

应用地理信息系统技术对江苏省艾滋病疫情的空间分析

彭志行 羊海涛 成月佳 王璐 喻荣彬 还锡萍 沈洪兵 陈峰 汪宁

【摘要】 目的 探讨江苏省各县(区)HIV感染者空间分布规律,试建立疫情预测模型。**方法** 建立江苏省各县(市)HIV感染者人数疫情数据库,根据地理信息分析HIV感染者相关的地理特征,对HIV感染者人数进行趋势分析、空间自相关分析、空间线性回归分析。**结果** (1)江苏省HIV感染者总体呈随机分布,但局部地区存在感染者聚集现象,发现南京市城区的感染者人数表现为负自相关,苏州和无锡市城区两地为正自相关。铜山县、吴江市、南京市浦口区和城区、溧水县、无锡市城区、苏州市城区为江苏省HIV感染者高度聚集的热点区域。(2)建立江苏省HIV感染者疫情分布图和趋势分析图,全省HIV感染者分布在地理东西方向上的变化不大,南北方向上,苏南地区疫情远远比苏北和苏中地区严重。(3)运用最小二乘法进行HIV感染者人数的空间线性回归分析,江苏省HIV感染者人数与地理经纬度之间的线性回归关系无统计学意义($t=-1.045\ 103$, $P=0.299\ 904$; $t=-1.443\ 668$, $P=0.153\ 714$)。**结论** 应根据江苏省HIV感染者空间分布特点,制定预防和控制疫情传播的措施。

【关键词】 艾滋病;地理信息系统;空间自相关分析;趋势分析

Study on the spatial distribution of AIDS based on geographic information system in Jiangsu province PENG Zhi-hang¹, YANG Hai-tao², CHENG Yue-jia¹, WANG Lu³, YU Rong-bin¹, HUAN Xi-ping², SHEN Hong-bing¹, CHEN Feng¹, WANG Ning³. 1 Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing 210029, China; 2 Center for Disease Control and Prevention of Jiangsu Province; 3 National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention

Corresponding author: WANG Ning, Email: wangnbj@163.com; CHEN Feng, Email: dr.chenfeng@njmu.edu.cn

This work was supported by grants from the National Science and Technology Support Projects for the "Eleventh Five-Year Plan" of China (No. 2008ZX10001-003, 2009ZX10004-904), the National Natural Science Foundation of China (No. 81001288) and the Higher Education Department of Jiangsu Provincial Natural Science Research Project (No. 09KJB330004).

【Abstract】 Objective To analyze the spatial distribution of AIDS in every city of Jiangsu province, trying to describe the geographic characteristics of AIDS and to develop a prediction model. **Methods** Numbers of patients in Jiangsu province were collected, to establish the database for the geographic information system, then setting up a 'risk map' of the disease. Spatial, autocorrelation. Linear spatial analyses were used to study the patients' numbers. **Results** (1) Results from the autocorrelation analysis showed that the distribution of AIDS was clustered at some places and was at random on the whole. The results also indicated that the distribution of AIDS in Nanjing was of negative correlation, while that in Suzhou and Wuxi were of positive correlation but in Tongshan, Wujiang, Pukou, Nanjing, Lishui, Wuxi and Suzhou showed seven locations of clusters with significantly higher numbers of patients. (2) The trend of 'surface analysis map' indicated that the disease was more severe in the southern than in the northern parts of Jiangsu province. (3) Ordinary Least Squares method was finally used in the linear spatial regression and the results were: $t=-1.045\ 103$ ($P=0.299\ 904$); $t=-1.443\ 668$ ($P=0.153\ 714$) respectively. **Conclusion** According to the feature spatial distribution of the disease, effective measures should be taken to prevent and to keep the prevalence of AIDS under control.

