

¿Odds ratio o razón de proporciones? Su utilización en estudios transversales

A. Schiaffino^a / M. Rodríguez^b / M.I. Pasarín^b / E. Regidor^c / C. Borrell^b / E. Fernández^a

^aServei de Prevenció i Control del Càncer. Institut Català d'Oncologia. Barcelona. España.

^bServei d'Informació Sanitària. Institut Municipal de Salut Pública de Barcelona. Barcelona. España.

^cDepartamento de Medicina Preventiva y Salud Pública. Universidad Complutense de Madrid. Madrid. España.

Correspondencia: Dra. Anna Schiaffino. Servei de Prevenció i Control del Càncer. Institut Català d'Oncologia. Avda. Gran Via, s/n, km. 2,7. 08907 L'Hospitalet de Llobregat. Barcelona. España.

Recibido: 29 de mayo de 2002.
Aceptado: 4 de noviembre de 2002.

(Odds ratio or prevalence ratio? Their use in cross-sectional studies)

Resumen

Antecedentes: En los estudios transversales las medidas de asociación clásicamente descritas son la razón de *odds* (*odds ratio*, OR) y la razón de prevalencias (*prevalence ratio*, PR). Algunos estudios epidemiológicos con diseño transversal expresan sus resultados en forma de OR, pero utilizan la definición de PR. El objetivo principal de este trabajo es describir y comparar diferentes métodos de cálculo de la PR discutidos en la bibliografía reciente en dos escenarios (prevalencia < 20% y prevalencia > 20%).

Material y métodos: Se realizó una búsqueda bibliográfica para conocer las técnicas más utilizadas para la estimación de la PR. Los 4 procedimientos más empleados fueron: a) seguir obteniendo OR mediante regresión logística no condicional, pero utilizando su definición correcta de OR; b) utilizar una regresión de Breslow-Cox; c) utilizar un modelo lineal generalizado con la transformación logaritmo y familia binomial, y d) utilizar una fórmula de conversión de una OR, obtenida mediante regresión logística tradicional, a una PR. Se han replicado para cada uno de los dos escenarios (prevalencia < 20% y prevalencia > 20%) los modelos hallados utilizando datos reales de la Encuesta de Salud de Catalunya de 1994.

Resultados: No se observan grandes diferencias entre las estimaciones ni entre los errores estándar obtenidos al utilizar una u otra técnica cuando la prevalencia es baja. Cuando la prevalencia es alta existen diferencias entre los estimadores y entre los intervalos de confianza, aunque todas las medidas mantienen la significación estadística.

Conclusión: Todos los métodos propuestos tienen sus pros y sus contras, y debe ser el propio investigador/a quien escoja la técnica que mejor se adapte a sus datos y ser coherente a la hora de utilizar un estimador y su interpretación.

Palabras clave: *Odds ratio*. Razón de prevalencias. Estudio transversal.

Abstract

Background: The most commonly used measures of association in cross-sectional studies are the odds ratio (OR) and the prevalence ratio (PR). Some cross-sectional epidemiologic studies describe their results as OR but use the definition of PR. The main aim of this study was to describe and compare different calculation methods for PR described in literature using two situations (prevalence < 20% and prevalence > 20%).

Material and methods: A literature search was carried out to determine the most commonly used techniques for estimating the PR. The four most frequent methods were: 1) obtaining the OR using non-conditional logistic regression but using the correct definition; 2) using Breslow-Cox regression; 3) using a generalized linear model with logarithmic transformation and binomial family, and 4) using the conversion formula from OR into PR. The models found were replicated for both situations (prevalence less than 20% and greater than 20%) using real data from the 1994 Catalan Health Interview Survey.

Results: When prevalence was low, no substantial differences were observed in either the estimators or standard errors obtained using the four procedures. When prevalence was high, differences were found between estimators and confidence intervals although all the measures maintained statistical significance.

Conclusion: All the methods have advantages and disadvantages. Individual researchers should decide which technique is the most appropriate for their data and should be consistent when using an estimator and interpreting it.

Key words: Odds ratio. Prevalence ratio. Cross-sectional study.

