

Obtaining adjusted prevalence ratios from logistic regression models in cross-sectional studies

Obtendo razões de chance prevalentes de modelos de regressão logística em estudos transversais

La obtención de las prevalencias ajustadas a partir de los modelos de regresión logística en los estudios transversales

Leonardo Soares Bastos ¹
Raquel de Vasconcellos Carvalho de Oliveira²
Luciane de Souza Velasque ³

Abstract

In the last decades, the use of the epidemiological prevalence ratio (PR) instead of the odds ratio has been debated as a measure of association in cross-sectional studies. This article addresses the main difficulties in the use of statistical models for the calculation of PR: convergence problems, availability of tools and inappropriate assumptions. We implement the direct approach to estimate the PR from binary regression models based on two methods proposed by Wilcosky & Chambless and compare with different methods. We used three examples and compared the crude and adjusted estimate of PR, with the estimates obtained by use of log-binomial, Poisson regression and the prevalence odds ratio (POR). PRs obtained from the direct approach resulted in values close enough to those obtained by log-binomial and Poisson, while the POR overestimated the PR. The model implemented here showed the following advantages: no numerical instability; assumes adequate probability distribution and, is available through the R statistical package.

*Prevalence Ratio; Logistic Models;
Cross-Sectional Studies*

Resumo

Nas últimas décadas, tem sido discutido o uso da razão de prevalência (RP) ao invés da razão de chance como a medida de associação a ser estimada em estudos transversais. Discute-se as principais dificuldades no uso de modelos estatísticos para o cálculo da RP: problemas de convergência, disponibilidade de ferramentas e pressupostos não apropriados. O objetivo deste estudo é implementar uma abordagem direta para estimar a RP com base em modelos logísticos binários baseados em dois métodos propostos por Wilcosky & Chambless, e comparar com outros métodos. Utilizou-se três exemplos e comparou-se as estimativas bruta e ajustada da RP obtidas pela função com as estimativas obtidas pelos modelos log-binomial, Poisson e razão de chance prevalente (RCP). As RP da abordagem proposta resultaram em valores próximos aos obtidos pelos modelos log-binomial e Poisson, e a RCP superestimou a RP. O modelo aqui implementado apresentou as seguintes vantagens: não apresenta instabilidade numérica; assume a distribuição de probabilidades adequada; e está disponível no programa estatístico R.

*Razão de Prevalências; Modelos Logísticos;
Estudos Transversais*

¹ Programa de Computação Científica, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil.

² Instituto de Pesquisa Clínica Evandro Chagas, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil.

³ Departamento de Matemática e Estatística, Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, Brasil.

Correspondência

L. S. Bastos
Programa de Computação Científica, Fundação Oswaldo Cruz.
Av. Brasil 4365, Rio de Janeiro, RJ
21040-360, Brasil.
lsbastos@fiocruz.br

Introduction

Over the past few decades, several authors ^{1,2,3,4,5,6,7} have tried to determine the most appropriate association measure to be used in cross-sectional studies. The consensus is that the prevalence odds ratio (POR) is only a good approximation of the prevalence ratio (PR) when the event of interest is rare ⁸. Logistic regression is the most popular statistical model used in estimating POR due to ease of interpretation and computational implementation. However, when the choice of association measure is the PR, and the event of interest is not rare, this model produces poor estimates. In such cases, several authors have proposed alternatives to logistic regression models to estimate the true PR.

Lee & Chia ⁹ were the first authors to use Cox models with Breslow's modification (Breslow-Cox model) to estimate prevalence ratios, but in the study, standard errors and, consequently, confidence intervals, were not calculated correctly. The correction for standard errors obtained by Cox models had already been proposed ¹⁰, but were not considered. Barros & Hirakata ⁵ used the fact that Breslow-Cox and Poisson models estimate the same effects ¹¹ and used Poisson regression models with robust variance to estimate the PR. Zou ¹² published a simulation study demonstrating the reliability of the Poisson model with robust variance to estimate PR in 2 by 2 tables. The main issue with using a Poisson model to estimate PR is the misuse of a specific counting probability distribution to describe a response variable that is dichotomous (presence or absence of an outcome).

