

甲流感疫情时空聚集性的GIS分析

徐敏^{1,3}, 曹春香^{1*}, 程锦泉², 吴永胜², 谢旭², 李小文¹

(1. 中国科学院遥感应用研究所 遥感科学国家重点实验室, 北京 100101;

2. 深圳市疾病预防控制中心, 深圳 518020; 3. 中国科学院研究生院, 北京 100049)

摘要: 2009年肆虐全球的甲流感疫情是由一种突变的猪流感病毒引发的流感,故又被称为猪流感。自2009年5月11日发现首例甲流感病例传入我国以来,在短短几个月的时间里,疫情迅速蔓延并呈现全国大爆发的态势。深圳因地理位置特殊,流动人口众多,一直是我国流行病的高发地区。本文以深圳市为例,对2009年5月26日至2009年11月15日间通过传染病疫情信息网络直报系统所上报的甲型H1N1流感确诊病例,分别依据患者的性别、年龄、职业等属性进行了统计,同时以日发病数为基本单位对这期间的甲流感疫情进行了时序与特征分析;并将病例数据输入地理信息系统进行地理空间定位,选取病例患者的家庭住址作为地理空间定位的基本单元,利用回顾性时空重排扫描统计量的方法对这期间深圳市的甲流感疫情进行时空聚集性分析。结果显示,深圳市的甲流感疫情的时空聚集性重点表现在9月份上与香港接壤的南部地区,对深圳市疫情的防控要重点布置在与香港往来的几个通关口岸处。

关键词: 甲型H1N1流感;扫描统计量;时空聚集性;地理信息系统

1 引言

2009年3月18日墨西哥发现首例甲型H1N1流感(又名猪流感,以下皆简称为甲流感),在随后短短数月的时间内疫情开始在全世界范围内蔓延并呈现大爆发的趋势。我国自2009年5月11日发现首例甲流感病例以来,截止至2009年11月30日,全国已累计发现甲流感病例90720例。

对此,我们将流行病学数据,利用地理信息系统强大的制图,以及时空分析模型探测了流行病爆发的时空聚集性,以及传播模式,为流行病的早期预警提供支持^[1]。

本文对2009年5月26日至2009年11月15日间通过疫情网络直报系统所上报的深圳市甲型H1N1流感确诊病例,进行了属性特征分析,对病例数据进行地理空间定位,根据病例患者的家庭住址属性,通过google earth工具获取病例个案的经纬度信息,利用时空重排扫描统计量的方法对这期间深圳市的甲流感疫情进行时空聚集性分析。

深圳市位于中国广东省南部地区,南以香港相邻,北与东莞、惠州接壤,如图1所示。深圳行政辖区内的土地面积为1952.84 km²,据国家统计局的数据显示,2008年深圳市共有人口855万,其中流动人口675万。深圳是我国第一个经济特区,由于地理位置特殊,流动人口众多,深圳一直以来都是我国流行病的高发地区,从2004年的SARS到今年的甲流感,深圳市疫情的爆发时间及规模都位于全国的前列,因此,研究深圳市的甲流感疫情发展情况对全国疫情的研究及防控具有非常重要的指导性意义。

2 甲流感疫情基本特征的统计与分析

2.1 数据来源及属性统计

本文所使用的甲流感病例数据来源于“中国疾病预防控制中心信息系统”的子系统“疾病监测信息报告管理系统”(又称传染病疫情信息网络直报系统)。每条信息记录了患者的姓名、性别、出生日

收稿日期: 2010-06-17;修回日期: 2010-09-06.

