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

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

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.

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.

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.

Received: April 2015; Accepted: July 2015

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

Cross-sectional designs are used extensively in substance use research. Substance use researchers usually use this type of design to estimate the association between a dichotomous dependent variable and one or more independent variables. Although the Odds Ratio (OR) or the Prevalence Ratio (PR) could be good estimators of this association, traditionally most studies have used OR, calculated with logistic regression, to estimate the association (Barros & Hirakata, 2003). The PR is defined as the prevalence in exposed population divided by the prevalence in non-exposed, while OR is the odds of disease or condition among exposed individuals divided by the odds of disease or condition among unexposed. In this sense, in cross-sectional designs, when the dependent variable is dichotomous, we usually obtain the prevalence in the descriptive analysis and therefore, PR is more intuitive and easy to understand than OR. Although OR is a good estimator of PR when the prevalence is low, it is known that OR overestimates PR when the prevalence is moderate or high (e.g. prevalence rates above 10%) (Szklo & Nieto, 2012). This could be a problem because OR has usually been treated and interpreted as a PR, independently of the prevalence of the illness (e.g in a paper about predictors of driving under the influence of alcohol among Spanish adolescents, the authors treated the OR as probabilities although the prevalence rate was above 10% in some categories) (Barlés-Arizón, Escario & Sánchez-Ventura, 2014). For those reasons, several studies have come up with alternative methods to estimate associations between a dichotomous dependent variable and several independent variables in cross-sectional designs, which yielded PR (Barros et al., 2003; Coutinho, Scazufca & Menezes, 2008; Deddens & Petersen, 2008; Schiaffino et al., 2003; Thompson, Myers & Kriebel, 1998). One of the simplest methods consists in using the following formula to calculate PR from a given OR (Schiaffino et al., 2003):

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

where p_1 is the prevalence of the illness or condition in the reference group (non-exposed).

In this case, although point-estimates are correct, there could be a problem when estimating confidence intervals, especially if the models have been adjusted for many variables. The intuitive method to calculate PR would be to use log-binomial regression. However, log-binomial regression often has convergence problems when any of the independent variables is continuous (Cummings, 2009; Deddens et al., 2008). As a result, alternative methods of modelling have been studied (e.g. cox regression models or Poisson regression models) (Barros et al., 2003; Deddens et al., 2008). In addition, although there is evidence supporting the use of Poisson regression models with robust variance to estimate PR in cross-sectional studies (Barros et al., 2003; Coutinho et al., 2008; Deddens et al., 2008), the optimal solution would be to use a log-binomial regression model (Deddens et al., 2008), if it converged. But, if we only take into account one decimal, the results using different regression models do not vary regardless of prevalence of the illness or condition (Cummings, 2009). In this sense, the estimation of PR using Poisson regression models with robust variance, based on the Huber sandwich estimate, has proved to be correct and robust in different experimental situations, such as using different prevalence values (low, moderate or high prevalence) or fitting several models (crude and adjusted) (Barros et al., 2003; Coutinho et al., 2008; Deddens et al., 2008).

Although there appears to be a tendency in recent years to use PR instead of OR in cross-sectional studies (Bosque-Prous et al., 2014; Espelt et al., 2013; Palencia et al., 2010), knowledge among substance use researchers about how to perform these analyses tends to be scarce. For this reason, the objectives of this brief report are to examine the differences between PR and OR in a cross-national study and to provide the tools to calculate PR using log-binomial and Poisson regression models with robust variance with two statistical packages commonly used in substance use research [STATA and R (free software)].

Methods

Design and participants

We used the database of the Survey of Health, Ageing and Retirement in Europe project (SHARE) (Börsch-Supan et al., 2013). The study population consisted of people over 50 years from 16 European countries who participated in wave 4 (2010-2012) of SHARE (n=41,263). Although the database contained sampling weights, they were not used in this study as it was not intended to do a population study. Moreover, participants with missing values in any of the variables were excluded.