Skov et al. ⁴ used a generalized linear model with a binomial distribution and a log link (log-binomial model) to directly estimate PR ¹³. Although this model allows for directly estimating the PR and assumes a probability distribution that agrees with the type of the response variable, the lack of convergence in the presence of continuous variables remains a problem. To solve this issue, Deddens et al. ⁶ introduced the COPY method for finding an approximation to the MLE when the log-binomial model fails to converge. Due to the convergence problem of the log-binomial model, Schouten et al. ¹⁴ proposed a simple data manipulation that allows for the use of logistic regression to obtain the PR. It consists in modifying the data set by duplicating the lines where the event occurs and replacing the outcome from event to non-event ^{14,15,16}.

Another approach – proposed by Wilcosky & Chambless ¹⁷, using the conditional and marginal methods ¹⁸ – involves a direct adjustment of epidemiological measures through binary re-

gression. An advantage of these methods is that they assume a probability distribution for a variable with a binomial response, matching the nature of the observed response variable in cross-sectional studies. We find one article ¹⁹ that uses the Wilcosky & Chambless ¹⁷ method to estimate PR, but it did not mention the software implementation. Recently, R Core Team developed a software package in R (R package version 1.2; The R Foundation for Statistical Computing, Viena, Austria; <http://www.r-project.org>) for estimating marginal and conditional PRs and confidence intervals via bootstrap and delta methods, but they have yet to publish a scientific article explaining the details of this package and the differences between the methods it utilizes to estimate PR.

In this article, we use a direct approach to estimate the prevalence ratio from binary regression models based on methods proposed by Wilcosky & Chambless ¹⁷, and we compare the results to those obtained through different methods presented in the literature. Three different data sets are used to illustrate our study.

Methods

Based on the approach proposed by Wilcosky & Chambless ¹⁷, we use real and simulated data to compare PR estimates obtained by the marginal and conditional methods. Those estimates are also compared with the estimates obtained by the binomial, log-binomial and robust Poisson/Cox models.

Using a logistic model, it is straightforward to estimate the probability of occurrence of a disease (denominated prevalence in transversal studies) adjusted for two or more variables. Suppose, for example, that one has information about the diabetes status (1: yes/0: no), age (continuous) and obesity status (1: yes/0: no) of a defined population. With this information, one can obtain the adjusted probability of diabetes through the following equation:

$$P = \frac{1}{1 + \exp[-(\beta_0 + \beta_1 AGE + \beta_2 OBESITY)]} \quad (\text{Equation 1}),$$

where P is the probability that DIABETES = 1, β_0 , β_1 , β_2 are regression coefficients estimated from the data. Note that $\exp(\beta_2)$ estimates the odds ratio for diabetes in obese individuals compared to non-obese individuals, adjusted by age. However, if we are interested in obtaining the estimated PR for diabetes, adjusted by age, in obese and non-obese individuals, we can proceed in two ways, as described below:

1) Marginal method: in each stratum of the variable OBESITY (yes or no), the diabetes preva-

lence is calculated for each age value included in the dataset using Equation 1. The PR is the ratio between the average of the prevalences in each stratum. Wilcosky & Chambless¹⁷ refer to this estimate as the marginal prevalence ratio (MPR); 2) Conditional method: in each stratum of the variable OBESITY, the diabetes prevalence is calculated using Equation 1, setting age as an average value obtained from the dataset. Thus, the ratio of the two prevalences can be calculated. Wilcosky & Chambless¹⁷ refer to this estimate as the conditional prevalence ratio (CPR).

In the linear regression model, both methods estimate the same value. However, in the logistic regression model, we observed significant differences between the estimates of the two methods when p is close to zero or one.

According to Lee & Chia⁹, the marginal method provides an internally adjusted measure, making invalid any comparisons to external values of PR. With the conditional method, on the other hand, one can use default values as the average values of covariates, allowing for comparisons with other population studies that used the same default values. More details about the marginal and conditional methods can be found in Lee²⁰ and Wilcosky & Chambless¹⁷.

Asymptotic confidence intervals for the conditional and marginal prevalence ratios were proposed by Flanders & Rhodes²¹. The authors also presented an SAS (SAS Inc., Cary, USA) script for estimating and calculating the intervals of the conditional and marginal prevalence ratio. To the best of our knowledge, this was the only implementation of these measures to date.