基金项目: 艾滋病和肝炎等国家科技重大专项课题(2008ZX10004-012);国家自然科学基金项目(40871173)。

作者简介: 徐敏(1985-),男,博士研究生,江西樟树人,目前主要从事地理信息与公共卫生方面的研究工作。

E-mail: xumin@irsa.ac.cn

* 通讯作者: 曹春香(1964-),女,研究员,内蒙古呼和浩特人,主要从事森林结构参数定量遥感反演方面的研究工作。

E-mail: cao413@irsa.ac.cn

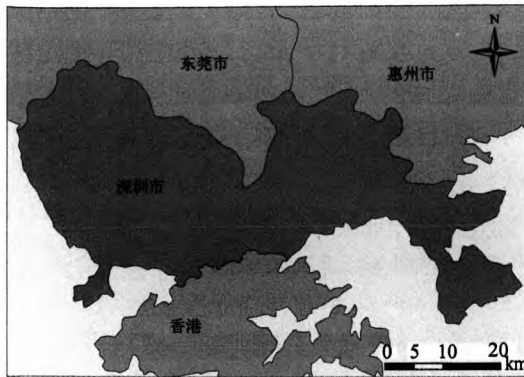


图1 深圳市地理位置

Fig. 1 The location of Shenzhen City

期、职业、家庭住址、发病日期、住址编码等。

根据统计,自2009年5月26日发现首例甲流感病例以来,截止至2009年11月15日为止,深圳市共确诊甲流感病例1119例,其中,男性病例674例,女性病例445例。疫情在性别属性上的分布没有呈现明显的聚集性,男性患者与女性患者的比例大致比较均衡,男性患者稍多于女性患者。

根据患者的职业属性,我们对从事不同职业的患者数量进行了统计,深圳市的甲流感患者在职业上的分布呈现出明显的聚集性,患者中绝大多数是学生,共有847例,占患者总数的75.69%,这表明,学生是甲流疫情的高发人群,应重点加强对学校甲流疫情的控制和防疫工作。

从年龄结构来看,深圳市甲流感患者的年龄段主要聚集在10-20岁之间,超过患者总人数的一半,这主要是因为10-20岁年龄段间的青少年主要是以学生为主,而甲流感是一种呼吸道疾病,其传播主要是通过人与人之间的飞沫,以及口腔、鼻腔、眼睛等处黏膜直接或间接接触,学生们长时间在封闭、通风条件不佳的教室内活动,其读书、下课嬉闹等都将大大增加其传播感染甲流感的概率。而50岁以上的老人大多退休在家,与外界人群接触不多,因此,感染甲流感的概率很小,这也从另一个侧面反映了甲流感的传播途径,目前还主要是以从人到人的感染为主。

2.2 病例时序分析

为了分析甲流疫情的发展趋势,本文统计了深圳市从5月26日至11月15日间的每日确诊病例,统计结果如图2所示。

从图4中可以看出,深圳市甲流感确诊病例从

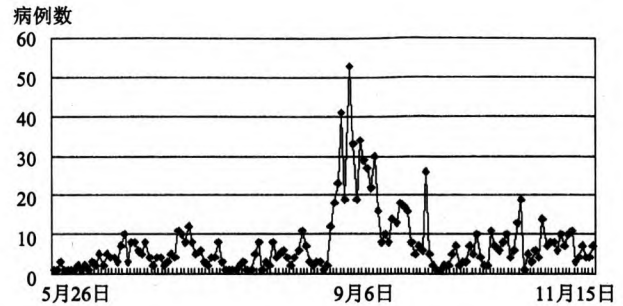


图2 深圳市甲流感病例日分布曲线

Fig. 2 Distribution curve of daily cases in Shenzhen

5月26日发现首例确诊病例后,一直处以较低的发病水平,直到9月份开始急剧增加,至9月6日达到最高值53例,随后又开始波动性的呈下降趋势。这是由于9月初是学生们开学的高峰期,学生们在学校相互感染传播,导致甲流感病例迅猛增加,随后的两个月内,由于政府开始采取积极防控措施,为在校学生接种甲流感疫苗,病例数随后慢慢开始回落。因此,随着越来越多的人进行了甲流感疫苗的接种,甲流疫情将逐渐稳定下降,并最终得到有效地控制。

3 甲流疫情时空聚集性的GIS分析

根据专家估计,我国的甲流疫情在相当长一段时间内其主要传播模式都将以聚集性感染和社区爆发流行为主^[2]。因此,探测疫情在时间和空间上的聚集性,有利于及早发现疫情的时空传播模式,可以起到疫情的早期预警作用,为政府的疫情防控工作提供支持。

