posiblemente el factor estudiado sea un factor *protector* para la enfermedad (siempre y cuando se cumplan los criterios de causalidad).

Con el objeto de valorar si la *asociación* encontrada es *estadísticamente significativa*, calcularemos el *intervalo de confianza* (generalmente del 95%) para la *RP* y el *valor - p* del contraste $H_0: RP = 1$ vs. $H_1: RP \neq 1$ con un *nivel de significación* (α , generalmente, de 0.05) apoyándonos en algún *software* estadístico (*R*, *Open Epi*...).

Índice

INTRODUCCIÓN.....	7
0. ESTUDIOS TRANSVERSALES.....	9
0.1. Conceptos básicos.....	9
0.2. Características.....	10
0.3. Ventajas y desventajas.....	11
0.4. Sesgos.....	12
0.5. Análisis.....	12
0.6. Ejercicios resueltos.....	21
0.7. Ejercicios propuestos.....	42
1. ESTUDIOS DE CASOS Y CONTROLES.....	43
1.1. Conceptos básicos.....	43
1.2. Ventajas y desventajas.....	44
1.3. Sesgos.....	46
1.4. Análisis.....	48
1.5. Ejercicios resueltos.....	51
1.6. Ejercicios propuestos.....	62
2. ESTUDIOS DE COHORTES.....	63
2.1. Conceptos básicos.....	63
2.2. Ventajas y desventajas.....	64
2.3. Sesgos.....	65
2.4. Análisis.....	66
2.5. Ejercicios resueltos.....	77
2.6. Ejercicios propuestos.....	92
3. ENSAYOS CLÍNICOS.....	95
3.1. Conceptos básicos.....	95
3.2. Análisis.....	96
3.3. Ejercicios resueltos.....	98
3.4. Ejercicios propuestos.....	105
BIBLIOGRAFÍA.....	107

ISBN 978-84-1340-539-1



9 7884 13 40539 1

BIOMÉDICAS



colección
textos docentes



Prensas de la Universidad
Universidad Zaragoza

1474