

Estimación de la Razón de Prevalencia con distintos modelos de Regresión: Ejemplo de un estudio internacional en investigación de las adicciones

Applied Prevalence Ratio estimation with different Regression models: An example from a cross-national study on substance use research

ALBERT ESPELT^{*,**,***,****}; MARC MARÍ-DELL'OLMO^{***,*,****}; EVA PENELO^{**}; MARINA BOSQUE-PROUS^{*,****}.

*Agència de Salut Pública de Barcelona, Plaça Lesseps 1, 08023, Barcelona, Spain; **Departament de Psicobiologia i de Metodologia de les Ciències de la Salut, Universitat Autònoma de Barcelona, 08193, Bellaterra, Spain; ***CIBER de Epidemiologia y Salud Pública (CIBERESP), Instituto de Salud Carlos III, Avda. Monforte de Lemos, 5, 28029, Madrid, Spain; ****Institut d'Investigació Biomèdica de Sant Pau (IIB Sant Pau), C/ Sant Antoni Maria Claret, 167, 08025, Barcelona, Spain.

Resumen

Objetivo: Examinar las diferencias entre la Razón de Prevalencia (RP) y la Odds Ratio (OR) en un estudio transversal y proporcionar herramientas para calcular la RP usando dos paquetes estadísticos ampliamente utilizados en la investigación de adicciones (STATA y R). *Métodos:* Se utilizaron los datos de un estudio transversal de 41.263 participantes de 16 países de Europa que participaron en la Encuesta sobre Salud y Envejecimiento en Europa (SHARE). La variable dependiente, consumo de riesgo de alcohol, se calculó a partir del Alcohol Use Disorders Identification Test – Consumption (AUDIT-C). Como principal variable independiente se utilizó el género. Otras variables fueron la edad, el nivel de estudios y el país de residencia. Las RP de consumo de riesgo de alcohol entre hombres y mujeres se estimaron a partir del método de Mantel Haenzel, de modelos de regresión log-binomial y de modelos de regresión de Poisson con varianza robusta. Estas estimaciones fueron comparadas con las OR obtenidas a partir de modelos de regresión logística. *Resultados:* La prevalencia de consumidores de riesgo de alcohol varía según país. En general los hombres tienen un mayor consumo de riesgo que las mujeres [RP=1.43 (1.38-1.47)]. La RP estimada no varía, independientemente del método o paquete estadístico utilizado. Sin embargo, dependiendo de la prevalencia del consumo de riesgo del país, la OR entre los consumidores de riesgo y el género sobrestima la RP. *Conclusiones:* En estudios transversales en los que se comparan distintos países con diferente prevalencia de una determinada enfermedad o condición es recomendable utilizar la RP en lugar de la OR.

Palabras clave: Regresión de Poisson; Regresión Log-binomial; Razón de Prevalencia; Odds Ratio; Estudios transversales.

Abstract

Objective: To examine the differences between Prevalence Ratio (PR) and Odds Ratio (OR) in a cross-sectional study and to provide tools to calculate PR using two statistical packages widely used in substance use research (STATA and R). *Methods:* We used cross-sectional data from 41,263 participants of 16 European countries participating in the Survey on Health, Ageing and Retirement in Europe (SHARE). The dependent variable, hazardous drinking, was calculated using the Alcohol Use Disorders Identification Test – Consumption (AUDIT-C). The main independent variable was gender. Other variables used were: age, educational level and country of residence. PR of hazardous drinking in men with relation to women was estimated using Mantel-Haenszel method, log-binomial regression models and poisson regression models with robust variance. These estimations were compared to the OR calculated using logistic regression models. *Results:* Prevalence of hazardous drinkers varied among countries. Generally, men have higher prevalence of hazardous drinking than women [PR=1.43 (1.38-1.47)]. Estimated PR was identical independently of the method and the statistical package used. However, OR overestimated PR, depending on the prevalence of hazardous drinking in the country. *Conclusions:* In cross-sectional studies, where comparisons between countries with differences in the prevalence of the disease or condition are made, it is advisable to use PR instead of OR.

Keywords: Poisson regression; Log-binomial regression; Prevalence Ratio; Odds Ratio; Cross-sectional studies.

Recibido: Abril 2015; *Aceptado:* Julio 2015

Enviar correspondencia a:

Albert Espelt. Agència de Salut Pública de Barcelona. Plaça Lesseps, 1. 08023 Barcelona. Tel: 93.202.77.01
E-mail: aespelt@aspb.cat

