SCAALA

Social changes, asthma and allergy in Latin America.

Estimating Adjusted Prevalence Ratio in Clustered Cross-Sectional Epidemiological Data

Carlos A. Teles, Leila D. Amorim, Rosemeire L. Fiaccone, Sérgio Cunha,Maurício L. Barreto, Maria Beatriz B. do Carmo, Ana Lucia M.Benalcazar and Nelson F. Oliveira

INTRODUCTION

- It is well known that OR can provide a good approximation to prevalence ratios (PR) in crosssectional studies when the underlying outcome is rare.

- It is also known that the OR overestimates the RR(or PR) when the outcome of interest is common (larger than 10%, for instance).

- The major limitation of using OR in such scenarios is related to its misinterpretation as a measure of risk.

INTRODUCTION

- There are several statistical models that can provide adjusted estimates for PR, including logistic model, Poisson regression and log-binomial regression.

- There is no concensus about the best approach to obtain the adjusted PR and these methods may lead to different conclusions.

OBJECTIVES

- In this paper we **describe methods of estimation of adjusted PR** for the settings of **clustered** cross-sectional studies:
  - taking into account the degree of similarity between subjects within the same cluster;
  - using random-effects models.

- We also **evaluated the performance of different approaches** to estimate adjusted PR.

METHODS

- Definition of statistical approaches.
- Evaluation of methods using simulation studies.
- Illustration of use of methods through two data analysis using data from SCAALA Study and Esmeraldas-Ecuador Study.

METHODS

- Consider that p_1 and p_0 denote, respectively, the prevalence of an event in exposed and unexposed groups.

- Thus, OR and PR can be defined, respectively, as:

$$OR = \frac{\left(\frac{p_1}{1-p_1}\right)}{\left(\frac{p_o}{1-p_o}\right)}$$

$$PR = \frac{p_1}{p_o}$$

- The interpretation of these two measures, therefore, are not the same, unless as mentioned before, the event is rare.

METHODS

- Statistical models are used in order to obtain adjusted measures of association, controlling for the effect of several potential confounders.

- Using logistic regression (the most popular model), the adjusted OR and PR can be defined as:

$$OR = \exp(\beta_1)$$

METHODS

- The estimation of PR, on the other hand, requires a more complicated mathematical expression that relates the effects and the values of the factors of interest.

- For instance, suppose it is of interest to evaluate the effect of an exposure ($X1$) to the occurrence of an outcome while controlling for k - 1 confounders ($X2;..;Xk$). In such case, the PR between exposed and unexposed subjects could be expressed as:

$$PR = \frac{1 + \exp\{-\beta_0 - \beta_2 X_2 - \ldots - \beta_k X_k\}}{1 + \exp\{-\beta_0 - \beta_1 - \beta_2 X_2 - \ldots - \beta_k X_k\}}$$

✓Note that the PR depends on the values of the covariates in the model

METHODS

- **Alternative Procedures: Poisson and log-binomial models**

➢ A major limitation for the Poisson is that it allows prediction of probabilities out of the interval [0; 1].

➢ The log-binomial models usually do not converge.

- **Approach for Clustered Data: Logistic Model with Random Effects**

➢ A well-established approach to modeling clustered/correlated data introduces random effects in the model of interest.

➢ It takes into account adjustment on non-observed individual characteristics reflecting a natural heterogeneity across subjects.

METHODS

- **Confidence Intervals for Prevalence Ratios using Logistic Model with Random Effects:**

➢ Methods for obtaining large sample confidence intervals for prevalence ratios include the **delta method and bootstrap**:

✓ The **delta method** produces an approximated standard error for the RP.

✓ The **bootstrap** is based on resampling the data with replacement to obtain an empirical standard error.

SIMULATION STUDIES

- Simulation studies were done to evaluate the performance of delta and bootstrap methods to obtain CI for PR.

- Configuration of the simulation studies:

➢ variation of **degree of dependency** (**intraclass correlation coefficient – ICC=0.03 , 0.29 and 0.71)**

➢ and levels of clustering **number (k=10, 30 and 100) and size of clusters (m=10 and 30)** .

EPIDEMIOLOGICAL STUDIES

- **SCAALA- Salvador Study**

  ✓ Analysis considered data from 758 children, aged 4 to 12 years-old.

  ✓ Aim of the analysis: estimated the effect of maternal mental health status on asthma occurrence.

  ✓ Models considered: random effects logistic regression, robust Poisson regression and robust log-binomial regression.

  ✓ Data analysis was done using STATA v.8 and R v.2.6.0 software.

RESULTS-EPIDEMIOLOGICAL STUDIES

Among these children, 38.2% are suspect of having asthma.

Table I. Comparison of Prevalence Ratio (PR) Estimates using Random Effects Logistic Regression with Different Standardization Procedures: Impact of Maternal Mental Health Status on Child's Asthma in Brazil.

