Applying Zero-inflated Mixed Model to School Absenteeism Surveillance in Rural China

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

[1]School of Public Health, Fudan University, Shanghai, China; [2]Jiangxi Provincial Center for Disease Prevention and Control, nanchang, China; [3]Future Position X, Gavle, Sweden; [4]ICHAR, Karolinska Institute, Stockholm, Sweden

Objective

To describe and explore the spatial and temporal variability via ZIMM for absenteeism surveillance in primary school for early detection of infectious disease outbreak in rural China.

Introduction

Absenteeism has great advantages in promoting the early detection of epidemics1. Since August 2011, an integrated syndromic surveillance project (ISSC) has been implemented in China2. Distribution of the absenteeism generally are asymmetry, zero inflation, truncation and non-independence3. For handling these encumbrances, we should apply the Zero-inflated Mixed Model (ZIMM).

Methods

Data for this study was obtained from the web-based data of ISSC in 62 primary schools in two counties of Jiangxi province, China from April 1th, 2012 to June 30st, 2012. The ZIMM was used to explore: 1)the temporal and spatial variability regarding occurrence and intensity of absenteeism simultaneously, and 2) the heterogeneity among the reporting primary schools by introducing random effects into the intercepts. The analyse was processed in the SAS procedure NLMIXED4.

Results

The total 4914 absenteeism events were reported in the 62 primary schools in the study period. The rate of zero report was 49.88% (Fig.1). According to ZIMM, there are fixed and random effect parameters in this model (Table 1). Firstly, for the fixed parameters, the spatial variable (county) was not significantly different both the occurrence and intensity model, while for the temporal variable (month), the probability of absenteeism occurrence was significantly different over three months (β=-0.165, p =0.026), suggesting a decreasing of school absenteeism from April to June. Meanwhile, a statistical significant difference in the intensity of absenteeism was also found over the three months (β=-0.073, p=0.007). Secondly, the random effect of intensity model was statistically significance (p=0.008), which strongly indicated a heterogeneity in intensity of absenteeism among the surveillance schools. Whereas the random effect of occurrence model by logistic regression showed a non-statistical difference (p=0.774) among the schools suggesting the homogeneity in the occurrence of absenteeism.

Conclusions

School absenteeism data has greater uncertain than many other sources and easier fluctuate by some factors such as holiday, season, family status and geographic distribution. Thus, the spatial and temporal dynamics should be taken into account in controlling fluctuate of absenteeism. Moreover, school absenteeism data are correlated within each school due to repeated measures. Applying the ZIMM, the occurrences and intensity of absenteeism could be evaluated to reduce the bias and improve the prediction precision. The ZIMM is an appropriate tool for health authorities in decision making for public health events.

Table 1 Fixed parameters and variance components estimates for the absenteeism using ZIMM

		Logistic regression parameter with occurrence			lognormal regression parameter with intensity		
parameters		β	Std Err	p value	β	Std Err	p value
Fixed parameters	Intercept	-0.733	0.262	0.005	0.718	0.039	0.000
	county	-0.188	0.103	0.068	-0.020	0.042	0.632
	month	-0.165	0.074	0.026	-0.073	0.027	0.007
Variance components	Var(Randdm Effect)	0.548	1.906	0.774	0.316	0.120	0.009
	Residual				0.120	0.119	0.313

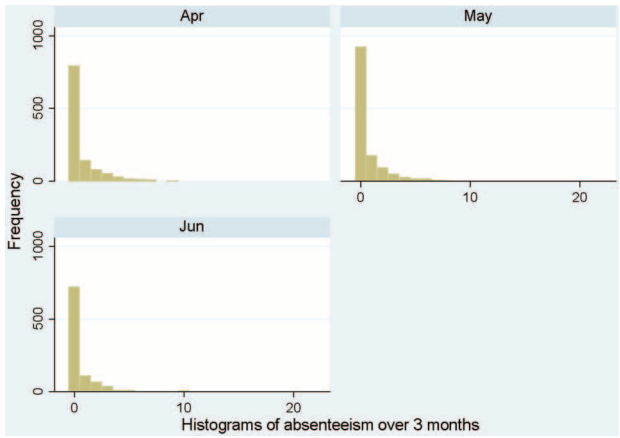

Fig. 1 Absenteeism from Apr. 1st to Jun. 30th 2012

Keywords

surveillance; absenteeism; zero-inflated mixed model; occurrence; intensity

References

1. Lenert L, Kirsh D, Johnson J, Aryel RM. Absenteeism. In: Wagner MM, Moore AW, Aryel RM, editors. Handbook of Biosurveillance Burlington: Academic Press, 2006:361-68.
2. 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 Making 2012;12.
3. Calama R, Mutke S, Tome J, Gordo J, Montero G, Tome M. Modelling spatial and temporal variability in a zero-inflated variable: The case of stone pine (Pinus pinea L.) cone production. Ecological Modelling 2011;222(3):606-18.
4. Tooze JA, Grunwald GK, Jones RH. Analysis of repeated measures data with clumping at zero. Statistical Methods in Medical Research 2002;11(4):341-55.

*Biao Xu
E-mail: bxu@shmu.edu.cn