国内相关专家早期主要是从时间上来探测流行病的聚集性^[3]。这种单纯的时间聚集性探测方法存在很大的局限性:若某次传染病爆发发生于城市的某个小区域,若单纯地用时间聚集性分析对每日的发病例数进行统计,则即使该区域发病例数急剧增加,也很可能由于该区域发病例数的增加不足以使整个城市的发病例数曲线上升,从而无法判别到该聚集性^[4-5]。因此,只有将传染病数据通过地理信息系统,从时间和空间上进行聚集性分析,才可能探测到该地区的聚集性。

3.1 时空重排扫描统计量方法

扫描统计量是一种广泛应用于流行病聚集性分析的方法^[6],它能够有效地探测到病例局部时间

和(或)空间发生数的增加,并且检验出这种增加是否由于随机变异造成的,它不仅可以探测某区域是否存在聚集性,而且还可以对聚集性进行精确定位^[7-9]。

时空扫描统计量是美国哈佛医学院教授 Kuldorf 于 1997 年提出的空间扫描统计量的延伸扩展^[10],它在原来空间扫描统计量的基础上加入了时间维,使扫描统计量能够同时在时间和空间上对聚集性进行探测。因此,相比较空间扫描统计量的圆形窗口,其扫描窗口也相应地变成为圆柱体,其中,圆柱体的底对应的是空间范围,而高则对应着一定长度的时间片段,由于该圆柱体扫描窗口的大小和位置都是在不断变化的,因此,时空扫描统计量能够对流行病发病的时间、地点及其规模的大小进行深入地分析,从而实现流行病爆发的早期识别^[4]。

时空扫描统计量的具体探测过程是:首先,需要随机地在研究区域内选择一个空间点作为圆柱体扫描窗口底面的中心;然后,逐渐增加扫描窗口圆柱体的底面半径及其高,圆柱体的底面积不断变化即对应扫描窗口所覆盖的地理区域变化,圆柱的高也不断变化即对应时间片段的长度的变化,直到达到预先设定的扫描窗口最大的空间和时间上限,圆柱体扫描窗口在研究区范围内的所有位置都重复相同的扫描过程;对于每一个扫描窗口,可以根据实际发病案例数和人口数来计算出预期发病数,最后,利用扫描窗口内和扫描窗口外的实际发病数和预期发病数可以构造检验统计量对数似然比(Log Likelihood Ratio, LLR),利用对数似然比扫描窗口内发病数的异常程度进行评价;由于在扫描窗口中会产生大量的扫描窗口,因此,需要选出对数似然比最大的窗口作为发病数异常程度最高的窗口,最后,利用蒙特卡罗模拟方法对该窗口的统计学意义进行评价^[4-5]。

由于本文将扫描所需要的基本地理单元精确到了患者的家庭住址,难以像通常的只精确到某行政边界那样获取到各区域确切的人口数据。因此,本文采用时空重排扫描统计量来进行聚集性分析。时空重排扫描统计量与通常所用的泊松分布的扫描统计量不同,它不需要基本单元的人口数据^[11-12],具体方法如下:

首先,假设在某扫描圆柱体对应的空间区域 z 在 d 天的时间片段内的发病数为 C_{zd} ,则整个研究区域在所有时间内的总发病例数 C 为:

$$C = \sum_z \sum_d C_{zd} \quad (1)$$

那么,每天每个区域的疫情发病数 μ_{zd} 为:

$$\mu_{zd} = \frac{(\sum_z C_{zd})(\sum_d C_{zd})}{C} \quad (2)$$

因此,就可以根据单位发病数 μ_{zd} 计算出每个扫描窗口圆柱体 A 内的预期发病数 μ_A 为:

$$\mu_A = \sum_{(z,d) \in A} \mu_{zd} \quad (3)$$

假设圆柱体 A 中的实际发病数为 C_A ,那么 C_A 服从均数为 μ_A 的超几何体分布的, C_A 的概率函数可以计算为:

$$P(C_A) = \frac{\binom{\sum_{z \in A} C_{zd}}{C_A} \binom{C - \sum_{z \in A} C_{zd}}{\sum_{d \in A} C_{zd} - C_A}}{\binom{C}{\sum_{d \in A} C_{zd}}} \quad (4)$$