【Key words】 AIDS; Geographic information system; Spatial autocorrelation analysis; Trend surface analysis

DOI: 10.3760/cma.j.issn.0254-6450.2011.01.011

基金项目:“十一五”国家科技重大专项(2008ZX10001-003, 2009ZX10004-904); 国家自然科学基金(81001288); 江苏省高校自然科学基金(09KJB330004)

作者单位:210029 南京医科大学公共卫生学院流行病与卫生统计学系(彭志行、成月佳、喻荣彬、沈洪兵、陈峰); 江苏省疾病预防控制中心(羊海涛、还锡萍); 中国疾病预防控制中心艾滋病性病预防控制中心(王璐、汪宁)

通信作者:汪宁, Email: wangnbj@163.com; 陈峰, Email: dr.chenfeng@njmu.edu.cn

1991 年江苏省确诊首例艾滋病患者,此后全省 HIV 感染者数不断上升。艾滋病同大多数传染病一样均存在空间分布特征,其传播和流行与地理因素存在非常紧密的联系。探索和寻找地理因素与疾病间的关系已成为研究传染病的热点。详细分析传染病发病的地理不平衡和分布状态可以有利于医疗卫生机构统筹资源分配和有效利用^[1,2]。地理信息系统(GIS)以地理空间数据库为基础,可对有关空间数据按地理空间位置进行采集、管理、储存、维护、模拟、分析研究和校正,进一步输出地理信息或以位置为基础的信息^[3,4]。本研究应用 Arc GIS 9.3 软件建立了江苏省县(市)级 HIV 感染者人数资料数据库,对省内 HIV 感染者的空间分布规律进行分析和研究,为进一步做好其防治工作提供理论依据。

资料与方法

1. 数据来源和研究区域:收集江苏省传染病疫情报告系统以县(区)为单位的 HIV 发病数据,并以此作为原始数据建立数据库。研究区域包括江苏省 106 个县级行政区划,以县(区)为研究单位,地理范围为东经 116.60°~121.67°、北纬 31.01°~34.89°。

2. 建立 GIS 数据库和分析图像:以江苏省 2006 年县级电子地图为基础地图,以“HIV 感染者人数”为标题建立新字段,将全省各县(区)数据输入其中,建立 DBF 4(dBASE IV)文件,作为 GIS 地图数据库。根据江苏省 HIV 感染者人数多少将各县(区)病例数分类,利用 Arc GIS 9.3 软件对江苏省的 HIV 感染者分布情况进行图形处理。用疫情分布图和趋势面图直观显示并进行相应分析,预测疾病发展趋势。趋势面分析是借助最小二乘法运用数学曲面拟合样本数据,建立二项多元回归模型^[5]。Arc GIS 9.3 软件工具可以在此基础上做出趋势面分析图,借助它可以分析出数据的系统性变化。

3. GIS 空间自相关分析:空间流行病学中的空间自相关是指空间位置上靠近的事物或现象在空间位置上的依赖关系^[6]。空间自相关分析的对象是点或面数据,如果没有空间自相关,那么事物或现象的分布就是随机的,空间分布规律就无法表现出来。空间自相关分析分为全局自相关分析和局部自相关分析两部分。

本研究采用 Moran's I 和 Getis 两种自相关分析法,从不同角度分析江苏省 HIV 感染者分布情况。其中 Moran's I 可揭示空间是否有集聚性以及自相关程度的强弱,其计算结果可得出三种结论,即正相

关(距离越近,相似程度越高,如高值与高值相邻或低值与低值相邻)、负相关(相邻区域观测值彼此不同,如低值与高值相邻)和零相关(即随机分布,不表现相关性)。Getis 即热点分析,则可以确定聚集的类型,即为高值集聚性或低值集聚性,从而可以探求具有实际意义的发病“热点区域”。

(1)全局自相关:依此可以分析整个研究范围内某种属性(本研究指 HIV 感染者人数)是否存在空间自相关性,默认整个空间为同质^[7]。但缺点是,即使全局存在自相关,也无法确切指出聚集区域,且可能掩盖其间部分区域的独特自相关性。