The prevalence ratio estimation methods are illustrated in three different databases, all containing a binary outcome Y , a binary exposure X , and at least one control variable Z .

Application 1: the first database was a toy example with 1,000 simulated observations and one continuous control variable. In this example, we simulated 1,000 binary outcomes with a binary exposure, X , and a continuous confounding variable, Z . The exposure was sampled from a Bernoulli distribution with probability 0.5, the confounding variable was sampled from a Normal distribution with mean zero and unit variance. The outcome was sampled from a Bernoulli distribution with probabilities such that: the baseline prevalence was 20%; the conditional prevalence ratio for X at $Z = 0$ was equal to 2; and the conditional prevalence ratio for X at $Z = 1$ was equal to 1.919 (regression coefficient $\beta_2 = 0.20$). There were several possible values for the conditional prevalence ratio for X depending on Z .

Application 2: the second database referred to a cohort of 1,273 live births in 1993 in the city

of Pelotas, Rio Grande do Sul, Brazil, studied with the aim of linking sociodemographic factors and reproductive health, informed by the responsible female, to the nutritional condition of their children after 4-5 years⁵. The analysis considered underweight in 4-5 year old children (with a prevalence of 4.1%) as the outcome of interest, Y ; previous hospitalization as the exposure, X ; and birth weight (normal or low birth weight) as a control variable, Z . For this application, because all variables are binary, we were able to calculate the prevalence ratio applying the Mantel-Haenszel method, considered here to be the gold standard.

Application 3: the third database analyzed 703 sexually active, HIV-infected women, treated between 1996 and 2007 in Rio de Janeiro, Brazil, with no history of hysterectomy. The data was collected in order to assess factors associated with high-grade squamous intraepithelial lesions (HSIL), lesions that can develop into cancer of the cervix²². Five variables pertaining to the HIV-infected women were included in analysis: presence of HPV was the exposure variable, X ; presence of cervical cytological abnormalities was the outcome, Y ; and the control variables, Z , were age, number of pregnancies and the time since the last gynecological examination. Variables X and Y were binary variables and the others were continuous variables. The prevalence of the outcome was 4.1%.

Adjusted prevalence ratios and prevalence odds ratios were estimated by several different methods. Prevalence ratios were estimated by robust Poisson and log-binomial models, and by the conditional and marginal methods proposed by Wilcosky & Chambless¹⁷. POR were also calculated using the usual logistic regression model.

The different methods to obtain prevalence ratios were coded in R (The R Foundation for Statistical Computing, Vienna, Austria; <http://www.r-project.org>). An R function to estimate the conditional and marginal prevalence ratios, as proposed by Wilcosky & Chambless¹⁷, is available (Figure 1).

Results

Application 1: toy example

Table 1 presents estimates obtained through different methods for the prevalence ratio of the variable X . The true prevalence ratio depends on Z , which follows a standard normal distribution. Hence, the true conditional prevalence ratio varies from 1.71 to 2.25 for Z varying from

Figure 1

R code developed to estimate the conditional and the marginal prevalence ratio, as proposed by Wilcosky & Chambless¹⁷.

```
#####
### Prevalence ratio from logistic regression models
#####

PR = function( model, type="marginal", level=0.95 ){
  # model: output of a glm funtion family=binomial
  # type: "marginal", "conditional", "both"

  p = model$rank-1
  n = length(model$y)

  xx=rep(0,model$rank-1)

  modeloAux = update(model, x=T)
  modeloAux2 = glm(modeloAux$y ~ modeloAux$x[, -1],
    family=binomial(), weights=modeloAux$prior.weights)

  labelCI = paste( (1 + c(-level,level))/2 * 100, "%")
  auxMatrix = data.frame(PR = model$coefficients[-1], sd = NA)
  auxMatrix[, labelCI[1]] = NA
  auxMatrix[, labelCI[2]] = NA

  auxList = list(Conditional=auxMatrix, Marginal=auxMatrix)

  # Condicional

  xmean = apply(modeloAux$x, 2, weighted.mean, w=modeloAux$prior.weights)

  betaAux = modeloAux2$coefficients
  varbetaAux = vcov(model)

  for(k in 1:p){
    ##### Conditional
    XcondAuxXk = xmean

    # Xk = x+1
    XcondAuxXk[k+1] = xx[k]+1
    etaxkC = drop(betaAux %*% XcondAuxXk)

    # P(X_k = x+1)
    PXk1C = 1/(1+exp(-etaxkC))
    # \Delta P(X_k = x+1)
    DeltaPXk1C = XcondAuxXk * PXk1C * (1 - PXk1C)

    # X_k = x
    XcondAuxXk[k+1] = xx[k]
    etaxkC = drop(betaAux %*% XcondAuxXk)

    # P(X_k = x)
    PXk0C = 1/(1+exp(-etaxkC))
    # \Delta P(X_k = x)
    DeltaPXk0C = XcondAuxXk * PXk0C * (1 - PXk0C)#exp(-etaxkC) / (1+exp(-etaxkC))^2
  }
}
```