Variables

The dependent variable was the prevalence of hazardous drinking, which was constructed using an adaptation of the Alcohol Use Disorders Identification Test Consumption (AUDIT-C test) (Meneses-Gaya et al., 2010). It was constructed based on three questions: two assessing regular drinking in terms of frequency and quantity and one assessing binge drinking (six or more alcoholic drinks in a single occasion, at least once a month in the preceding 3 months). Each answer was ranked from 0 to 4 points, and a final score was calculated as the sum of scores from each question. Hazardous drinking was built as a dichotomous variable (hazardous/non-hazardous drinking), considering drinking to be hazardous when the score was 5 or more among men, and 4 or more among women (Gual, Segura, Contel, Heather & Colom, 2002) [variable name: *auditc*]. The independent

Table 1. *Explanation of the steps to estimate PR using log-binomial regression models or Poisson regression models with robust variance in two statistical packages (STATA and R), using the tool bar or the specific syntax*

Package	Tool Bar (step by step)	Syntax unadjusted	Syntax adjusted
Log-Binomial regression Model			
STATA	Menu tools→Statistics→Generalized Linear Models→Generalized Lineal Models (GLM)→Model [(dependent variable: auditc; independent variable: sex)/ (family: binomial; link choices: log)] →Reporting [report exponentiated coefficients.	glm auditc sex, family(binomial 1) link(log) eform	glm auditc sex educ age, family(binomial 1) link(log) eform
R		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)	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)
Poisson regression model with robust variance			
STATA	Menu tools→Statistics→Generalized Linear Models→Generalized Lineal Models (GLM)→Model [(dependent variable: auditc; independent variable: sex)/ (family: poisson; link choices: log)] →SE/Robust [standard error type: Robust]→Reporting [report exponentiated coefficients.	glm auditc sex, family(poisson) link(log) robust eform	glm auditc sex educ age, family(poisson) link(log) robust eform

R		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←coeftest(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)exp(B + qnorm(0.05 / 2) * SE)
		# lower 95% CI
		exp(B + qnorm(1 - 0.05 / 2) * SE)
			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←coeftest(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)

variable used was gender [variable name: *sex*] and two different covariables were used to adjust: age, as a continuous variable [variable name: *age*], and educational level (less than secondary studies or secondary or tertiary studies), as a categorical variable [variable name: *educ*]. Finally, we took into account the country of residence, as a stratification variable.

Analysis

We calculated the prevalence of hazardous drinking by gender in each country, using STATA. PR of being a hazardous drinker in men with respect to women was estimated with Mantel-Haenszel method in STATA [syntax: cs *auditc sex*], and with log-binomial regression models and Poisson regression models with robust variance, stratified by country, in STATA and R (table 1). To estimate Poisson regression models, it is necessary to have individual data and to satisfy the following two conditions in order to obtain a realistic point-estimate and confidence intervals of reasonable width (Barros et al., 2003). First, the dependent variable has to be dichotomous with values 0 and 1 (other values cannot be used) when estimating the Poisson models. Value 1 is assigned to the individuals with the disease or condition (hazardous drinkers in our example) and 0 to the remaining participants. And second, the variance of the estimations has to be robust. All the models were performed using Generalized Linear Models with Poisson or binomial families with log link function.

Finally, we also calculated the association between gender and hazardous drinking for each country using logistic regression models in STATA [logit *auditc sex*, or], which yielded OR. Overestimations of OR with respect to PR for each country were calculated, using the following formula: [Overestimation=(OR-PR)/(OR-1)] (Brotman, 2006; Espelt et al., 2013; Shishehbor, Litaker, Lauer, 2006). To perform all the analyses, we used STATA13.0 and R 3.0.2.