全局 Moran's I 计算公式:

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \times \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1)$$

式中, n 是参与分析的空间单元数, x_i 和 x_j 分别表示某现象 x 在空间单元 i 和 j 上的观测值, $w_{ij}(d)$ 为在 d 距离内的空间相邻权重矩阵。对于区域数据而言,若 i 和 j 相邻,该 $w_{ij}(d)$ 为 1,不相邻为 0。为了便于解释结果,采用随机条件下近似正态分布假设的标准差对指标进行标准化,通常将统计量结果转化为标准化 Z 值,即用该统计量减去其理论期望值再除以相应的标准差而得到。对 Moran's I 进行假设检验, Z 值 ≥ 1.96 或 ≤ -1.96 ,Moran's I $\neq 0$,认为具有空间自相关性,否则 Moran's I = 0,不具有空间自相关性。

全局 Getis 系数,又称 G 统计量,其统计公式:

$$G_i(d) = \frac{\sum_{j=1, j \neq i}^n w_{ij}(d) x_i x_j}{\sum_{j=1, j \neq i}^n x_i x_j} \quad (2)$$

式中通常也要计算 G 统计的期望值和方差,在此计算基础上计算 Z 统计量值,与界值比较。同样对 G 进行假设检验,若 $Z \geq 1.96$ 且 $G > 0$,则拒绝零假设,说明研究区域存在高值聚集,若 $Z \leq -1.96$ 且 $G < 0$,则不拒绝零假设,说明研究区域存在低值聚集。

(2)局部空间自相关:局部空间自相关弥补了全局自相关的缺点,分析特定局部地区某属性是否存在自相关。结果可以解释和寻找存在的空间按聚集性或“焦点”,也可在全局自相关的基础上,找出异质性区域或与全局自相关性相反的区域。

Local Moran's I 指数计算公式:

$$I_i = \frac{n^2}{\sum_i \sum_j w_{ij}} \times \frac{(x_i - \bar{x}) \sum_j w_{ij} (x_j - \bar{x})}{\sum_j (x_i - \bar{x})^2} \quad (3)$$

局部热点分析 Getis-ord G_i^* 计算公式:

$$G_i(d) = \sum_{j=1}^n \sum_{j=1, j \neq i}^n w_{ij}(d) x_j / \sum_{i=1}^n \sum_{j=1, j \neq i}^n x_i \quad i \neq j \quad (4)$$

局部空间自相关同样通过 Z 检验验证疾病分布在 i 至 j 区域内的集聚性是否有统计学意义。

4. 空间回归分析 (spatial regression model): 空间回归模型或称空间模型, 是从地理的角度研究疾病发病空间分布与自变量 (环境因素, 如水、空气、土壤等及社会经济学因素等) 间的关系。它在传统回归方法中加入了随机效应项来解释可能存在的空间相关性的影响, 并进一步明确其保护效应或危险效应。

空间统计分析最常用的是空间线性回归模型, 起源于经典的线性回归模型。这一模型必须满足的条件之一: 参与回归变量的残差没有空间自相关性^[8]。所以在进行最小二乘法回归 (ordinary least squares) 时, 必须对其残差进行空间自相关分析。在 Arc GIS 中建立回归模型是一个反复的过程, 涉及寻找有效的独立变量, 解释过程中, 需要模拟/理解, 然后运行回归工具以确定哪些变量是有效的预测, 然后删除/添加变量, 直到找到最佳模式。

空间线性回归模型的一般形式为:

$$Y = \rho W_1 Y + X\beta + \varepsilon$$
$$\varepsilon = \lambda W_2 \varepsilon + \mu, \mu \sim N(0, \sigma^2 I)$$

(5)

