Policy Forum

Association between Macroscopic-factors and Identified HIV/AIDS Cases among Injecting Drug Users: An Analysis Using Geographically Weighted Regression Model*

XING Jian Nan[1,†], GUO Wei[1,†], QIAN Sha Sha[1], DING Zheng Wei[1], CHEN Fang Fang[1], PENG Zhi Hang[2], QIN Qian Qian[1], and WANG Lu[1,#]

Drug use (DU), particularly injecting drug use (IDU) has been the main route of transmission and spread of Human Immunodeficiency Virus (HIV)/Acquired Immune Deficiency Syndrome (AIDS) among injecting drug users (IDUs)[1]. Previous studies have proven that needles or cottons sharing during drug injection were major risk factors for HIV/AIDS transmission at the personal level[2-4]. Being a social behavioral issue, HIV/AIDS related risk factors should be far beyond the personal level. Therefore, studies on HIV/AIDS related risk factors should focus not only on the individual factors, but also on the association between HIV/AIDS cases and macroscopic-factors, such as economic status, transportation, health care services, etc[1]. The impact of the macroscopic-factors on HIV/AIDS status might be either positive or negative, which are potentially reflected in promoting, delaying or detecting HIV/AIDS epidemics.

China has reported HIV/AIDS infection among IDUs since domestic HIV/AIDS cases were first identified among them in Yunnan Province in 1989[5]. In response to HIV/AIDS epidemic, China has established a real name based HIV/AIDS case reporting system (CRS) since 1985[6]. All hospitals and centers for diseases control and prevention (CDCs) are obligated to report identified HIV/AIDS cases. According to the legislation, once a HIV/AIDS case was identified by Western Blot assay, an interview will be conducted by health care personnel to provide counseling to him/her within 10 work days. A case report form (CRF) should be completed and delivered to the National Center for AIDS/STD Control and Prevention (NCAIDS) after a face to face interview. Demographic information (age, gender, occupation, address, etc.) and the history of high risk behaviors of the interviewed cases were collected based on CRF. And the most likely route of HIV infections was judged by healthcare personnel according to the history of their high risk behaviors. By collecting demographic information of people living with HIV/AIDS, CRS provided both the number of HIV/AIDS cases and the time and their spatial distribution, which enabled us to explore the spatial distribution characteristics of HIV/AIDS cases among IDUs[7].

During spatial analysis, autocorrelation, instability and heterogeneity of the spatial data should be considered carefully[8-9]. General spatial autocorrelation is a technique which is used to detect the spatial cluster of data, measure and analyze the degree of dependency among observations in a geographic space[8]. In order to analyze the variance and instability of spatial parameters, an effective and sample technique-geographically weighted regression (GWR) was initiated by Brunsdon et al., which allowed the value of coefficients to change at different spatial sites. GWR is produced to generate parameters disaggregated by the spatial units of analysis by local spatial regression. This allows assessment of the spatial heterogeneity and instability in the estimated associations between the independent and dependent variables by expanding ordinarily linearity regression through embedding spatial data structure into the regression model, and in this way the GWR model can reflect the spatial instability of the coefficients at different spaces[9-12].

In this study, we focus on the spatial autocorrelation of HIV/AIDS cases among IDUs in China from 2007 to 2011, and explore the association between the spatial distribution and

doi: 10.3967/ bes2014.055

1. National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, China; 2. Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing 210029, Jiangsu, China

macroscopic socio-economic factors, instead of personal level factors, by using general spatial autocorrelation and GWR model.

Data Management

HIV/AIDS positive IDUs, identified by CRF from 2007 to 2011, were downloaded from the web based China's National HIV/AIDS CRS, which was managed by the NCAIDS, China CDC. Name linked information, such as name, ID number, home address, etc., were deleted before data analysis. National standard geocode at the provincial level was used to identify which province the HIV/AIDS positive IDUs came from.

Population size and twenty-one original demographic and socio-economic indicators, including economic status, transportation, social security and healthcare services, etc., were collected at the provincial level from China Statistical Yearbook from 2007 to 2011, released by the National Bureau of Statistics of China. To demonstrate the average level of these socio-economic indicators during 2007-2011, arithmetical mean of these indicators during the five years was used to extract four synthesized variables by principal component analysis.

Variable Definition

Identified HIV/AIDS cases density for each province was the key variable in spatial autocorrelation analysis and acted as the dependent variable in GWR. Identified HIV/AIDS cases density was calculated at the provincial level as below to reflect the HIV/AIDS epidemic situation among IDUs.

$$\text{Identified cases density} = \frac{\text{Identified HIV/AIDS cases among IDUs from 2007-2011}}{\text{mean of population size from 2007-2011}/10\ 000} \quad (1)$$

By using principal component analysis[13], four synthesized variables were extracted from 21 original socio-economic indicators to reflect economic status, transportation, social security and health care services in each province. These four provincial synthesized variables acted as independent variables in GWR.

Principal Component Analysis

Principal component analysis using the method of dimension reduction, not only can extract a principal component from many variables in the premise of losing little information of variables, but can also handle the possible multicollinearity among variables. Cumulative variance contribution is a statistic parameter which assesses the result of principal component analysis, when the cumulative variance contribution≥70%, which denotes that the result is reliable and the extracted principal component can represent most of the information of variables. Factor loading represents the association between the extracted principal component and variables; the higher absolute value of factor loading is, the more information about variables is represented by the extracted principal component.

In this paper, we used principal component analysis to extract four provincial synthesized variables (macroscopic-factors), from each four genres of original variables (indicators): economic status, means of transportation, social security, and health care services. These four synthesized variables were expected to represent the overall situation and information which were interesting to us. The process of principal component analysis was conducted by SPSS20.0® and the four macroscopic-factors were embedded in the GWR model as independent variables.

General Spatial Autocorrelation

We used the spatial code in the record of each HIV/AIDS cases among IDUs to match the geographic information system. It was assumed that the provinces in the whole country did not differ from each other, and autocorrelation was applied to analyze the country as a whole in this paper. *Moran's Index* was used to analyze the general spatial autocorrelation and the value of *Moran's Index* was between [-1, 1]. When the value of *Moran's Index*>0 and *Z*-value>1.96 or the *Moran's Index* <0 and *Z*-value <-1.96, it represented clustered HIV/AIDS cases; if the value of *Moran's Index* was close to zero and *Z*-value was between -1.96 and 1.96, then it represented randomly distributed HIV/AIDS cases[8]. In this study, we conducted general spatial autocorrelation for identified cases density. If $P<0.05$ and *Moran's Index*≠0, it indicated that HIV/AIDS cases were clustered among IDUs in China, but the data were not independent, and it was too serious to use the traditional statistical model to analyze them, and so the GWR model was better to be used for analyzing these data[9].

Geographically Weighted Regression Model

The GWR model allows the value of coefficients to change at different spatial sites (in this paper, spatial sites were defined as 31 provinces) and it is