Results

Table 1 shows the steps to calculate PR by fitting log-binomial regression models and Poisson regression models with robust variance through the toolbar and the specific syntax, using STATA and R. Data to perform all the analyses are available in STATA format (supplementary data). To get these data and to execute all the analyses properly with R statistical package, the user needs to have previously installed “foreign”, “Epi” and “sandwich” libraries (table 1). To read STATA data in R the instruction is data<- read.dta(“C:/Users_directori/bbdd.dta”, convert.factors=F).

Table 2 shows hazardous drinking prevalence in men and women for each country and the associations between variables calculated using STATA. Hazardous drinking prevalence varied from one country to another. For example, hazardous drinking prevalence in Slovenia was low in both men and women (14% and 11%, respectively), while it was high in both genders in Denmark (39% in men and 35% in women) and in Estonia it was moderate in men (17%) but low in women (4%). PR estimates and their 95% confidence intervals (95%CI) calculated using STATA and R were the same as those calculated using the Mantel-Haenszel method. However, OR overestimated PR in almost all analyses. For example, PR of being a hazardous drinker in men with respect to women in Austria was 1.49 (95%CI: 1.34-1.66) while the corresponding OR was 1.66 (95%CI: 1.45-1.90). Moreover, PR was 1.33 (95%CI: 1.22-1.46) in France, while OR was 1.47 (95%CI: 1.30-1.66). If OR was interpreted as a PR, the overestimation of OR was high in some countries (e.g. 40% in Denmark or 33% in Belgium). The degree of this overestimation depended on the prevalence of hazardous drinking among men and women in each country. However, when the prevalence was similar for men and women, no differences between PR and OR were observed in Netherlands and Switzerland but 33% of overestimation was found in Italy (table 3).

In general, PR calculated using log-binomial or Poisson regression models with robust variance do not vary among them in the unadjusted analysis. However, in the adjusted analysis controlling for educational level and age, some differences in the second decimal were found. PR obtained using different packages were not statistically different.

Discussion

The results show that there is no reason to systematically use OR instead of PR in cross-sectional studies, especially if the prevalence of the disease or condition is moderate or high, since PR are calculated easily and there are methods to obtain robust estimations of PR and their 95%CI. Moreover, our findings are in line with other published articles (Barros et al., 2003; Coutinho et al., 2008; Deddens et al., 2008; Schiaffino et al., 2003; Thompson et al., 1998). As stated in this methodological study, statistical packages used in most epidemiological studies allow researchers to calculate PR easily. However, if we use Poisson regression models we have to be sure that we have used robust methods to estimate their variance, otherwise the Poisson regression would produce wider confidence intervals compared to a log-binomial regression model (McNutt, Wu, Xue, & Hafner, 2003).

One advantage of using PR is that the results are much more intuitive. For example, prevalence of hazardous drinkers in men and women in Austria is 25.6% and 17.2%, respectively. When dividing the prevalence in men by prevalence in women we obtain a PR of 1.49, which is the same PR that was estimated using the various statistical packages. Moreover, we found that the degree of overestimation of PR (using OR) varied among countries and depended on the prevalence of the disease or condition (i.e. hazardous drinking) in exposed and non-exposed participants (in

Table 2. *Unadjusted Prevalence, prevalence ratio, odds ratio and overestimation of OR with respect to PR estimates of being hazardous drinker between men and women in several European countries.*

	Men	Women	Hazardous drinking prevalence		$PR_{Men/Women}$ Mantel-Haenszel		$PR_{Men/Women}$ log-binomial		$PR_{Men/Women}$ robust Poisson		$OR_{Men/Women}$ logistic regression		Over-estimation*
	N	N	Men	Women	PR	95%CI	PR	95%CI	PR	95%CI	OR	95%CI	%
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%
Belgium	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%
Czech Republic	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%
Denmark	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%
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%
France	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%
Germany	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%
Hungary	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%
Italy	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%
Netherlands	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%
Poland	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%
Slovenia	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%
Spain	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%
Sweden	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%
Switzerland	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%