式(5)中 Y 为 N 个区域数据组成的因变量, X 为 N 个区域的自变量, ρ 为 Y 的空间自回归系数, β 为 X 的空间自回归系数, λ 为 ε 的空间自回归系数, W₁ 和 W₂ 为空间权重矩阵, μ 为均值是 0、协方差矩阵为 σ²I 的正态分布白噪声。

结 果

由图 1 可见, 苏州市城区 (328 例)、常州市城区 (361 例)、南京市城区 (588 例) 和徐州市城区 (321 例) 成为 HIV 感染者人数最多的 4 个县 (区)。总体趋势分析, 苏南地区感染者人数普遍偏高, 尤其在东部地区 HIV 感染者高发。苏中地区的感染者人数不多, 属于低发地带。苏北地区的各县 (市) 则差别较大, 如徐州市城区感染者人数很多, 铜山县、连云港市等都有比较高的感染者数, 邳州市、东海县感染者人数较低, 而新沂市、宿迁市城区等则感染者数最低, 即各个等级的感染者人数均在苏北地区有所呈现。

图 2 为江苏省各县 (区) HIV 感染例数趋势面分析。其中 X 轴表示地理经度变化, Y 轴表示地理纬度的变化, Z 轴表示 HIV 感染者人数。XZ 平面上的曲线表示 HIV 感染者人数在南北方向上的变化, YZ

平面上的曲线表示其在东西方向上的走势。平面上各点可以表示各县 (区) 发病人数的投影。总体而言, 江苏省大部分县 (区) 都处于低感染人数的状态, 只有少数地区感染者人数较高。可以发现, YZ 平面上的曲线比较平直, 显示 HIV 感染者人数在东西方向上的变化趋势不大, 而 XZ 平面上的曲线则很直观地呈现出 HIV 感染者人数由北向南先降低再上升的趋势。

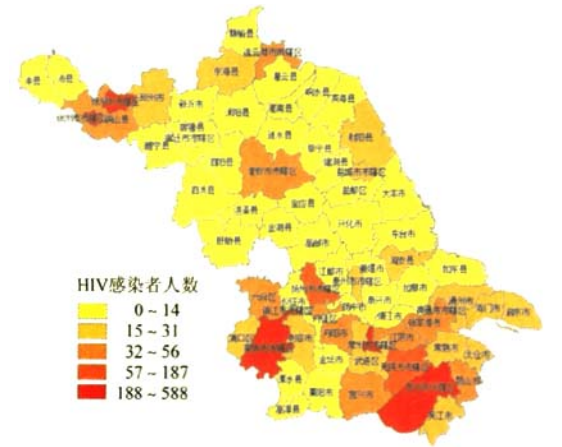


图 1 江苏省 HIV 感染例数分布

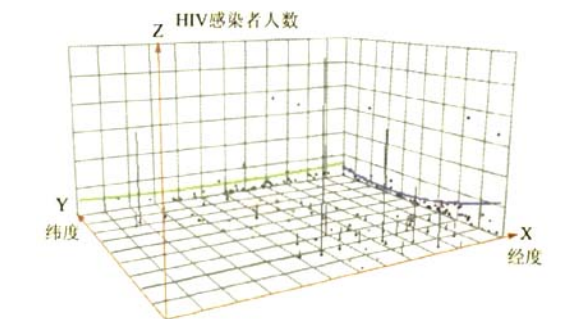


图 2 江苏省各县 (市) HIV 感染例数趋势面分析

全局自相关分析表明 (表 1), 江苏省 HIV 感染者人数在全局范围呈随机分布, 不呈现聚集性。

表 1 江苏省 HIV 感染者人数全局 Moran's I 和 Getis 分析结果

方法	观察值	期望值	方差	Z 值	P 值
Moran's I	0.030 648	-0.014 286	0.007 867	0.506 618	0.612 423
Getis	0.000 002	0.000 002	0.000 000	0.798 764	0.424 427