(continues)

Figure 1 (continued)

```

##### Conditional PR_Xk
PRXkC = PXk1C / PXk0C

DeltaPRxkC = (DeltaPXk1C * PXk0C - DeltaPXk0C * PXk1C) / PXk0C^2
varPRxkC = crossprod(DeltaPRxkC, varbetaAux) %*% DeltaPRxkC

auxList$Conditional[k,] = c( PRXkC, sqrt(varPRxkC), PRXkC*exp(
qnorm(c(.025,.975))*sqrt(varPRxkC)/PRXkC ))

##### Marginal
XmargAux = modeloAux$x

# Xk = 1
XmargAux[(k+1)] = xx[k]+1
etaxkM = drop(XmargAux %*% betaAux)

PXk1M = mean(1/(1+exp(-etaxkM)))

auxVarMarg = exp(-etaxkM) / (1+exp(-etaxkM))^2

DeltaXk1M = 0
for(j in 1:n)
  DeltaXk1M = DeltaXk1M + XmargAux[j,] * auxVarMarg[j] / n

# Xk = 0
XmargAux[(k+1)] = xx[k]
etaxkM = drop(XmargAux %*% betaAux)

PXk0M = mean(1/(1+exp(-etaxkM)))

auxVarMarg = exp(-etaxkM) / (1+exp(-etaxkM))^2

DeltaXk0M = 0
for(j in 1:n)
  DeltaXk0M = DeltaXk0M + XmargAux[j,] * auxVarMarg[j] / n

##### Marginal PR_Xk
PRXkM = PXk1M / PXk0M

DeltaPRXkM = (DeltaXk1M * PXk0M - DeltaXk0M * PXk1M) / PXk0M^2
varPRXkM = crossprod(DeltaPRXkM, varbetaAux) %*% DeltaPRXkM

auxList$Marginal[k,] = c( PRXkM, sqrt(varPRXkM), PRXkM*exp(
qnorm(c(.025,.975))*sqrt(varPRXkM)/PRXkM ))
}

switch(type,
  marginal = return(auxList$Marginal),
  conditional = return(auxList$Conditional),
  both = return(auxList)
)
}

```

-1.96 to 1.96. The crude prevalence ratio (1.477) underestimates this range of the true prevalence ratio, whereas the crude and the adjusted prevalence odds ratio (2.528 and 2.537) overestimate the true range (although their confidence inter-

vals overlap with some of the true range) (Table 1). The adjusted prevalence ratios are all very similar, and all provide reasonable estimates (Table 1). The estimates differ only in the second or third decimal places, with the smallest estimated

Table 1

Adjusted prevalence ratios and respective 95% confidence interval (95%CI) estimates in the analysis of toy data using Y as the outcome, X as the risk factor and the continuous covariate Z as a control factor.

Regression model (measure)	Estimate	95%CI
Robust poisson (PR)	1.950	1.573, 2.416
Log-binomial (PR)	1.942	1.575, 2.418
Logistic regression (POR)	2.537	1.905, 3.398
Logistic regression (CPR)	1.956	1.578, 2.425
Logistic regression (MPR)	1.949	1.574, 2.414

CPR: conditional prevalence ratio; MPR: marginal prevalence ratio; POR: prevalence odds ratio; PR: prevalence ratio.