Los diseños transversales se utilizan extensamente en la investigación de las adicciones. Las personas que investigan en el campo de las adicciones generalmente utilizan este tipo de diseño para estimar la asociación entre una variable dependiente dicotómica y una o más variables independientes. Aunque la Odds-Ratio (OR) o la Razón de Prevalencia (RP) podrían ser buenos estimadores de esta asociación, tradicionalmente la mayoría de los estudios ha utilizado la OR, calculada mediante la regresión logística, para estimar la asociación (Barros y Hirakata, 2003). La RP se define como la prevalencia en la población expuesta dividida por la prevalencia en la no expuesta, mientras que la OR se define como la odds de la enfermedad o condición entre las individuos expuestos dividido por la odds de la enfermedad o condición entre los no expuestos. En este sentido, en los diseños transversales, cuando la variable dependiente es dicotómica, generalmente obtenemos la prevalencia en el análisis descriptivo y, por lo tanto, la RP es más intuitiva y fácil de entender que la OR. Aunque la OR es un buen estimador de la RP cuando la prevalencia es baja, se sabe que la OR sobrestima la RP cuando la prevalencia es moderada o alta (por ejemplo, RPs superiores al 10%) (Szklo y Nieto, 2012). Esto podría ser un problema porque la OR generalmente ha sido tratada e interpretada como una RP, independientemente de la prevalencia de la enfermedad (por ejemplo, en un artículo sobre los predictores de la conducción bajo la influencia del alcohol entre los adolescentes españoles, las personas autoras consideraron la OR como probabilidades aunque la RP fue superior al 10% en algunas categorías) (Barlés-Arizona, Escario y Sánchez-Ventura, 2014). Por estas razones, se han ideado métodos alternativos para estimar la asociación entre una variable dependiente dicotómica y varias variables independientes en diseños transversales, obteniéndose RPs (Barros et al., 2003; Coutinho, Sczufca y Menezes, 2008; Deddens y Petersen, 2008; Schiaffino et al., 2003; Thompson, Myers y Kriebel, 1998). Uno de los métodos más simples consiste en utilizar la siguiente fórmula para calcular la RP a partir de una OR dada (Schiaffino et al., 2003):

$$PR = \frac{OR}{(1 + p_1 * [OR - 1])}$$

donde p_1 es la prevalencia de la enfermedad o condición en el grupo de referencia (no expuestos).

En este caso, aunque las estimaciones puntuales son correctas, podría haber un problema de estimación de los intervalos de confianza, especialmente si los modelos se han ajustado para muchas variables. El método intuitivo para calcular la RP sería utilizar la regresión log-binomial. Sin embargo, la regresión log-binomial a menudo tiene problemas de convergencia cuando alguna de las variables independientes es continua (Cummings, 2009; Deddens et al., 2008). Como consecuencia, se han estudiado métodos

alternativos de estimación (por ejemplo, los modelos de regresión de Cox o los modelos de regresión de Poisson) (Barros et al., 2003; Deddens et al., 2008). Además, aunque existe evidencia apoyando el uso de los modelos de regresión Poisson con varianza robusta para estimar la RP en estudios transversales (Barros et al., 2003; Coutinho et al., 2008; Deddens et al., 2008), la solución óptima sería utilizar un modelo de regresión log-binomial (Deddens et al., 2008) si éste convergiera. Pero si sólo consideramos un decimal, los resultados utilizando modelos de regresión diferentes no varían, independientemente de la prevalencia de la enfermedad o condición (Cummings, 2009). En este sentido, la estimación de la RP utilizando modelos de regresión Poisson con varianza robusta, basada en la estimación sándwich de Huber, ha demostrado ser correcta y robusta en diferentes situaciones experimentales, tales como el uso de diferentes valores de prevalencia (prevalencia baja, moderada o alta) o estimando varios modelos (crudos y ajustados) (Barros et al., 2003; Coutinho et al., 2008; Deddens et al., 2008).

Aunque parece haber una tendencia en los últimos años a usar la RP en vez de la OR en estudios transversales (Bosque-Prous et al., 2014; Espelt et al., 2013; Palencia et al., 2010), el conocimiento las personas investigadoras en el campo de las adicciones acerca de cómo llevar a cabo estos análisis tiende a ser escaso. Por esta razón, los objetivos de este artículo son: examinar las diferencias entre la RP y la OR en un estudio internacional y proveer las herramientas para calcular la RP utilizando modelos de regresión log-binomial y de Poisson con varianza robusta con dos paquetes estadísticos utilizados en la investigación de las adicciones [STATA y R (software libre)].

Métodos

Diseño y participantes

Se utilizó la base de datos del proyecto "Survey of Health, Ageing and Retirement in Europe" (SHARE; Encuesta de Salud, Envejecimiento y Jubilación en Europa) (Börsch-Supan et al., 2013). La población de estudio estaba formada por personas mayores de 50 años de 16 países europeos que participaron en la ola 4 (2010-2012) de SHARE ($n = 41,263$). Aunque la base de datos contenía pesos muestrales, no se usaron en este estudio ya que no se pretendía hacer un estudio poblacional. Por otra parte, se excluyeron los participantes con falta de información en cualquiera de las variables.