表 2 为 HIV 感染者人数 Local Moran's I 具有局部自相关性的城区计算结果, 差异有统计学意义 (P<0.05)。将 Local Moran's I 分析中的 Z 值计算结果分为 7 个值域, 即 Z<-2.58、-2.58≤Z<-1.96、-1.96≤Z<-1.65、-1.65≤Z<1.65、1.65≤Z<1.96、1.96≤Z<2.58 和 Z≥2.58, 其中 Z≤-1.96 表示

区域疫情呈负相关, $Z \geq 1.96$ 表示区域疫情呈正相关, 其余部分则为无统计学意义的过度地区, 表现为疫情的随机分布。对具有显著相关性的区域进行研究, 其中 H-H (高-高相邻) 表示 HIV 感染者人数高于均值的地区被高于均值的邻域包围, 表现为正相关性, 如无锡市城区和苏州市城区; H-L (高-低相邻) 表示 HIV 感染者人数高于均值的区域被低于均值的邻域包围, 表现为负相关性。在本研究得出南京市城区为负相关区域。

表2 江苏省HIV感染者人数局部Moran's I分析结果

地名	I值	Z值	P值
南京市城区(H-L)	-0.000 124	-2.799 570	0.005 117
无锡市城区(H-H)	0.000 234	5.165 080	0.000 000
苏州市城区(H-H)	0.000 113	3.145 610	0.001 657

局部热点分析将各地 Z 值分为 $Z < -2.58$ 、 $-2.58 \leq Z < -1.96$ 、 $-1.96 \leq Z < -1.65$ 、 $-1.65 \leq Z < 1.65$ 、 $1.65 \leq Z < 1.96$ 、 $1.96 \leq Z < 2.58$ 和 $Z \geq 2.58$ 共7个部分。图3中红色区域表示该地区 HIV 感染者人数有很高的聚集性, 为 HIV 感染者“热点区域”(表3), 如浦口区、无锡市城区、南京市城区、苏州市城区, 都是 HIV 感染者高度集聚的区域。而全省不存在低感染人数的聚集区域(冷点区), 大部分地区的感染情况呈现随机分布状态。

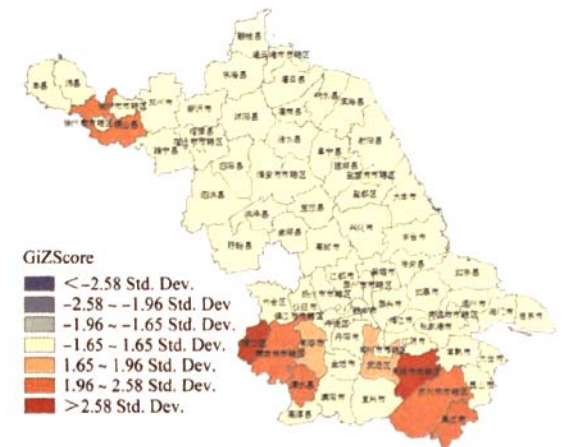


图3 江苏省各县(市)HIV感染者人数显著性分布的局部热点分析

表3 江苏省各县(市)HIV感染者人数局部热点分析结果

地区(县、市)	G值	P值
铜山县	2.106 05	0.035 20
吴江市	2.053 57	0.040 02
南京市浦口区	3.950 43	0.000 08
南京市城区	2.532 95	0.011 31
溧水县	2.023 01	0.043 07
无锡市城区	3.545 40	0.000 39
苏州市城区	2.579 77	0.009 89

以江苏省各县(区)HIV 感染者人数作为因变量, 经度和纬度为自变量, 建立空间线性回归模型。运用 Arc GIS 9.3 程序进行最小二乘法运算, 首次分析后得到各县(区)残差。对各县(区)残差进一步进行空间自相关分析, 发现南京市城区 ($Z = -2.799\ 570$)、无锡市城区 ($Z = 5.165\ 080$)、苏州市城区 ($Z = 3.145\ 610$) 的残差值存在空间自相关性。将其去除后, 重新行空间线性回归分析, 结果见表4。由此可见, 根据本研究的回归分析结果, 无法得出具有统计学意义的 HIV 感染者人数与地理经纬度之间的有效线性回归方程。