Note. *Overestimation of OR with respect to PR was calculated using the formula: [Overestimation = (OR-PR)/(OR-1)] (Brotman, 2006; Espelt et al., 2013; Shishehbor, Litaker, Lauer, 2006)

Table 3. *Comparison of adjusted prevalence ratio, adjusted odds ratio and overestimation of adjusted OR with respect to adjusted PR estimates of being hazardous drinker between men and women in several European countries.*

	$PR_{Men/Women}$ log-binomial[1]		$PR_{Men/Women}$ robust Poisson[1]		$OR_{Men/Women}$ logistic regression[1]		Overestimation*
	PR	95%CI	PR	95%CI	OR	95%CI	%
Austria	1.48	(1.33-1.65)	1.48	(1.32-1.64)	1.64	(1.43-1.88)	25.0%
Belgium	1.11	(1.02-1.20)	1.10	(1.01-1.19)	1.15	(1.02-1.30)	33.3%
Czech Republic	2.08	(1.89-2.29)	2.08	(1.89-2.30)	2.65	(2.33-3.01)	34.5%
Denmark	1.11	(0.99-1.24)	1.10	(0.99-1.23)	1.17	(0.98-1.39)	41.2%
Estonia	3.87	(3.27-4.57)	3.87	(3.28-4.58)	4.67	(3.87-5.63)	21.8%
France	1.30	(1.19-1.43)	1.31	(1.19-1.44)	1.44	(1.27-1.62)	29.5%
Germany	1.39	(1.13-1.71)	1.38	(1.12-1.70)	1.49	(1.15-1.94)	22.4%
Hungary	3.07	(2.56-3.69)	3.07	(2.55-3.68)	3.78	(3.05-4.68)	25.5%
Italy	1.05	(0.93-1.18)	1.05	(0.93-1.17)	1.06	(0.91-1.24)	16.7%
Netherlands	1.01	(0.91-1.13)	1.01	(0.90-1.12)	1.01	(0.85-1.19)	0.0%
Poland	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%
Slovenia	1.28	(1.05-1.57)	1.28	(1.05-1.56)	1.33	(1.06-1.67)	15.2%
Spain	1.52	(1.30-1.78)	1.51	(1.28-1.77)	1.63	(1.35-1.98)	19.0%
Sweden	0.92	(0.73-1.15)	0.91	(0.72-1.15)	0.90	(0.69-1.18)	10.0%
Switzerland	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%

Note. *Overestimation of OR with respect to PR was calculated using the formula: [Overestimation = (OR-PR)/(OR-1)](Espelt et al., 2013). [1] Adjusted by age and educational level.

this study, men were considered as exposed and women as non-exposed). For that reason, if we interpret OR as an estimation of PR, we could be misinterpreting the results, as we have seen in the results section. The fact that OR could overestimate PR depending on the prevalence of the condition or disease analysed in each country leads OR to be similar to PR in some countries, while in others the estimations of OR and PR are quite different. As a result, when OR are used to make comparisons among countries, the interpretation of the results could be a problem for researchers that intuitively interpret OR as PR. For those reasons, in cross-national studies, where comparisons between countries with large differences in the prevalence of the disease or condition are made, it is advisable to use PR instead of OR. It is especially relevant because, as we said, people usually read the OR estimate as a PR. The overestimation may inappropriately affect clinical decisions-making or policy development and therefore may lead to unintentional errors in the economic analysis of potential intervention programs or treatments (McNutt et al., 2003).

Nowadays, some substance use studies are starting to use regression models to obtain PR as estimators of the association between a dichotomous dependent variable and several independent variables. In this sense, in substance abuse research some studies have calculated PR to estimate which factors could be associated to illicit drug consumption (Jamieson et al., 2010; Sarasa-Renedo et al., 2014) or to licit drug use (Bosque-Prous et al., 2014; Font-Ribera et al., 2013; Jamieson et al., 2010). However, the use of regression methods to estimate PR is still scarce. For example, if we compare the studies published in Pubmed in 2013 that have used PR or OR using the following strategies: PR = (["cross-sectional"] and ["prevalence ratio" or "log-binomial" or "poisson regression model with robust variance"]); OR = (["cross-sectional"] and ["odds ratio" or "logistic regression"]), we found 132 papers that used PR and 4886 that used OR.