Variables

La variable dependiente fue la prevalencia de consumo de riesgo de alcohol, que se construyó usando una adaptación del "Alcohol Use Disorders Identification Test Consumption" (AUDIT-C, test de cribado para identificar el trastorno por consumo de alcohol) (Meneses-Gaya et al., 2010). Se basó en tres preguntas: dos que evaluaban el consumo regular de alcohol en términos de frecuencia y cantidad y

Tabla 1. Explicación de los pasos para estimar la RP utilizando modelos de regresión log-binomial o modelos de regresión de Poisson con varianza robusta en dos paquetes estadísticos (STATA y R), utilizando la barra de herramientas o la sintaxis específica.

Paquete	Barra de Herramientas (paso a paso)	Sintaxis modelo crudo	Sintaxis modelo ajustado
Modelo de regresión Log-Binomial			
	Menú herramientas→Estadísticos→Modelos Lineales Generalizados →Modelos Lineales Generalizados (GLM)→Modelos [(variable dependiente: auditc; variable independiente: sex)/(family: binomial; selecciones de link: log)]→Reportar [reportar coeficientes exponenciados.	glm auditc sex, family(binomial 1) link(log) eform	glm auditc sex educ age, family(binomial 1) link(log) eform
STATA		<pre> install.packages(pkgs = c("Epi", "foreign")) library(Epi) library(foreign) data<-read.dta("C:/BBDD.dta", convert.factors=F) model<-glm(auditc ~ sex, data=data, family=binomial(link=log)) summary(model) round(ci.lin(model, Exp=T), 2) </pre>	<pre> install.packages(pkgs = c("Epi", "foreign")) library(Epi) library(foreign) data<-read.dta("C:/BBDD.dta", convert.factors=F) model<-glm(auditc ~ sex + educ + age, data=data, family=binomial(link=log)) summary(model) round(ci.lin(model, Exp=T), 2) </pre>
R			
Modelo de regresión Poisson con varianza robusta			
	Menú herramientas→Estadísticos→Modelos Lineales Generalizados→Modelos Lineales Generalizados (GLM)→Modelo [(variable dependiente: auditc; variable independiente: sex)/(family: poisson; selecciones de link: log)]→SE/Robusto [tipo de error estándar: Robusto]→Reportar [reportar coeficientes exponenciados.	glm auditc sex, family(poisson) link(log) robust eform	glm auditc sex educ age, family(poisson) link(log) robust eform
STATA		<pre> install.packages(pkgs = c("Epi", "foreign", "sandwich", "lmtest")) library(Epi) library(foreign) library(sandwich) # to get robust estimators library(lmtest) # to test coefficients data<-read.dta("C:/BBDD.dta", convert.factors=F) model<-glm(auditc ~ sex, data=data, family=poisson(link=log)) summary(model) coef<-coefest(model, vcov = sandwich) ## Sex Coefficient B<-coef["sex", "Estimate"] ## Sex coefficient Standard Error SE<-coef["sex", "Std. Error"] ## PR point estimation exp(B) ## PR 95% Confidence Interval # upper 95% CI exp(B + qnorm(0.05 / 2) * SE) # lower 95% CI exp(B - qnorm(1 - 0.05 / 2) * SE) </pre>	<pre> install.packages(pkgs = c("Epi", "foreign", "sandwich", "lmtest")) library(Epi) library(foreign) library(sandwich) # to get robust estimators library(lmtest) # to test coefficients data<-read.dta("C:/BBDD.dta", convert.factors=F) model<-glm(auditc ~ sex + educ + age, data=data, family=poisson(link=log)) summary(model) coef<-coefest(model, vcov = sandwich) ## Sex Coefficient B<-coef["sex", "Estimate"] ## Sex coefficient Standard Error SE<-coef["sex", "Std. Error"] ## PR point estimation exp(B) ## PR 95% Confidence Interval # upper 95% CI exp(B + qnorm(0.05 / 2) * SE) # lower 95% CI exp(B - qnorm(1 - 0.05 / 2) * SE) </pre>
R			

una que evaluaba el consumo excesivo de alcohol (seis o más bebidas alcohólicas en una sola ocasión, por lo menos una vez al mes en los últimos 3 meses). Cada respuesta recibía entre 0 a 4 puntos, y la puntuación final se calculó como la suma de las puntuaciones de cada pregunta. El consumo de riesgo se construyó como una variable dicotómica (consumo de riesgo/no consumo de riesgo), considerando que era de riesgo cuando la puntuación era de 5 o más para los hombres y 4 o más para las mujeres (Gual, Segura, Contel, Heather y Colom, 2002) [nombre de la variable: *auditc*]. La variable independiente utilizada fue género [nombre de la variable: *sex*] y se emplearon dos covariables diferentes para el ajuste: edad, como variable continua [nombre de la variable: *age*] y nivel educativo (no estudios secundarios o estudios secundarios o terciarios), como una variable categórica [nombre de la variable: *educ*]. Finalmente, se tomó en cuenta el país de residencia, como una variable de estratificación.