Note: the true conditional prevalence ratio for X varies from 1.71 to 2.25 depending on the value of Z (-1.96 to 1.96),

value in the log-binomial model and the largest in the conditional prevalence ratio.

Application 2: underweight in 4-5 year-old children in Pelotas

Table 2 presents the adjusted prevalence ratio of the occurrence of underweight in 4-5 year-old children (outcome) by previous hospitalization (exposure) controlled by birth weight (normal or low birth weight).

Despite the low prevalence of the outcome (4.1%), a difference was observed of 0.169 between the crude PR (2.902) and the crude POR (3.071) for previous hospitalization. According to the crude PR, there was a greater prevalence of underweight among children who were previously hospitalized when compared with those without previous hospitalization. The adjusted prevalence ratios of the log-binomial, robust Poisson, marginal prevalence ratio and Mantel-Haenszel method presented similar estimates (2.481, 2.479, 2.460, and 2.483, respectively). The largest adjusted estimates were the POR (2.641) and the conditional prevalence ratio (2.532).

Application 3: cervical cytological abnormalities in HIV-infected women

Table 3 shows the influence of high risk HPV (exposure) in the occurrence of cervical cytological abnormalities in HIV-infected women, controlled by age, number of pregnancies and time since last gynecological examination. Despite the low prevalence, the crude POR (7.909) differed from the crude PR (7.360) by 0.6. Those women with high risk HPV presented 640% more cytological abnormalities. The adjusted POR was the highest estimated value (7.990). The adjusted preva-

lence ratios obtained by the log-binomial, robust Poisson approach and the marginal prevalence ratio were very similar. The conditional prevalence method led to a ratio up to 46% greater than those obtained from other adjusted methods.

Discussion

Difficulties in obtaining prevalence ratios in cross-sectional studies have been investigated by several authors in recent years. Several authors use strategies for indirect calculation of the PR using the Breslow-Cox and Poisson models (with and without robust variance), while others interpret the prevalence odds ratio obtained in logistic regression models as a prevalence ratio.

Lee & Chia⁹ were the first authors to discuss methods proposed for estimating the PR. Until then, most cross-sectional studies in health used the logistic regression model estimate (POR), since it has the advantage of adjusting for the confounding or modifying effects of other variables. When the outcome is prevalent, however, the POR is a poor estimate of the prevalence ratio, overpredicting the PR by up to 27 times²³.

Regarding the estimation of the adjusted prevalence ratio, in our examples the log-binomial model, robust Poisson model, and marginal prevalence ratio provided similar estimates. The conditional prevalence ratio (CPR) differed from the other estimates but was still smaller than the adjusted POR. The CPR proposed by Wacholder¹³ is the prevalence ratio conditional on the mean values of the covariates, yet one could condition on any value for the confounding variables (higher or lower risk scenario). For instance, the prevalence of cervical cytological abnormalities in the 703 HIV-infected women (Application 3)

Table 2

Adjusted prevalence ratios and respective 95% confidence interval (95%CI) estimates in the analysis of the data using underweight as the outcome (Y), previous hospitalization as the risk factor (X) and birth weight as the control factor (Z).

Regression model (measure)	Estimate	95%CI
Mantel-Haenszel (PR)	2.483	1.456, 4.235
Robust poisson (PR)	2.479	1.454, 4.226
Log-binomial (PR)	2.481	1.447, 4.226
Logistic regression (POR)	2.641	1.481, 4.671
Logistic regression (CPR)	2.532	1.471, 4.357
Logistic regression (MPR)	2.460	1.451, 4.171

CPR: conditional prevalence ratio; MPR: marginal prevalence ratio; POR: prevalence odds ratio; PR: prevalence ratio.

Table 3

Adjusted prevalence ratios and respective 95% confidence interval (95%CI) estimates in the analysis of cervical cytological abnormalities in HIV-infected women considering high risk HPV as the exposure variable (X), and controlling for three other variables (Z).

