Considerations on the use of Odds Ratio versus Prevalence or Proportion Ratio

Reflexiones sobre el uso de la Odds Ratio o la Razón de Prevalencias o Proporciones

Albert Espelt*,**,***, Marina Bosque-Prous*****, Marc Marí-Dell'Olmo**,****

* Facultat de Ciències de la Salut de Manresa. Universitat de Vic Universitat Central de Catalunya (UVicUCC). Av. Universitària 46, 08242, Manresa, Spain
** Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP). C/ Monforte de Lemos 3 Pabellón 11, 28029, Madrid, Spain
***Departament de Psicobiologia i Metodologia en Ciències de la Salut. Universitat Autònoma de Barcelona (UAB). C/ de Ca n'Altayó s/n. 08193 Bellaterra. Spain
****Agència de Salut Pública de Catalunya, Barcelona, España; Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, España.
***** Faculty of Health Sciences, Universitat Oberta de Catalunya, Barcelona, Spain.

In 2017, we published an article explaining how to estimate prevalence ratios using different regression models (Espelt, Mari-Dell'Olmo, Penelo & Bosque-Prous, 2017). In health science research, we often work with cross-sectional or longitudinal studies. The variables of interest in these studies may be dichotomous variables (i.e. yes vs no) related to a disease or health condition, and we usually represent them as proportions, such as prevalence or cumulative incidence (e.g. percentage of at-risk drinkers or percentage of new cases of hazardous drinking in a given period, respectively) (Hernandez-Avila, Garrido-Latorre & Lopez-Moreno, 2000; Moreno-Altamirano, López-Moreno & Corcho-Berdugo, 2000). In this sense, when we work with dichotomous dependent variables and proportions, the first estimate we show is the prevalence or cumulative incidence of the disease or unhealthy behavior. Once the prevalence or cumulative incidence has been estimated, what will interest us is to ascertain whether this prevalence or cumulative incidence changes depending on the different independent or explanatory variables. For this reason, after a first table describing the sample, a second table is usually presented showing the proportions according to the different independent or explanatory variables (with their respective statistical tests). Up to this point, we would all agree. However, we are often interested in showing the measures of association between these dichotomous dependent variables and the independent or explanatory variables in both a crude and an adjusted way. It is at this point that some questions arise which we already tried to answer in the article published in 2017 with data from a European study on alcohol (Espelt et al., 2017). Which measure of association is better: the odds ratio (OR) or the prevalence or proportions ratio (PR)? Which should we select? If we simply want to see whether there is an association between variables, the answer would be that both measures perform equally well. However, if we wish to interpret the magnitude of this association, we must be aware that OR and PR are not interpreted in the same way. The problem is that the OR is difficult to understand and is often interpreted as if it were a PR. This misinterpretation is worse in the cases when the prevalence, proportion or cumulative incidence of disease or health behavior is high (Szklo & Nieto, 2012) because in these cases, the OR is not similar to the PR, and by using an OR instead of a PR we would be overestimating the association. For example, an OR of 2.65 can perfectly well be a PR of 2.08 (Espelt et al., 2017), and here the problems begin. It is not the same to say that there is 265% more disease in one category rather

Received: August 2019; Accepted: September 2019.

Send correspondence to:
Marina Bosque-Prous. Facultat de Ciències de la Salut. Universitat Oberta de Catalunya. Rambla del Poblenou, 156, 08018 Barcelona
E-mail: mbosquep@uoc.edu

than 208%. If we review the scientific literature, we will find that most cross-sectional or longitudinal studies published with dichotomous dependent variables show associations with OR while few present associations with PR (Espelt et al., 2017). From our point of view, in order to give coherence to the design so that the descriptive tables of proportions are totally comparable with the crude data of associations, thus ensuring that there are no possible errors in the interpretation of the data of these associations, the most reasonable solution would be to use PR as a measure of association whenever possible for designs with proportions (cross-sectional or longitudinal). This is especially important when ORs are shown in tables because non-specialists are highly likely to misinterpret OR as PR and understand OR as higher probability, higher risk or higher prevalence when they should really think of it as a comparison of probabilities: the odds of an outcome happening compared to the odds of the non-occurrence of that outcome, i.e. the *relative* odds. Nevertheless, epidemiologists or researchers also fall into these errors sometimes, and we find articles with inadequate vocabulary to explain associations (Barlés Arizón, Escario & Galbe Sánchez-Ventura, 2014; Díaz Geada, Busto Miramontes & Caamaño Isorna, 2018; Mori-Gamarra et al., 2018).

To this end, our previous article provided the steps to calculate prevalence ratios with two statistical packages, STATA and R. On this occasion, we would also like to facilitate the procedure with the statistical package SPSS, which is widely employed in our environment, using the menus. Table 1 shows how to obtain the PRs from the binomial regression model using SPSS. It is very important that the dependent variable is coded as 0 and 1, with 1 being the health outcome or disease of interest.