Análisis

Se calculó la prevalencia del consumo de riesgo de alcohol por género en cada país, utilizando el programa estadístico STATA. La RP de ser un bebedor de riesgo en los hombres con respecto a las mujeres se estimó con el método de Mantel-Haenszel en STATA [sintaxis: `cs auditc sex`] y con modelos de regresión log-binomial y modelos de regresión de Poisson con varianza robusta, estratificado por país, en STATA y en R (Tabla 1). Para estimar los modelos de regresión de Poisson, es necesario disponer de datos individuales y satisfacer las dos condiciones siguientes para obtener una estimación puntual realista e intervalos de confianza de amplitud razonable (Barros et al., 2003). En primer lugar, la variable dependiente tiene que ser dicotómica con valores 0 y 1 (no pueden utilizarse otros valores) al estimar los modelos de Poisson. El valor 1 se asigna a los individuos con la enfermedad o condición (bebedores de riesgo en nuestro ejemplo) y 0 para el resto de los participantes. Y en segundo lugar, la varianza de las estimaciones debe ser robusta. Todos los modelos se realizaron utilizando Modelos Lineales Generalizados con las familias Poisson o binomial con función log-link.

Finalmente, también se calculó la asociación entre el género y el consumo de riesgo de alcohol para cada país usando modelos de regresión logística en STATA [`logit auditc sex, or`], que proporcionaban la OR. Se calcularon las sobrestimaciones de OR con respecto a la RP para cada país, utilizando la siguiente fórmula: [Sobrestimación = (OR-PR)/(OR-1)] (Brotman, 2006; Espelt et al., 2013; Shishehbor, Litaiker y Lauer, 2006). Se utilizaron STATA13.0 y R 3.0.2 para realizar todos los análisis.

Resultados

La Tabla 1 muestra los pasos para calcular la RP ajustando modelos de regresión log-binomial y modelos de regresión

de Poisson con varianza robusta a través de la barra de herramientas y la sintaxis específica, utilizando STATA y R. Los datos para llevar a cabo todos los análisis están disponibles en formato STATA (datos complementarios). Para obtener estos datos y realizar todos los análisis correctamente con el paquete estadístico R, el usuario debe tener previamente instaladas las bibliotecas "foreign", "Epi" y "sandwich" (Tabla 1). Para leer los datos en formato STATA en R, la instrucción es `data<- read.dta("C:/Users_directori/bbdd.dta", convert.factors=F)`.

La Tabla 2 muestra la prevalencia de consumo de riesgo en hombres y mujeres en cada país y las asociaciones entre las variables, calculadas usando STATA. La prevalencia de consumo de riesgo variaba de un país a otro. Por ejemplo, la prevalencia de consumo de riesgo en Eslovenia fue baja tanto en hombres como en mujeres (14% y 11%, respectivamente), mientras que fue alta en ambos sexos en Dinamarca (39% en hombres y 35% en mujeres), y en Estonia, fue moderada en los hombres (17%) pero baja en las mujeres (4%). Las estimaciones de la RP y sus intervalos de confianza al 95% (IC95%), calculadas usando STATA y R, fueron las mismas que las que se calcularon usando el método Mantel-Haenszel. Sin embargo, la OR sobreestimó la RP en casi todos los análisis. Por ejemplo, la RP de ser un bebedor de riesgo en los hombres con respecto a las mujeres en Austria fue de 1,49 (IC95%: 1,34-1,66) mientras que la OR correspondiente fue de 1,66 (IC95%: 1,45-1,90). Por otra parte, la RP era de 1,33 (IC95%: 1,22-1,46) en Francia, mientras que la OR era de 1,47 (IC95%: 1,30-1,66). Si la OR se interpretaba como una RP, la sobrestimación de la OR era alta en algunos países (por ejemplo, del 40% en Dinamarca o del 33% en Bélgica). El grado de esta sobrestimación dependía de la prevalencia de consumo de riesgo de alcohol entre los hombres y las mujeres en cada país. Sin embargo, cuando la prevalencia era similar para los hombres y las mujeres, no se observaron diferencias entre la RP y la OR en los Países Bajos y Suiza, mientras que se encontró un 33% de sobrestimación en Italia (Tabla 3).

En general, las RPs calculadas utilizando modelos de regresión log-binomial o modelos de regresión de Poisson con varianza robusta no varían entre sí en el análisis crudo. Sin embargo, en el análisis ajustado, controlando por la edad y el nivel educativo, se encontraron algunas diferencias en el segundo decimal. Las RPs obtenidas mediante diferentes paquetes estadísticos no fueron estadísticamente diferentes.

Discusión

Los resultados muestran que la utilización sistemática de la OR en lugar de la RP en estudios transversales no debería realizarse, sobre todo si la prevalencia de la enfermedad o condición es moderada o alta, dado que las RPs se calculan fácilmente y existen métodos para obtener estimaciones robustas de las RPs y sus IC95%. Por otra parte, nuestros

Tabla 2. Prevalencia sin ajustar, razón de prevalencia, odds ratio y sobreestimación de la OR con respecto a las estimaciones de la RP de consumo de riesgo de alcohol entre hombres y mujeres en varios países europeos.