表4 修正后的空间线性回归结果

变量	r	s _e	t值	P值
经度	-8.801 184	8.421 356	-1.045 103	0.299 904
纬度	-11.860 574	8.215 582	-1.443 668	0.153 714

注: 截距=1470.005 448

讨 论

本研究分析发现, 江苏省 HIV 感染者的空间分布在东西方向上的变化不大, 在南北方向上则存在着差异。其中经济发达的苏南地区如南京市城区、无锡市城区、苏州市城区等成为全省 HIV 感染者最聚集的热点区域。HIV 感染者的聚集性提示了当地存在一定的诱发因素和地理特点, 造成艾滋病的流行。近年来, 流动人口、性乱人群、吸毒人群的比例上升, 造成 HIV 传播蔓延的趋势。结合江苏省境内艾滋病高危人群特点, 根据本研究结果, 得出造成江苏省 HIV 感染者热点区域的可能原因有①人口经济因素。苏南地区经济发达、人口流动大, 易引入大量外来感染源, 引起大规模流行。②病毒变异。自 2003 年以来, 江苏省 HIV-1 的亚型已经发生明显变化, 且更加复杂^[9]。江苏省自 2008 年以来发现 HIV 多种亚型 (CRF01_AE、CRF07_BC、CRF08_BC、B、C、CRF01_AG), 有 6 种亚型均存在于苏南地区性传播和注射吸毒传播途径中, 而该两种传播途径感染是江苏省 HIV-1 亚型日趋复杂化以及艾滋病治疗困难的重要原因。③传播方式转变, 性传播比例上升。截止 2009 年底江苏省累计报告 HIV 感染者 4103 例, 艾滋病患者 918 例。注射吸毒途径感染占的比例由 2003 年的 66.84% (262/392, 截止 2003 年底累计报告感染者 392 例) 下降至 16.40% (142/868), 性途径感染的比例由 2003 年的 21.68% (85/392) 上升至 77.40% (672/868) [其中同性性传播占 39.10% (339/868)、异性性传播占 38.30% (333/868)]。

对 HIV 感染者分布特点的研究对于 HIV 的预

防控制具有重要意义^[10-13]。本研究运用 Arc GIS 9.3 软件分析系统,按照 HIV 感染者空间分布特征,探究其在不同地理位置的不同流行特点,寻找 HIV 感染者人数与环境因素的相关和聚集程度,并寻找热点聚集区域,为进一步有重点、有方向地做好防治工作提供基础。江苏省的艾滋病监测网络自 1995 年开始建设,现已建成 80 个覆盖不同地区和各类重点人群的艾滋病性病综合监测哨点,其中国家级监测点 28 个,省级监测点 52 个,这些流行病学实时动态监测可以更新研究数据、提高研究的准确性、为防控措施的运用提供依据,引导防治方向。

空间流行病学的主要目的是建立相应的空间预测模型,来实现空间上的流行病学预测,其中,回归模型为其主要研究对象。回归模型不仅能够与风险地图结合,显示以往发病的空间规律,还能够借此预测未来发展趋势。但本研究回归模型由于数据源的限制,还不能对江苏省累计的 HIV 疫情进行空间回归分析。未来的研究应当力求结合与疾病更相关的经济发展方式、对外开放程度、文化传播特点等因素,实现多角度分析,从更加具体的角度来分析疫情的蔓延并预测其发展趋势,为更科学有效地做好疾病预防工作打好基础,对这些因素的分析将是未来重要的研究方向。