Table 1. *Explanation of the steps to estimate PR using log-binomial regression models with SPSS*

Tool Bar (step by step)
Menu tools → Analyze → Generalized Linear Models → Generalized Linear Models → Type of Model [Custom; Distribution: Binomial; Link function: Log] → Response [Dependent variable: auditc; Type of Dependent Variable: Binary; Reference Category: First (lower value)] → Predictors [Factors: educ, sex; Covariates: age] → Model [Model: educ, sex, age] → Estimation [maximum iterations: 100000] → Statistics [report exponentiated coefficients]

In many articles, when we analyze the association between a dependent variable and one or more independent or explanatory variables, we try to control for possible variables that could be confounding factors (Babyak, 2009). One of the statistical techniques used to control for other variables is to fit regression models. In the different studies using regression models, we found crude associations as well as associations adjusted for some of these confounding variables. For example, living in an urban environment is protective against alcohol consumption compared to living in a rural environment [PR=0.86 (95%CI = 0.78-0.95)]. However, this protective effect disappears [PR=0.91 (95%CI = 0.80-1.05)] when different individual variables and factors such as sports centers, unemployment rate, number of pubs, and other accessibility variables are accounted for (Obradors-Rial, Ariza, Continente & Muntaner, 2019).

In the cited article, it is easy to note from the prevalence table that the prevalence of high-risk drinking in urban areas (51.1%) is lower than in rural areas (59.3%) (Obradors-Rial et al., 2019). A simple division of the prevalence of high-risk drinking in the urban environment by the prevalence of high-risk drinking in the rural environment reveals the association given by the regression model (RP = 0.86). However, we stated that the adjusted association between high-risk drinking and urban/rural environment is 0.91. Can we determine the adjusted prevalence of alcohol consumption in the urban and rural environment which gives us this association? With the statistical program STATA, we can estimate the adjusted prevalence for each of the categories of the independent variable of interest (Muller & MacLehose, 2014). If we return to the article published in 2017 (Espelt et al., 2017), we can observe in Table 2 that the prevalence of risk drinkers in Estonian men was 16.75% while in women it was 4.24%, which implies a PR of 3.95. Estonia's age- and education level adjusted prevalence ratio was 3.87. However, by applying the regression model and estimating its marginals (Table 2), we find that the prevalence of high-risk drinkers in men adjusted for age and level of education was 16.48% and that of women was 4.26%. If we divide the prevalence of high-risk drinking in men versus women, we obtain the same adjusted PR as in the model.

Table 2. *Syntax to obtain crude and adjusted PR with STATA and R*

Adjusted Model Syntax

STATA

```
glm auditc i.sex i.educ age, family(binomial 1) link(log) eform
margins i.sex i.educ
```

R

```
install.packages(pkgs = c("Epi", "foreign"))
library(Epi)
library(foreign)
model<-glm(auditc ~ sex + educ + age, data=data,
family=binomial(link=log))
summary(model)
round(ci.lin(model, Exp=T),2)


library(prediction)
pred <- prediction(model, at = list(sex=c("Women","Men")), type =
"response")
```

In conclusion, from our point of view, in studies in which the independent variable or variable of interest is a proportion, it is more useful to use the prevalence or proportion ratios as a measure of association since this implies greater coherence and avoids errors of interpretation.

Conflict of interests

The authors of this article declare no conflict of interest. Albert Espelt is associate editor of the journal Adicciones, but this did not play any role in the process of publication.

References

Babyak, M. A. (2009). Understanding confounding and mediation. *Evidence-based Mental Health*, *12*, 68-71. doi:10.1136/ebmh.12.3.68.

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

Díaz Geada, A., Busto Miramontes, A. & Caamaño Isorna, F. (2018). Alcohol, tobacco and cannabis consumption in adolescents from a multicultural population (Burela, Lugo). *Adicciones*, *30*, 264-270. doi:10.20882/adicciones.915.

Espelt, A., Mari-Dell'Olmo, M., Penelo, E. & Bosque-Prous, M. (2017). Applied Prevalence Ratio estimation with different Regression models: An example from a cross-national study on substance use research. *Adicciones*, *29*, 105-112. doi:10.20882/adicciones.823.

Hernandez-Avila, M., Garrido-Latorre, F. & Lopez-Moreno, S. (2000). Epidemiologic study design. *Salud Pública de México*, *42*, 144-154.

Moreno-Altamirano, A., López-Moreno, S. & Corcho-Berdugo, A. (2000). Principales medidas en epidemiología. *Salud Pública de México*, *42*, 337–348.

Mori-Gamarra, F., Moure-Rodríguez, L., Sureda, X., Carbia, C., Royé, D., Montes-Martínez, A., … Caamaño-Isorna, F. (2018). Alcohol outlet density and alcohol consumption in Galician youth. *Gaceta Sanitaria,* pii: S0213-9111(18)30238-3. doi:10.1016/j.gaceta.2018.09.005.

Muller, C. J. & MacLehose, R. F. (2014). Estimating predicted probabilities from logistic regression: Different methods correspond to different target populations. *International Journal of Epidemiology, 43*, 962-970. doi:10.1093/ije/dyu029.

Obradors-Rial, N., Ariza, C., Continente, X. & Muntaner, C. (2019). School and town factors associated with risky alcohol consumption among Catalan adolescents. *Alcohol*, pii: S0741-8329(17)30987-4. doi:10.1016/j.alcohol.2019.04.005.

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