	Hombres	Mujeres	Prevalencia de consumo de riesgo de alcohol		RP _{Hombres/Mujeres} Mantel-Haenszel		RP _{Hombres/Mujeres} log-binomial		RP _{Hombres/Mujeres} Poisson robusta		OR _{Hombres/Mujeres} regresión logística		Sobre-estimación*
	N	N	Hombres	Mujeres	RP	IC 95%	RP	IC 95%	RP	IC 95%	OR	IC 95%	%
Alemania	697	796	22,53	17,09	1,32	(1,07-1,62)	1,32	(1,07-1,62)	1,32	(1,07-1,62)	1,41	(1,09-1,82)	22,0%
Austria	2.159	2.945	25,61	17,18	1,49	(1,34-1,66)	1,49	(1,34-1,66)	1,49	(1,34-1,66)	1,66	(1,45-1,90)	25,8%
Bélgica	2.256	2.789	33,73	30,62	1,10	(1,02-1,19)	1,10	(1,02-1,19)	1,10	(1,02-1,19)	1,15	(1,02-1,30)	33,3%
Dinamarca	1.006	1.191	38,97	34,76	1,12	(1,00-1,25)	1,12	(1,00-1,25)	1,12	(1,00-1,25)	1,20	(1,01-1,43)	40,0%
Eslovenia	1.181	1.549	14,31	11,17	1,28	(1,05-1,56)	1,28	(1,05-1,56)	1,28	(1,05-1,56)	1,33	(1,06-1,67)	15,2%
España	1.510	1.878	18,34	12,51	1,47	(1,25-1,72)	1,47	(1,25-1,72)	1,47	(1,25-1,72)	1,57	(1,30-1,90)	17,5%
Estonia	2.692	4.030	16,75	4,24	3,95	(3,33-4,68)	3,95	(3,33-4,68)	3,95	(3,33-4,68)	4,54	(3,78-5,46)	16,7%
Francia	2.380	3.164	28,91	21,68	1,33	(1,22-1,46)	1,33	(1,22-1,46)	1,33	(1,22-1,46)	1,47	(1,30-1,66)	29,8%
Hungría	1.302	1.730	25,04	8,21	3,05	(2,54-3,66)	3,05	(2,54-3,66)	3,05	(2,54-3,66)	3,74	(3,02-4,62)	25,2%
Italia	1.577	1.940	25,94	25,31	1,02	(0,92-1,15)	1,02	(0,92-1,15)	1,02	(0,92-1,15)	1,03	(0,89-1,20)	33,3%
P. Bajos	1.148	1.469	32,32	32,81	0,98	(0,88-1,10)	0,98	(0,88-1,10)	0,98	(0,88-1,10)	0,98	(0,83-1,15)	0,0%
Polonia	651	874	14,59	2,75	5,31	(3,44-8,22)	5,31	(3,44-8,22)	5,31	(3,44-8,22)	6,05	(3,82-9,59)	14,7%
Portugal	857	1.129	31,74	20,99	1,51	(1,30-1,76)	1,51	(1,30-1,76)	1,51	(1,30-1,76)	1,75	(1,43-2,14)	32,0%
Rep. Checa	2.482	3.420	32,96	15,67	2,10	(1,91-2,31)	2,10	(1,91-2,31)	2,10	(1,91-2,31)	2,65	(2,34-3,00)	33,3%
Suecia	848	1.002	12,85	14,57	0,88	(0,70-1,11)	0,88	(0,70-1,11)	0,88	(0,70-1,11)	0,86	(0,66-1,13)	14,3%
Suiza	1.634	1.987	28,21	27,98	1,01	(0,91-1,12)	1,01	(0,91-1,12)	1,01	(0,91-1,12)	1,01	(0,87-1,17)	0,0%
Total	24.380	31.893	25,88	18,15	1,43	(1,38-1,47)	1,43	(1,38-1,47)	1,43	(1,38-1,48)	1,57	(1,51-1,64)	24,6%

Nota. *La sobrestimación de la OR con respecto a la RP se calculó con la fórmula: [Sobrestimación = (OR-RP)/(OR-1)] (Brotman, 2006; Espelt et al., 2013; Shishehbor, Litaker, y Lauer, 2006)

Tabla 3. Comparación de la razón de prevalencia ajustada, la odds ratio ajustada y la sobrestimación de la OR ajustada con respecto a las estimaciones ajustadas de la RP de consumo de riesgo de alcohol entre hombres y mujeres en varios países europeos.