Regression model (measure)	Estimate	95%CI
Robust poisson (PR)	7.123	2.489, 20.388
Log-binomial (PR)	7.192	2.849, 24.135
Logistic regression (POR)	7.990	3.029, 27.531
Logistic regression (CPR)	7.529	2.617; 21.665
Logistic regression (MPR)	7.118	2.518; 20.124

CPR: conditional prevalence ratio; MPR: marginal prevalence ratio; POR: prevalence odds ratio; PR: prevalence ratio.

Note: Z variables = age, number of pregnancies, and time since last gynecological examination.

was estimated for those women who had high risk HPV ($X = 1$), based on their respective values for age, number of pregnancies, and time since last gynecological examination (Z variables), and the mean value of the prevalence was computed. Similar calculations were performed for those women who were not diagnosed with high risk HPV ($X=0$). More detailed information on conditional and marginal methods is well described in Wacholder¹³ and Wilcosky & Chambless¹⁷.

The main advantage of the log-binomial and robust Poisson models is that they are already implemented in most popular statistical packages. The log-binomial has the disadvantage of not using a proper link function, leading to numerical instability in the estimation process and resulting in non-convergence issues. The COPY method⁶ was proposed to achieve convergence with the log-binomial model, but this method is

only available in SAS, which is a proprietary software. The robust Poisson model assumes that all the events in the database occurred at the same time. In addition, use of the Poisson distribution is not appropriate for modeling a binary outcome⁶. However, it is important to highlight that the likelihood of the Poisson model has been used only to obtain an estimation equation and not for the purpose of modeling a binary response variable. The Schouten et al.¹⁴ approach can be implemented easily on any statistical package, but in changing the database, it brings extra uncertainty that should be properly treated. The approach used by Wilcosky & Chambless¹⁷, unlike the log-binomial model, does not suffer from convergence problems.

One limitation of our results is that there is no "gold standard" for choosing the best method, especially when there is a continuous explana-

tory variable. In Application 2, where all explanatory variables were binary, the Mantel-Haenszel method was used as the “gold standard”. For this application, we found that the prevalence ratio estimated by the log-binomial model, Poisson robust model and marginal prevalence ratio showed estimates similar to the one obtained by the Mantel-Haenszel method. We thus conclude that, in this application, the equivalence of the models applied. In this paper we have not explored robust methods based on quasi-likelihood estimation¹⁵.

In summary, we recommend the use of the direct approach proposed by Wilcosky & Chamb-

less¹⁷ because it is suitable for a binary response when using a variable binomial model, it has no convergence difficulties and it is now available as a package for the open source statistical software R. The estimates of the marginal prevalence ratio are similar to those of other methods, while the conditional prevalence ratio shows the prevalence ratio for an average person in the database. If one is interested in a particular set of control variables, one only need specify the values of those variables. The authors are developing an R package with the Wilcosky & Chamberless approach, which will be available along with this article.

Resumen

En las últimas décadas, se ha discutido el uso de la razón de prevalencia (RP), en lugar del odds ratio como medida de asociación que se estima en estudios transversales. Se analizan las principales dificultades en el uso de modelos estadísticos para el cálculo de la RP: problemas de convergencia, disponibilidad de herramientas y supuestos no apropiados. El objetivo es realizar un enfoque directo para estimar la RP desde modelos logísticos binarios, basados en dos métodos propuestos por Wilcosky y Chambler y compararlos con otros métodos. Se han utilizado 3 ejemplos y comparamos las estimaciones crudas y ajustadas de RP con las estimaciones obtenidas por log-binomial, Poisson y odds ratio de prevalencia (ORP). Los RP obtenidos del enfoque directo dieron como resultado valores cercanos a los obtenidos mediante el log-binomial y de Poisson, mientras que la RCP sobreestimó la RP. El modelo que aquí se presenta implementó las siguientes ventajas: no presenta inestabilidad numérica, toma una distribución de probabilidad apropiada y está disponible en software estadístico libre R.

Razón de Prevalencias; Modelos Logísticos; Estudios Transversales

Contributors

L. S. Bastos participated in the drawing up of the R function, literature revision, and article write up. R. V. C. Oliveira and L. S. Velasque collaborated in the literature revision and article write up.