参 考 文 献

- [1] Moreira RF, Nico LS, Tomita NE. The relation between space and collective oral health: for a georeferenced epidemiology. *Cien Saude Colet*, 2007, 12(1):275-284.
- [2] Antunes JL, Frazão P, Narvai PC. Spatial analysis to identify differentials in dental needs by area-based measures. *Community Dent Oral Epidemiol*, 2002, 30(2):133-142.
- [3] Karsentg E, Leventhal A. Health Geographic Information Systems (HGIS)—a tool for health planning and epidemiology. *Harefuah*, 2002, 141(12):1070-1307.
- [4] Richards TB, Corner CM, Rushton G, et al. Geographic information systems and public health: mapping the future. *Public Health Rep*, 1999, 114(4):359-373.
- [5] Guo FY, Lu Z. The progress of disease in geography based on spatial analysis. *Geomatics World*, 2009, 7(6):22-26. (in Chinese)
郭风云, 路紫. 基于空间分析方法的疾病地理研究进展. *地理信息世界*, 2009, 7(6):22-26.
- [6] Zhou XN. *Spatial epidemiology*. Beijing: Science Press, 2009: 236-238. (in Chinese)
周晓农. *空间流行病学*. 北京: 科学出版社, 2009: 236-238.
- [7] He ZG, Han SM, Cui DY. Discussion on statistic of spatial autocorrelation analysis. *Chin J Schisto Control*, 2008, 20(4): 315-317. (in Chinese)
何宗贵, 韩世民, 崔道永. 空间自相关分析的统计量探讨. *中国血吸虫病防治杂志*, 2008, 20(4):315-317.
- [8] Wang HQ, Wang JF, Zhao YJ, et al. Space-time linear modeling method for GIS lattice data. *J China University of Petroleum*, 2008, 32(5):159-164. (in Chinese)
王海起, 王劲峰, 赵永军, 等. GIS 区域数据时空线性建模方法. *中国石油大学学报(自然科学版)*, 2008, 32(5):159-164.
- [9] Tang WM, Yan HJ, Liu XY, et al. Factors associated with HIV infection among men who have sex with men in Nanjing, Suzhou and Yangzhou: a 1:4 matched case-control study. *Chin J Epidemiol*, 2009, 30(5):448-451. (in Chinese)
唐卫明, 闫红静, 刘晓燕, 等. 江苏省南京、苏州、扬州市男男性行为人群 HIV 感染因素的配比病例对照研究. *中华流行病学杂志*, 2009, 30(5):448-451.
- [10] Peng ZH, Wang N, Wang L, et al. Development of methods for estimation and prediction on epidemic situation of HIV/AIDS. *Chin J Epidemiol*, 2009, 30(3):294-297. (in Chinese)
彭志行, 汪宁, 王璐, 等. 艾滋病疫情估计和预测方法研究进展. *中华流行病学杂志*, 2009, 30(3):294-297.
- [11] Wang J. Analysis on associated HIV risk behavior with influencing factors among injecting drug user. *Chin J Public Health*, 2007, 23(9):1025-1027. (in Chinese)
王君. 静脉吸毒人群 HIV 危险行为及影响因素分析. *中国公共卫生*, 2007, 23(9):1025-1027.
- [12] Peng ZH, Wang L, Yu RB, et al. The Asia epidemic model of AIDS and its' application to the AIDS epidemic forecast in China. *Chin J Prev*, 2010, 44(2):97-100. (in Chinese)
彭志行, 王璐, 喻荣彬, 等. 亚洲艾滋病流行模型及其在我国艾滋病疫情预测中的应用. *中华预防医学杂志*, 2010, 44(2): 97-100.
- [13] Ma L, Ye DQ, Zhang GD, et al. Study on the quality of life and social support among people living with HIV/AIDS and their family members. *Chin J Epidemiol*, 2007, 28(3): 254-257. (in Chinese)
马李, 叶冬青, 张广东, 等. 艾滋病病毒感染者/艾滋病患者及其家属生活质量和社会支持的研究. *中华流行病学杂志*, 2007, 28(3):254-257.

(收稿日期:2010-09-08)

(本文编辑:张林东)