Introducción

En los estudios epidemiológicos de diseño transversal las medidas de asociación clásicamente descritas son la razón de *odds* (*odds ratio*, OR) y la razón de prevalencias (*prevalence ratio*, PR)¹. Estas dos medidas muestran el grado de asociación que existe entre una enfermedad o condición de interés y cierta exposición, pero difieren notablemente en su interpretación. La PR se define en términos de cuántas veces es más probable que los individuos expuestos presenten la enfermedad o condición respecto a aquellos individuos no expuestos². En cambio, la OR se define como el exceso o defecto de ventaja («*odds*») que tienen los individuos expuestos de presentar la enfermedad o condición frente a no padecerla respecto a la ventaja de los individuos no expuestos de presentar la condición frente a no presentarla².

En los últimos 15 años, se puede encontrar resultados de estudios epidemiológicos y clínicos con diseño transversal expresados en forma de OR, pero que utilizan la definición de PR. Este posible error conceptual se ha visto favorecido por la dificultad de ajustar la PR por múltiples variables, mientras que por su relativa sencillez, una alternativa comúnmente utilizada ha sido calcular la OR mediante modelos de regresión logística como aproximación de la PR. Esta aproximación podría ser correcta cuando la prevalencia de la enfermedad o condición de interés que se está estudiando es pequeña (clásicamente conocido como «asunción de enfermedad rara»)¹, ya que la OR daría un valor próximo a la PR¹. No hay un consenso sobre qué se entiende por prevalencia baja, diferentes autores han propuesto distintos valores que van del 10 hasta el 30%^{1,3,4}. Sin embargo, el problema se presenta cuando no se da esta condición. En este caso la OR aparentemente sobreestima la asociación, pero lo que sucede en realidad es que la OR se mueve en otra escala de medida, ya que no compara proporciones sino *odds*². El objetivo principal de esta nota metodológica es describir, comparar y discutir diferentes métodos de cálculo de la PR aparecidos en la bibliografía durante los últimos años en las dos situaciones expuestas (prevalencia < 20% y prevalencia > 20%), mediante la aplicación en un conjunto de datos reales derivados de un estudio transversal.

Técnicas utilizadas para medir la razón de prevalencia

Se realizó una búsqueda bibliográfica exhaustiva para conocer las técnicas más utilizadas para la estimación de la PR. Los 4 procedimientos más emplea-

dos según la búsqueda realizada fueron: a) seguir obteniendo OR mediante regresión logística no condicional, pero utilizando su definición correcta de OR^{5,6}; b) utilizar una regresión de Breslow-Cox⁷⁻⁹; c) utilizar un modelo lineal generalizado con la conversión logaritmo y familia binomial (log-binomial)¹⁰⁻¹³, y d) utilizar una fórmula de conversión de una OR, obtenida mediante regresión logística tradicional, a una PR^{6,14}. En la bibliografía se plantean otros métodos de obtención de la PR, pero son defendidos actualmente por una minoría, como el análisis GEE (*Generalized Estimated Equations*)¹⁵ o modelos lineales generalizados usando otra transformación y otra familia de distribuciones¹⁶, o utilizar la regresión de Poisson¹⁷.

A continuación se describen las características principales de la regresión logística, la regresión de Breslow-Cox, el modelo lineal generalizado log-binomial y la fórmula de conversión.

Regresión logística

El modelo de regresión logística proporciona la OR de contraer la enfermedad en aquellos individuos que han sufrido una exposición (X_E) respecto a aquellos individuos que no la han sufrido, ajustada por un conjunto de características (X_1, \dots, X_k). El modelo utiliza la transformación logit (logaritmo neperiano de la *odds*) para evitar que la función obtenida pueda tomar valores negativos:

$$\text{Logit} = \log(p/1-p) = \beta_0 + \beta_E X_E + \beta_1 X_1 + \dots + \beta_k X_k \quad [1]$$

El problema de este modelo es que no se obtienen PR, sino que se obtienen razones de ventajas entre expuestos y no expuestos. Su principal ventaja es que es un método bien conocido, fácil de aplicar y de evaluar con cualquier paquete estadístico y, además, posee buenas propiedades estadísticas.