Regression Models	PR	95%CI
Random Effects Logistic	1.58	
Delta method		(1.29, 1.92)
Bootstrap Method		(1.15, 1.86)
Robust Poisson	1.55	(1.28, 1.89)

- The estimated odds ratio is 2.2 (95%CI=1.61, 3.01), which is larger than the estimated prevalence ratio (PR=1.58; 95%CI= (1.15;1.86)).

RESULTS-EPIDEMIOLOGICAL STUDIES

- **Comments:**

➢ The results obtained using the robust Poisson model(PR=1.55,95%CI=1.28,1.89) were very close from the random effects logistic regression;

➢ The convergence was not achieved using the log-binomial model;

➢ Note that the 95% bootstrap condence intervals are less precise than those obtained from the delta method.

EPIDEMIOLOGICAL STUDIES

- **Esmeralda-Ecuador Study**
  - ✓ Aim of this study was to compare the prevalence of infection for some geohelminths between areas where ivermectina was or was not administered (main effect).
  - ✓ Data from a subset of 2000 children aged 6 to 16 years-old was used.
  - ✓ Prevalence of Trichuris infection was 57.9% (very common event).

RESULTS-EPIDEMIOLOGICAL STUDIES

Table II. Prevalence Ratio (PR) Estimation using Standard and Random E_ects Logistic Regression with Conditional Standardization and Robust Poisson Model, with corresponding 95% con_dence intervals: Evaluation of E_ectiveness of a Health Program in Ecuador.

Regression Models		PR	95% CI
Standard Logistic		0.38	
	Delta Method		(0.34;0.42)
	Bootstrap Method		(0.34;0.42)
Random Effects Logistic		0.33	
	Delta Method		(0.27;0.42)
	Bootstrap Method		(0.30;0.39)
Robust Poisson		0.38	(0.31;0.47)

- OR overestimated effect of ivermectina (OR=0.07 [95%CI=0.05; 0.11]) compared to PR = 0.33 (95%CI=0.27; 0.42).

RESULTS-EPIDEMIOLOGICAL STUDIES

- **Comments:**

➢Bootstrap CI's were narrower than those obtained through delta method using random effects logistic model.

➢Convergence was not achieved using log-binomial models.

➢ Estimated ICC=0.415.

➢In general, CI's using random effects logistic model were wider than those from the standard logistic regression.

Simulation Results

We present the results for the simulation study comparing the coverage probability (CP) of the Wald 95% confidence interval obtained through delta method and bootstrap for random effects logistic model (Table 3).

The prevalence of disease in each of the configurations was between 55% and 60%, with a PR of 1.52

Table 3: Coverage probability of the Wald 95% confidence interval for delta method and bootstrap in the estimation of prevalence ratios considering the random effects logistic model for different degree of correlation, number and size of clusters.

Sample	ICC=0.03		ICC=0.29		ICC=0.71	
Size	Delta	Bootstrap	Delta	Bootstrap	Delta	Bootstrap
Number of clusters=10						
m=10	94.7%	88.3%	92.7%	88.0%	95.0%	87.3%
m=30	93.7%	93.0%	91.7%	88.3%	92.0%	89.0%
Number of clusters=30						
m=10	95.3%	94.0%	90.3%	91.0%	93.3%	92.0%
m=30	92.3%	90.0%	94.0%	93.3%	94.0%	93.6%
Number of clusters=100						
m=10	94.3 %	94.3%	92.7%	94.0%	91.8%	92.5%
m=30	93.9%	92.9%	95.7%	95.3%	95.0%	94.3%

Table 4: Comparison of coverage probability of the Wald 95% confidence interval for estimation of prevalence ratios using random effects logistic model (with delta method) and random effects Poisson model regression for different degree of correlation and number of clusters of size 10.

ICC and Numb.clusters (k)	Random effects logistic model	Random effects Poisson model
ICC=0.03		
k=15	.95 %	98%
k=30	.94%	95%
k=50	.96%	89%
ICC=0.29		
k=15	.92 %	93%
k=30	.93%	87%
k=50	.94%	75%
ICC=0.71		
k=15	.91 %	85%
k=30	.93%	81%
k=50	.92%	73%

According to the results, the logistic model generally has a better performance than the Poisson model.

DISCUSSION

- Logistic regression can be used to obtain adjusted prevalence ratios (PR).

- More than one approach to obtain CI's for PR using logistic regression.

- Other model-based approaches that has been commonly used to estimate PR are Poisson and log-binomial models, which have advantage related to the direct estimation of PR and confidence intervals.

DISCUSSION

Barros and Hirakata (2003) suggested that more than one modeling strategy should be used to evaluate the robustness of the results.

We are extending the macros to estimate PR proposed by Localio et al (2007) at STATA (version 7) for clustered data.

Implementation of these methods are already available using R software.