	RP _{Hombres/Mujeres} log-binomial ¹		RP _{Hombres/Mujeres} Poisson robusta ¹		OR _{Hombres/Mujeres} regresión logística ¹		Sobreestimación*
	RP	IC 95%	RP	IC 95%	OR	IC 95%	%
Alemania	1,39	(1,13-1,71)	1,38	(1,12-1,70)	1,49	(1,15-1,94)	22,4%
Austria	1,48	(1,33-1,65)	1,48	(1,32-1,64)	1,64	(1,43-1,88)	25,0%
Bélgica	1,11	(1,02-1,20)	1,10	(1,01-1,19)	1,15	(1,02-1,30)	33,3%
Dinamarca	1,11	(0,99-1,24)	1,10	(0,99-1,23)	1,17	(0,98-1,39)	41,2%
Eslovenia	1,28	(1,05-1,57)	1,28	(1,05-1,56)	1,33	(1,06-1,67)	15,2%
España	1,52	(1,30-1,78)	1,51	(1,28-1,77)	1,63	(1,35-1,98)	19,0%
Estonia	3,87	(3,27-4,57)	3,87	(3,28-4,58)	4,67	(3,87-5,63)	21,8%
Francia	1,30	(1,19-1,43)	1,31	(1,19-1,44)	1,44	(1,27-1,62)	29,5%
Hungría	3,07	(2,56-3,69)	3,07	(2,55-3,68)	3,78	(3,05-4,68)	25,5%
Italia	1,05	(0,93-1,18)	1,05	(0,93-1,17)	1,06	(0,91-1,24)	16,7%
P. Bajos	1,01	(0,91-1,13)	1,01	(0,90-1,12)	1,01	(0,85-1,19)	0,0%
Polonia	5,77	(3,73-8,94)	5,77	(3,73-8,93)	6,91	(4,31-11,07)	19,3%
Portugal	1,60	(1,38-1,85)	1,59	(1,36-1,84)	1,88	(1,53-2,32)	33,0%
Rep. Checa	2,08	(1,89-2,29)	2,08	(1,89-2,30)	2,65	(2,33-3,01)	34,5%
Suecia	0,92	(0,73-1,15)	0,91	(0,72-1,15)	0,90	(0,69-1,18)	10,0%
Suiza	1,02	(0,91-1,13)	1,02	(0,91-1,13)	1,02	(0,88-1,18)	0,0%
Total	1,44	(1,40-1,49)	1,43	(1,38-1,49)	1,59	(1,53-1,66)	27,1%

Nota. *La sobrestimación de la OR con respecto a la RP se calculó con la fórmula: [Sobrestimación = (OR-RP)/(OR-1)] (Espelt et al., 2013). ¹Ajustado por edad y nivel de estudios.

resultados concuerdan con otros artículos publicados (Barros et al., 2003; Coutinho et al., 2008; Deddens et al., 2008; Schiaffino et al., 2003; Thompson et al., 1998). Según este estudio metodológico, los paquetes estadísticos usados en la mayoría de los estudios epidemiológicos permiten a los investigadores calcular la RP fácilmente. Sin embargo, si utilizamos los modelos de regresión de Poisson, tenemos que estar seguros de que hayamos utilizado métodos robustos para estimar su varianza, de lo contrario la regresión de Poisson produciría intervalos de confianza más amplios en comparación con un modelo de regresión log-binomial (McNutt, Wu, Xue y Hafner, 2003).

Una ventaja de usar la RP es que los resultados son mucho más intuitivos. Por ejemplo, la prevalencia de bebedores de riesgo en hombres y mujeres en Austria es del 25,6% y del 17,2%, respectivamente. Al dividir la prevalencia de los hombres por la prevalencia de las mujeres, obtenemos una RP de 1,49, que es la misma RP que se estimó utilizando los diferentes paquetes estadísticos. Por otra parte, encontramos que el grado de sobreestimación de la RP (usando la OR) variaba entre los países y dependía de la prevalencia de la enfermedad o condición (es decir, consumidor de riesgo de alcohol) en los participantes expuestos y no expuestos (en este estudio, los hombres eran considerados como expuestos y las mujeres como no expuestas). Por esa razón, si interpretamos la OR como una estimación de la RP, podríamos estar malinterpretando los resultados, como hemos visto en la sección de resultados. El hecho de que la OR podría sobreestimar la RP dependiendo de la prevalencia de la condición o enfermedad analizada en cada país conlleva que la OR sea similar a la RP en algunos países, mientras que en otros, las estimaciones de la OR y de la RP son muy diferentes. Como resultado, cuando se usan las OR para hacer comparaciones entre países, la interpretación de los resultados podría ser un problema para los investigadores que interpretan la OR intuitivamente como una RP. Por ello, en los estudios internacionales, donde se realizan comparaciones entre países con grandes diferencias en la prevalencia de la enfermedad o condición, es recomendable utilizar la RP en lugar de la OR. Es especialmente relevante porque, como hemos dicho, las personas generalmente interpretan la estimación de la OR como una RP. La sobreestimación puede afectar inapropiadamente la toma de decisiones clínicas o el desarrollo de políticas y por lo tanto, puede conducir a errores involuntarios en el análisis económico de posibles programas de intervención o tratamientos (McNutt et al., 2003).