Regresión de Breslow-Cox

El modelo de riesgos proporcionales de Cox, utilizado esencialmente en estudios de seguimiento, estima el riesgo relativo (RR) de padecer una enfermedad en aquellos individuos que han sufrido una exposición (X_E) respecto a aquellos individuos que no la han sufrido, ajustado por un conjunto de características (X_1, \dots, X_k) y que tiene en cuenta el tiempo de exposición a partir de la función de riesgo acumulado:

$$\text{Log}(H(t)) = \log(h_0(t)) + \beta_0 + \beta_E X_E + \beta_1 X_1 + \dots + \beta_k X_k \quad [2]$$

La propuesta para los estudios transversales es fijar el tiempo de seguimiento (t) como constante y utilizar

la medida de asociación obtenida como estimación de la PR¹⁸. La principal limitación de este modelo es que asume que los errores del modelo siguen una distribución de Poisson cuando, según la naturaleza de la variable, estos errores siguen realmente una distribución binomial. Esto puede producir estimaciones puntuales sesgadas e inconsistentes.

Modelo lineal generalizado con vínculo logarítmico y familia binomial

El modelo lineal generalizado estima la prevalencia de la enfermedad o condición en individuos que han sufrido una exposición (X_E) respecto a aquellos que no la han sufrido, ajustada por un conjunto de variables (X₁,..., X_k) mediante la conversión log-binomial¹⁹:

$$\text{Log}(p) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k \quad [3]$$

Para asegurar que las estimaciones de las proporciones están entre 0 y 1, es necesario imponer la restricción de que la suma de coeficientes para todos los posibles valores de las variables independientes sea inferior a 0; pero con los paquetes estadísticos utilizados habitualmente no es fácil introducir esta restricción. La principal limitación de este modelo es que esta combinación de conversión log-binomial puede no proporcionar la estimación de máxima verosimilitud y producir estimaciones erróneas cuando la suma de coeficientes es > 0.

Conversión de odds ratio a prevalence ratio

A partir de la OR estimada mediante regresión logística no condicional, se puede calcular la PR utilizando la fórmula de conversión [4], así como su intervalo de confianza (IC) mediante la fórmula propuesta por Miettinen [5].

$$PR = \frac{OR}{(1 + p_1 * [OR - 1])} \quad [4] \quad \text{IC del 95\%: } OR^{(1 \pm \frac{1.96}{z})} \quad [5]$$

Donde p₁ es la prevalencia de la enfermedad en el grupo de referencia (no expuestos) y z es el coeficiente de regresión dividido por su error estándar. La principal limitación es que la obtención del IC no es siempre exacta, sino solamente aproximada (sobre todo si la prevalencia en el grupo de los no expuestos es muy elevada).

Descripción de dos ejemplos

Para evaluar estas cuatro técnicas se han replicado para cada uno de los dos escenarios (prevalencia

< 20% y prevalencia > 20%) todos los modelos descritos utilizando datos reales de la Encuesta de Salud de Catalunya de 1994 (ESCA-94)²⁰. Brevemente, la ESCA-94 es un estudio transversal por entrevista de una muestra aleatoria y representativa de la población no institucionalizada de Catalunya. Para este análisis se han utilizado los datos de la región sanitaria de Barcelona ciudad (n = 1.828).

Se estima la asociación entre haber sido hospitalizado en el último año (6,9% de los entrevistados declaran haber sido hospitalizados durante los últimos 12 meses) y la clase social (recodificada como 0 = trabajadores no manuales y 1 = trabajadores manuales) ajustando por edad (15-44 años, 45-64 años, > 64 años) y sexo.

Por otro lado, se estima un modelo similar en el que la variable dependiente es padecer algún trastorno crónico (el 59,1% de los entrevistados declararon padecer algún trastorno crónico).

En los dos modelos se ha incluido el sexo como variable de ajuste al estar en una situación experimental, aunque en un análisis de género seguramente se debería estratificar por esta variable. Todos los análisis se han realizado con el paquete estadístico Stata versión 6.0²¹.