Acknowledgments

The authors would like to thank the researchers Claudio Struchiner and Francisco Inácio Bastos for their valuable suggestions and Daniel Stokes by the English review.

References

1. Axelson O, Fredriksson M, Ekberg K. Use of the prevalence ratio v the prevalence odds ratio as a measure of risk in cross sectional studies. *Occup Environ Med* 1994; 51:574.
2. Stromberg U. Prevalence odds ratio v prevalence ratio. *Occup Environ Med* 1994; 51:143-4.
3. Strömberg U. Prevalence odds ratio v prevalence ratio: some further comments. *Occup Environ Med* 1995; 52:143.
4. Skov T, Deddens J, Petersen M, Endahl L. Prevalence proportion ratios: estimation and hypothesis testing. *Int J Epidemiol* 1998; 27:91-5.
5. Barros AJ, Hirakata VN. Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. *BMC Med Res Methodol* 2003; 3:21.
6. Deddens JA, Petersen MR, Lei X. Estimation of prevalence ratios when PROC GENMOD does not converge. In: *Proceedings of the 28th Annual SAS Users Group International Conference*. Cary: SAS Institute Inc.; 2003. p. 270-328.
7. Coutinho L, Scazufca M, Menezes P. Métodos para estimar razão de prevalência em estudos de corte transversal. *Rev Saúde Pública* 2008; 42:992-8.
8. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic research: principles and quantitative methods*. Hoboken: John Wiley & Sons; 1982.
9. Lee J, Chia K. Estimation of prevalence rate ratios for cross sectional data: an example in occupational epidemiology. *Br J Ind Med* 1993; 50:861-2.
10. Lin D, Wei LJ. The robust inference for the Cox proportional hazards model. *J Am Stat Assoc* 1989; 84:1074-8.
11. Clayton D, Hills M. *Statistical models in epidemiology*. Oxford: Oxford University Press; 1993.
12. Zou G. A modified Poisson regression approach to prospective studies with binary data. *Am J Epidemiol* 2004; 159:702-6.
13. Wacholder S. Binomial regression in GLIM: estimating risk ratios and risk differences. *Am J Epidemiol* 1986; 123:174-84.
14. Schouten E, Dekker J, Kok F, Cessie S, van Houwelingen H, Pool J, et al. Risk Ratio and Rate Ratio estimation in case-cohort designs: hypertension and cardiovascular mortality. *Stat Med* 1993; 12:1733-45.
15. Lumley T, Kronmal R, Ma S. Relative risk regression in medical research: models, contrasts, estimators, and algorithms. *UW Biostatistics Working Paper Series*. (Working Paper 293). <http://biostats.bepress.com/uwbiostat/paper293> (accessed on 20/Jul/2013).
16. Diaz-Quijano FA. A simple method for estimating relative risk using logistic regression. *BMC Med Res Methodol* 2012; 12:14.
17. Wilcosky TC, Chambless LE. A comparison of direct adjustment and regression adjustment of epidemiologic measures. *J Chronic Dis* 1985; 38:849-56.
18. Lane PW, Nelder JA. Analysis of covariance and standardization as instances of prediction. *Biometrics* 1982; 38:613-21.
19. Souza CL, Carvalho FM, Araújo TM, Reis EJFB, Lima VMC, Porto LA. Fatores associados a patologias de pregas vocais em professores. *Rev Saúde Pública* 2011; 45:914-21.
20. Lee J. Covariance adjustment of rates based on the multiple logistic regression model. *J Chronic Dis* 1981; 34:415-26.
21. Flanders WD, Rhodes PH. Large sample confidence intervals for regression standardized risks, risk ratios, and risk differences. *J Chronic Dis* 1987; 40:697-704.
22. Luz P, Velasque L, Friedman R, Russomano F, Andrade A, Moreira R, et al. Cervical cytological abnormalities and factors associated with high-grade squamous intraepithelial lesions among HIV-infected women from Rio de Janeiro, Brazil. *Int J STD AIDS* 2012; 23:12-7.
23. Lee J. Odds ratio or relative risk for cross-sectional data? *Int J Epidemiol* 1994; 23:201-3.

Submitted on 08/Oct/2013

Final version resubmitted on 18/Nov/2014

Approved on 01/Dec/2014