Modeling Spatial Heterogeneity with Excess Zeroes from School Absenteeism dSata

Xiaoxiao Song*[1, 2], Qi Zhao[1], Changming Zhou[1], Tao Tao[1], Lars Palm[3], Vinod Diwan[4], Hui Yuan[5] and Biao Xu[1]

[1]Dept. of Epidemiology, School of Public Health, Fudan University, Shanghai, China; [2]School of Public Health, Kunming Medical University, Kunming, China; [3]Future Position X, Gavle, Sweden; [4]ICHAR, Karolinska Instituet, Stockholm, Sweden; [5]Jiangxi Provincial Center for Disease Prevention and Control, Nanchang, China

Objective

To describe and explore the spatial heterogeneity via Random effects zero-inflated Poisson model (RE-ZIP) for absenteeism surveillance in primary school for early detection of infectious disease outbreak in rural China.

Introduction

Absenteeism has been considered as a potential indicator for the early detection of infectious disease outbreaks in population, especially in primary schools. However, in practice this data are often characterized by an excess of zeros and spatial heterogeneity. In a project on integrated syndromic surveillance system (ISSC) in rural China(1), Random effect zero-inflated Poisson (RE-ZIP) model was applied to simultaneously quantify the spatial heterogeneity for "occurrence" and "intensity" on school absenteeism data.

Methods

Daily health-related school absenteeism at 62 primary schools in two counties of Jiangxi Province were reported to the web-based platform of ISSC during 24 continuous months (from April 1th, 2012 to June 31st, 2014). The RE-ZIP has two components that correspond to two zero generating processes with two random intercept (μ1 and μ2). The first process is governed by a binary distribution that generates structural zeros. The second process is governed by a Poisson distribution that generates counts, some of which may be zero (named sampling zero) (2), see Eqation.1-2. Firstly, artificial data was generated to illustrate that the two random effects could appropriately describe the spatial heterogeneity for "occurrence" and "intensity" of school absenteeism (Figure_1). Then, the spatial heterogeneity for "occurrence" and "intensity" in real absenteeism data from the 62 primary schools were assessed and visualized through quantifying the two random effects in each individual school. All analyses were processed using SAS NLMIXED(3).

Results

In total, 32283 health-related absenteeism records were reported from the 62 primary schools in the study period. Figure_2 displayed clearly the spatial heterogeneity for both "occurrence" and "intensity" in these primary schools indicated by μ1 and μ2 of the two random effects.

The spatial variation in these two counties were also visualized graphically these on map in Figure_3 and Figure_4.

Conclusions

In infectious disease surveillance, the key questions are "whether an outbreak will happen" or/and "how large the epidemic will be", which could be answered through observing and monitoring the spatial pattern of health related school absenteeism. Findings from this study suggest that RE-ZIP model is an effective tool in the surveillance of health-related school absenteeism data with high spatial heterogeneity.

$$\log\left(\frac{\pi}{1-\pi}\right) = \text{logit}(\pi) = W\gamma + \mu_1 \qquad \text{Equation 1}$$

$$\log(\lambda) = X\beta + \mu_2 \qquad \text{Equation 2}$$

Figure_1 artificial data to illustrate the source of zero in ZIP.

Figure_2 Estimation the two random effects with 95% CIs for 62 schools.

Figure_3 Estimation u1 for each school.

Figure_4 Estimation u2 for each school.
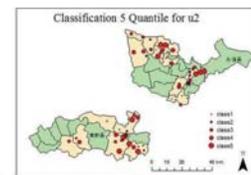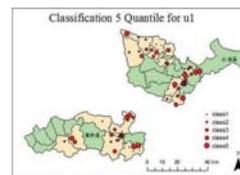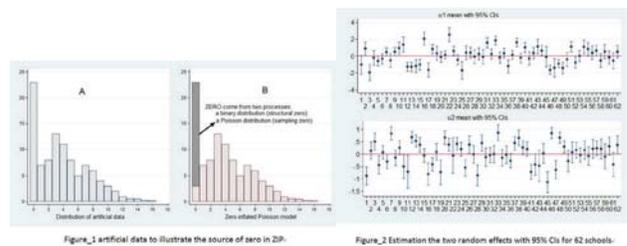
Keywords

School absenteeism; spatial heterogeneity; Random effects zero-inflated Poisson model; occurrence; intensity

References

1. Yan W-r, Nie S-f, Xu B, Dong H-j, Palm L, Diwan VK. Establishing a web-based integrated surveillance system for early detection of infectious disease epidemic in rural China: a field experimental study. BMC medical informatics and decision making2012;12(1):4.
2. Hu M-C, Pavlicova M, Nunes EV. Zero-inflated and hurdle models of count data with extra zeros: examples from an HIV-risk reduction intervention trial. The American journal of drug and alcohol abuse2011;37(5):367-75.
3. Wang J, Xie H, Fisher JF. Multilevel models: applications using SAS®: Walter de Gruyter; 2011.

*Xiaoxiao Song
E-mail: chinasxx@gmail.com