En la tabla 1 se observan los resultados en la situación donde la prevalencia de la enfermedad o condición de interés es baja (hospitalización en los últimos 12 meses). Se aprecia que no existen grandes diferencias entre las estimaciones ni entre los errores estándar obtenidos al utilizar una u otra técnica. Mediante regresión logística se obtiene una OR de 0,92 (IC del 95%, 0,63-1,34), mediante regresión de Breslow-Cox la estimación de la PR que se obtiene es de 0,93 (IC del 95%, 0,64-1,33), al utilizar un modelo log-binomial se obtiene una PR de 0,93 (IC del 95%, 0,66-1,32) y finalmente al utilizar la fórmula de conversión se estima una PR de 0,92 (IC del 95%, 0,65-1,32)

En la tabla 2 se observan los resultados cuando la prevalencia de la enfermedad es alta (padecer algún trastorno crónico). Mediante regresión logística se observa que los trabajadores manuales tienen una ventaja 1,59 veces superior (IC del 95%, 1,29-1,96) de tener algún trastorno crónico frente a no tenerlo respecto a la ventaja que tienen los trabajadores no manuales de

Tabla 1. Asociación entre haber estado hospitalizado durante el último año y la clase social. Estimaciones e intervalos de confianza (IC) del 95%

Técnicas	β	Error estándar (β)	Estimador ^{a,b}	IC del 95%
Regresión logística	-0,08	0,19	OR = 0,92	0,63-1,34
Regresión Breslow-Cox	-0,08	0,18	PR = 0,93	0,64-1,33
Modelo log-binomial	-0,07	0,18	PR = 0,93	0,66-1,32
Fórmula de conversión	-	-	PR = 0,92	0,65-1,32

^aDatos ajustados por edad y sexo. ^bClase social de referencia: trabajadores no manuales. OR: odds ratio; RR: riesgo relativo; PR: prevalence ratio.

Tabla 2. Asociación entre la presencia de algún trastorno crónico y la clase social. Estimaciones e intervalos de confianza (IC) del 95%

Técnicas	β	Error estándar (β)	Estimador ^{a,b}	IC del 95%
Regresión logística	0,46	0,11	OR = 1,59	1,29-1,96
Regresión Breslow-Cox	0,15	0,06	RR = 1,17	1,03-1,32
Modelo log-binomial	0,14	0,03	PR = 1,15	1,07-1,23
Fórmula de conversión	-	-	PR = 1,22	1,11-1,34

^aDatos ajustados por edad y sexo. ^bClase social de referencia: trabajadores no manuales. OR: odds ratio; RR: riesgo relativo; PR: prevalence ratio.

padecer algún trastorno crónico frente a no padecerlo. Al utilizar la regresión de Breslow-Cox los trabajadores manuales tienen un RR de 1,17 (IC del 95%, 1,03-1,32) de tener algún trastorno crónico en comparación con los no manuales. Si se usa la regresión log-binomial se obtiene que los trabajadores manuales tienen 1,15 veces (IC del 95%, 1,07-1,23) más probabilidad de padecer un trastorno crónico que los trabajadores no manuales. Finalmente, mediante la fórmula de conversión se obtiene que los trabajadores manuales tienen 1,22 veces (IC del 95%, 1,11-1,34) más probabilidad de padecer un trastorno crónico que los trabajadores no manuales. Se observa que, en esta situación, existen diferencias entre los estimadores y entre los IC hallados al usar los diferentes modelos, aunque todas las medidas mantienen la significación estadística.

Conclusión

Los resultados derivados de dos situaciones extremas, a partir de datos reales, muestran que según