One of the main strengths of our study is that we explain how to calculate PR using different regression models and also two different statistical packages (one of which is free software available to all researchers). However, this study could suffer some limitations. Its main limitation is that it was not designed as a simulation study, using different conditions to analyse the changes in PR with respect to OR. However, this was not the aim of this article. Nevertheless, relying on a cross-national study with substance use real data will be easier to understand. In fact, almost all scenarios are found in the different countries participating in the study (i.e. high prevalence in both sexes, low prevalence in both sexes, combinations of high and low prevalence), strengthening our results. Another limitation is that we only show the models most frequently used to calculate PR, using two different packages, but there are other methods that could also be used (Barros et al., 2003; Cummings, 2009) and other software. However, how to perform these analyses with other packages, as SAS, have been explained elsewhere (Deddens et al., 2008). In addition, given that R is free software, anyone could use the syntax that is provided to estimate associations using PR in their own studies.

Conclusion

In conclusion, although logistic regression is highly used in cross-sectional studies to estimate associations between variables, it is possible and easy to use other models in the analysis of cross-sectional data with binary dependent variables, which yield PR. One of the important advantages of these alternative methods is that PR, as a measure of association, is easier to interpret and communicate, especially to non-epidemiologists (Barros et al., 2003).

Funding

This work was supported by Health Strategic Action grant [grant number PI13/00183] and partially supported by the Spanish Network on Addictive Disorders [grant number RD12/0028/0018].

Acknowledgements

This paper uses data from SHARE wave 4 release 1.1.1, as of 28 March 2013. The SHARE data collection has been funded primarily by the European Commission through the 5th Framework Program (project QLK6-CT-2001-00360 in the thematic program Quality of Life), through the 6th Framework Program (projects SHARE-I3, RII-CT-2006-062193, COMPARE, CIT5-CT-2005-028857 and SHARELIFE, CIT4-CT-2006-028812) and through the 7th Framework Program (SHARE-PREP, no. 211909, SHARELEAP, no. 227822 and SHARE M4, no. 261982). Additional funding from the US National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, R21 AG025169, Y1-AG-4553-01, IAG BSR06-11 and OGHA 04-064) and the German Ministry of Education and Research as well as from various national sources is gratefully acknowledged (see http://www.share-project.org for a full list of funding institutions). This article forms part of the doctoral dissertation of Marina Bosque-Prous, at the Universitat Pompeu Fabra.

Conflict of interest

The authors declare no conflict of interest.

Authors' contributions

A. Espelt, M. Bosque-Prous, M. Marí-Dell'Olmo contributed to the conception and design of the study. M. Bosque-Prous contributed to data management. A. Espelt, M. Marí-Dell'Olmo, M. Bosque-Prous and E. Penelo discussed and contributed to interpretation of the results. A. Espelt wrote the first draft of the paper, which was revised with contributions by all authors.

References

Barlés-Arizón, J.J., Escario, J.J., & Sánchez-Ventura, J. G. (2014) Predictors of driving under the influence of alcohol among Spanish adolescents. *Adicciones, 26*, 96-105.

Barros, A. J. D., & 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., Korbmacher, 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., & 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: Th Jjournal 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., & 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., & 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., Gnavi, 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-Continente, X., Pérez, A., Torres, R., Sala, N., Espelt, A., & 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., & 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., & 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., & 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. *Internal 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., & 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.296.7.764-a

Szklo, M., & Nieto, J. (2012). *Epidemiology: Beyond the Basics* (3 edition.). Burlington, Mass: Jones & Bartlett Learning.

Thompson, M. L., Myers, J. E., & 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.