En la actualidad, algunos estudios en el campo de las adicciones están comenzando a utilizar modelos de regresión para obtener las RPs como estimadores de la asociación entre una variable dependiente dicotómica y varias variables independientes. En este sentido, en la investigación de adicciones, algunos estudios han calculado la RP para estimar qué factores podrían estar asociados al consumo de drogas ilegales (Jamieson et al., 2010; Sarasa-Renedo et al., 2014) o

legales (Bosque-Prous et al., 2014; Font-Ribera et al., 2013; Jamieson et al., 2010). Sin embargo, el uso de métodos de regresión para estimar la RP es todavía escaso. Por ejemplo, si comparamos los estudios publicados en Pubmed en 2013 que han utilizado la RP o la OR utilizando las siguientes estrategias: PR = ([“cross-sectional”] y [“prevalence ratio” o “log-binomial” o “poisson regression model with robust variance”]); OR = ([“cross-sectional”] y [“odds ratio” o “logistic regression”]), encontramos 132 publicaciones que usaron la PR y 4886 que usaron la OR.

Una de las principales fortalezas de nuestro estudio es que se explica cómo calcular la RP utilizando diferentes modelos de regresión y también dos diferentes paquetes estadísticos (uno de ellos es software de libre acceso para todos los investigadores). Sin embargo, este estudio podría presentar algunas limitaciones. Su principal limitación es que no se diseñó como un estudio de simulación, utilizando diferentes condiciones para analizar los cambios en la RP con respecto a la OR. Sin embargo, esto no era el objetivo de este artículo. No obstante, al tratarse de un estudio basado en datos reales de consumo de alcohol de diferentes países, el ejemplo es más fácil de entender. De hecho, casi todos los escenarios se encuentran en los diferentes países participantes en el estudio (es decir, alta prevalencia en ambos sexos, baja prevalencia en ambos sexos, combinaciones de alta y baja prevalencia), reforzando nuestros resultados. Otra limitación es que sólo mostramos los modelos de regresión más frecuentemente utilizados para calcular la RP, con dos paquetes estadísticos diferentes, pero hay otros métodos que también podrían utilizarse (Barros et al., 2003; Cummings, 2009) y otro software. Sin embargo, cómo llevar a cabo estos análisis con otros paquetes como el SAS, se ha explicado en otros artículos (Deddens et al., 2008). Además, dado que R es software libre, cualquiera podría utilizar la sintaxis que se proporciona para calcular las asociaciones usando RPs en sus propios estudios.

Conclusión

En conclusión, aunque la regresión logística se utiliza mucho en estudios transversales para estimar las asociaciones entre variables, es posible y fácil utilizar otros modelos en el análisis de datos transversales con variables dependientes binarias, obteniendo RPs. Una de las ventajas importantes de estos métodos alternativos es que la RP, como medida de asociación, es más fácil de interpretar y de comunicar, sobre todo a las personas no epidemiólogas (Barros et al., 2003).

Financiación

Este trabajo fue financiado por la subvención Health Strategic Action [número de concesión PI13/00183] y parcialmente por la Red Española de Trastornos Adictivos [número de concesión RD12/0028/0018].

Agradecimientos

Este trabajo utiliza los datos de SHARE ola 4 comunicado el 1.1.1, del 28 de marzo de 2013. La recolección de los datos SHARE ha sido financiada principalmente por la Comisión Europea a través del 5º Programa Marco (proyecto QLK6-CT-2001-00360 en el programa temático de Calidad de Vida), a través del 6º Programa Marco (proyectos SHARE-I3, RII-CT-2006-062193, COMPARE, CIT5-CT-2005-028857 y SHARELIFE, CIT4-CT-2006-028812) y por el 7º Programa Marco (SHARE-PREP, no. 211909, SHARELEAP, no. 227822 and SHARE M4, no. 261982). Se agradece también la financiación adicional del US National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, R21 AG025169, Y1-AG-4553-01, IAG BSR06-11 y OGH4 04-064) y del Ministerio Alemán de Educación e Investigación, así como de diversas fuentes nacionales (véase <http://www.share-project.org> para una lista completa de las instituciones de financiación). Este artículo forma parte de la tesis doctoral de Marina Bosque-Prous, en la Universitat Pompeu Fabra.

Conflicto de intereses

Las personas autoras declaran que no hay conflicto de intereses.

Contribuciones de los autores

A. Espelt, M. Bosque-Prous y M. Marí-Dell'Olmo contribuyeron a la concepción y el diseño del estudio. M. Bosque-Prous contribuyó a la gestión y análisis de datos. A. Espelt, M. Marí-Dell'Olmo, M. Bosque-Prous y E. Penelo comentaron y contribuyeron a la interpretación de los resultados. A. Espelt escribió el primer borrador del documento, que fue revisado con contribuciones por todos los autores.