la prevalencia de la enfermedad, la OR, aunque sigue siendo una buena medida de asociación en estudios transversales, puede no ser una buena aproximación de la PR, como algunas veces se está utilizando en el análisis, presentación e interpretación de resultados de estudios transversales. En los casos en los que la prevalencia es alta, la definición de la OR debería hacerse suficientemente explícita para evitar confundirla con la PR. Si, por el contrario, se quiere utilizar la PR, por su fácil comprensión e interpretación, la verdadera dificultad reside en escoger qué método es mejor para su estimación. Este tema ha creado en diferentes momentos un debate abierto que ha ocupado a metodólogos y epidemiólogos³⁻¹⁶. Algunos de los métodos propuestos aquí producen estimaciones sesgadas e inconsistentes (regresión de Breslow-Cox), por lo que de antemano no parece ser una buena alternativa. Del resto de métodos propuestos, todos tienen sus pros y sus contras, y debe ser el propio investigador/a quien escoja la técnica que mejor se adapte a sus datos y ser coherente a la hora de utilizar un estimador, y, sobre todo, estar muy atento a su interpretación¹⁴. Algunas sugerencias de procedimiento podrían ser las siguientes: a) seguir utilizando la OR calculada mediante regresión logística, aunque teniendo en cuenta su verdadera definición y no confundirla con la PR; b) calcular la OR y aplicar la fórmula de conversión, si las dos medidas no difieren se puede aproximar esta OR a una PR y utilizar la definición de esta última, y c) calcular la PR utilizando modelos log-binomiales, pero comprobando *a posteriori* que se cumple la restricción expuesta en el apartado sobre estos modelos. Sea cual fuere la elección, parece necesario que los investigadores/as se refieran con propiedad a la medida de asociación que escojan para el análisis de los estudios transversales sin caer en el fácil error de equiparar OR y PR.

Bibliografía

- Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research. Principles and quantitative methods. Belmont, CA: Lifetime Learning Publications; 1982.
- Martínez-González MA, De Irala-Estévez J, Guillén-Grima F. ¿Qué es una odds ratio? Med Clin (Barc) 1999;112:416-22.
- Axelsson O, Fredriksson M, Ekberg K. Use of prevalence ratio v prevalence odds ratio as a measure of risk in cross-sectional studies. Occup Environ Med 1994;51:574.
- Zocchetti C, Consonni D, Bertazzi PA. Relationship between prevalence rate ratio and odds-ratio in cross-sectional studies. Int J Epidemiol 1997;26:220-3.
- Stromberg U. Prevalence odds ratio v prevalence ratio. Occup Environ Med 1994;51:143-4.
- Stromberg U. Prevalence odds ratio v prevalence ratio - some further comments. Occup Environ Med 1995;52:143.
- Lee J, Chia KS. Estimation of prevalence rate ratios for cross-sectional data: an example in occupational epidemiology. Br J Ind Med 1993;50:861-4.
- Lee J. Odds ratio or relative risk for cross-sectional data? Int J Epidemiol 1994;23:201-3.
- Lee J, Chia KS. Prevalence odds ratio v prevalence ratio - a response. Occup Environ Med 1995;52:781-2.
- Wacholder S. Binomial regression in glm: estimating risks ratios and risk differences. Am J Epidemiol 1986;123:174-84.
- Zocchetti C, Consonni D, Bertazzi PA. Estimation of prevalence rate ratio from cross-sectional data. Int J Epidemiol 1995;24:1064-5.
- Skov T, Deddens J, Petersen MR, Endahl L. Prevalence proportion ratios: estimation and hypothesis testing. Int J Epidemiol 1998;27:91-5.
- Thompson ML, Myers JE, Kriebel D. Prevalence odds ratio

- or prevalence ratio in the analysis of cross-sectional data: What is to be done? *Occup Environ Med* 1998;55:272-7.
14. Osborn J, Cattaruzza MS. Odds ratio and relative risk for cross-sectional data. *Int J Epidemiol* 1995;24:464-5.
 15. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;73:13-22.
 16. Martuzzi M, Elliott P. Estimating the incidence rate ratio in cross-sectional studies using a simple alternative to logistic regression. *Ann Epidemiol* 1998;8:52-5.
 17. Breslow N E. Covariance analysis of censored survival data. *Biometrics* 1974;30:89-99.
 18. Nurminen M. To use or not to use the odds ratio in epidemiologic analyses. *Eur J Epidemiol* 1995;11:365-71.
 19. McCullagh P, Nelder JA. *Generalized linear models*. London: Chapman & Hall; 1989.
 20. Enquesta de Salut de Catalunya, ESCA-1994. Generalitat de Catalunya - Servei Català de la Salut.
 21. StataCorp. *Stata Statistical Software, Release 6.0*. College Station: Stata Corporation; 1999.
-