Referencias

- Barlés-Arizón, J.J., Escario, J.J., y Sánchez-Ventura, J. G. (2014) Predictors of driving under the influence of alcohol among Spanish adolescents. *Adicciones*, 26, 96-105.
- Barros, A. J. D., y Hirakata, V. N. (2003). Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. *BMC Medical Research Methodology*, 3, 21. doi:10.1186/1471-2288-3-21.
- Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmayer, J., Malter, F.,... Zuber, S. (2013). Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology*, 42, 992-1001. doi: 10.1093/ije/dyt088
- Bosque-Prous, M., Espelt, A., Guitart, A. M., Bartroli, M., Villalbí, J. R., y Brugal, M. T. (2014). Association between stricter alcohol advertising regulations and lower hazardous drinking across European countries. *Addiction*, 109, 1634-1643. doi:10.1111/add.12562.
- Brotman, D. J. (2006). Mediators of the association between mortality risk and socioeconomic status. *JAMA: The Journal of the American Medical Association*, 296, 763-764; author reply 764. doi: 10.1001/jama.296.7.763-b.
- Coutinho, L. M. S., Scazufca, M., y Menezes, P. R. (2008). Methods for estimating prevalence ratios in cross-sectional studies. *Revista de Saúde Pública*, 42, 992-998.
- Cummings, P. (2009). Methods for estimating adjusted risk ratios. *Stata Journal*, 9, 175-196.
- Deddens, J. A., y Petersen, M. R. (2008). Approaches for estimating prevalence ratios. *Occupational and Environmental Medicine*, 65, 501-506. doi: 10.1136/oem.2007.034777.
- Espelt, A., Borrell, C., Palència, L., Goday, A., Spadea, T., Gnani, R.,... Kunst, AE. (2013). Socioeconomic inequalities in the incidence and prevalence of type 2 diabetes mellitus in Europe. *Gaceta Sanitaria*, 27, 494-501. doi: 10.1016/j.gaceta.2013.03.002.
- Font-Ribera, L., Garcia-Contente, X., Pérez, A., Torres, R., Sala, N., Espelt, A., y Nebot, M. (2013). Driving under the influence of alcohol or drugs among adolescents: The role of urban and rural environments. *Accident Analysis & Prevention*, 60, 1-4. doi:10.1016/j.aap.2013.07.031.
- Gual, A., Segura, L., Contel, M., Heather, N., y Colom, J. (2002). Audit-3 and audit-4: effectiveness of two short forms of the alcohol use disorders identification test. *Alcohol Alcohol*, 37, 591-596. doi: 10.1093/alcalc/37.6.591.
- Jamieson, L. M., Gunthorpe, W., Cairney, S. J., Sayers, S. M., Roberts-Thomson, K. F., y Slade, G. D. (2010). Substance use and periodontal disease among Australian Aboriginal young adults. *Addiction*, 105, 719-726. doi: 10.1111/j.1360-0443.2009.02851.x.
- McNutt, L. A., Wu, C., Xue, X., y Hafner, J. P. (2003). Estimating the Relative Risk in Cohort Studies and Clinical Trials of Common Outcomes. *American Journal of Epidemiology*, 157, 940-943. doi: 10.1093/aje/kwg074
- Meneses-Gaya, C., Zuardi, A. W., Loureiro, S. R., Hallak, J. E. C., Trzesniak, C., de Azevedo Marques, J. M.,... Crippa, JA. (2010). Is the full version of the AUDIT really necessary? Study of the validity and internal construct of its abbreviated versions. *Alcoholism, Clinical and Experimental Research*, 34, 1417-1424. doi: 10.1111/j.1530-0277.2010.01225.x.
- Palencia, L., Espelt, A., Rodriguez-Sanz, M., Puigpinos, R., Pons-Vigues, M., Pasarin, M. I.,... Borrell, C. (2010). Socio-economic inequalities in breast and cervical cancer screening practices in Europe: influence of the type of screening program. *International Journal of Epidemiology*, 39, 757-765. doi: 10.1093/ije/dyq003.
- Sarasa-Renedo, A., Espelt, A., Folch, C., Vecino, C., Majó, X., Castellano, Y.,... Brugal, MT. (2014). Overdose prevention in injecting opioid users: the role of substance

- abuse treatment and training programs. *Gaceta Sanitaria*, 28, 146–54. doi:10.1016/j.gaceta.2013.10.012.
- Schiaffino, A., Rodriguez, M., Pasarin, M. I., Regidor, E., Borrell, C. y Fernandez, E. (2003). [Odds ratio or prevalence ratio? Their use in cross-sectional studies]. *Gaceta Sanitaria*, 17, 70–74.
- Shishehbor, M. H., Litaker, D., and Lauer, M. S. (2006). Mediators of the association between mortality risk and socioeconomic status - Reply. *JAMA: The Journal of the American Medical Association*, 296, 764–764. doi: 10.1001/jama.296.7.764a
- Szklo, M., & Nieto, J. (2012). *Epidemiology: Beyond the Basics* (3 edition.). Burlington, Mass: Jones & Bartlett Learning.
- Thompson, M. L., Myers, J. E. y Kriebel, D. (1998). Prevalence odds ratio or prevalence ratio in the analysis of cross sectional data: what is to be done? *Occupational and Environmental Medicine*, 55, 272–277.