当 $\sum_{d \in A} C_{zd}$ 和 $\sum_{z \in A} C_{zd}$ 相对于总发病数 C 而言非常小时, C_A 近似服从于均数为 μ_A 的泊松分布,基于这一近似,采用泊松广义似然函数 (Generalized Likelihood Ratio, GLR) 来衡量圆柱体 A 中的发病数是否异常^[4,5,12]:

$$GLR = \left\{ \frac{C_A}{\mu_A} \right\}^{C_A} \left\{ \frac{C - C_A}{C - \mu_A} \right\}^{C - C_A} \quad (5)$$

最后,利用蒙特卡罗的方法产生模拟数据集,并对模拟数据集采用同真实数据集相似的方法来进行计算,找到对数似然比最大的窗口作为发病数异常最高的窗口,即时空聚集性最大的时空区域。

3.2 数据处理与结果分析

由于本文获取的病例数据是以 excel 表格形式存储的,因此,需要对每条病例记录进行地理空间定位,给其赋予空间属性。目前,国内外对流行病数据进行地理空间定位时,大都选择行政边界作为基本的单元,然后,取行政区域的中心点作为扫描中心^[7-9]。这种编码方式具有很大的局限性,当爆发只波及该行政边界内的一个小区,而该行政区域内整体发病水平较高时,则很难探测到爆发的存在。另外,该方法探测的敏感性也很低。因此,本文选取患者的家庭住址作为编码的基本单元,利用 google earth 工具查找每一位患者家庭住址的经纬度作为该病例个案的地理空间定位。

在进行地理空间定位时,本文删除了 136 条家庭住址记录不详或者家庭住址不在深圳境内的记录,剩余 983 条记录依次编码入库。将这些数据导入 GIS 系统中,与深圳市行政边界图进行叠加显示。图 3 是编码入库的病例在空间上的分布图。

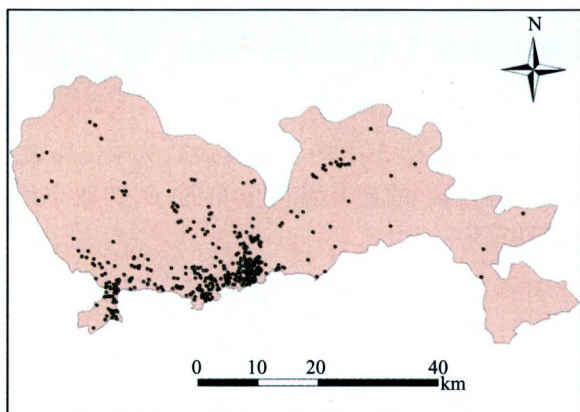


图 3 深圳市甲流感病例空间分布图

Fig. 3 The spatial distribution of H1N1 influenza A cases in Shenzhen

时空重排扫描统计量分为回顾性和前瞻性两种,由于本文的主要目的是发现甲流感疫情在深圳市各区域的时空聚集性,探索甲流感疫情的传播模式,因此,选择回顾性时空重排扫描统计量作为聚集性探测方法。甲流感的潜伏期大约为 1-7 天,故此选取最长的潜伏期 7 天作为最大时间聚集长度。空间扫描半径上限的确定原则是覆盖的人数小于或等于总人数的 50%,本文根据这一原则,通过对深圳市各街道办事处的人口统计分析,本文设定最大空间扫描半径为 10 km。时间步距采用的是一天,而 x 和 y 方向的步距由于本文是以圆形窗口进行扫描,所以,空间的步距采用扫描窗口距窗口外病例点的最小距离。得到的扫描结果如表 1 所示。同时将扫描得到疫情聚集点所在位置为中心,以该轮扫描中对数似然比最大的窗口半径作为缓冲区半径生成甲流感疫情聚集区的图层,并与疫情病例分布图和深圳市行政区域图叠加显示生成深圳市甲流感疫情时空聚集性分布图如图 4 所示。

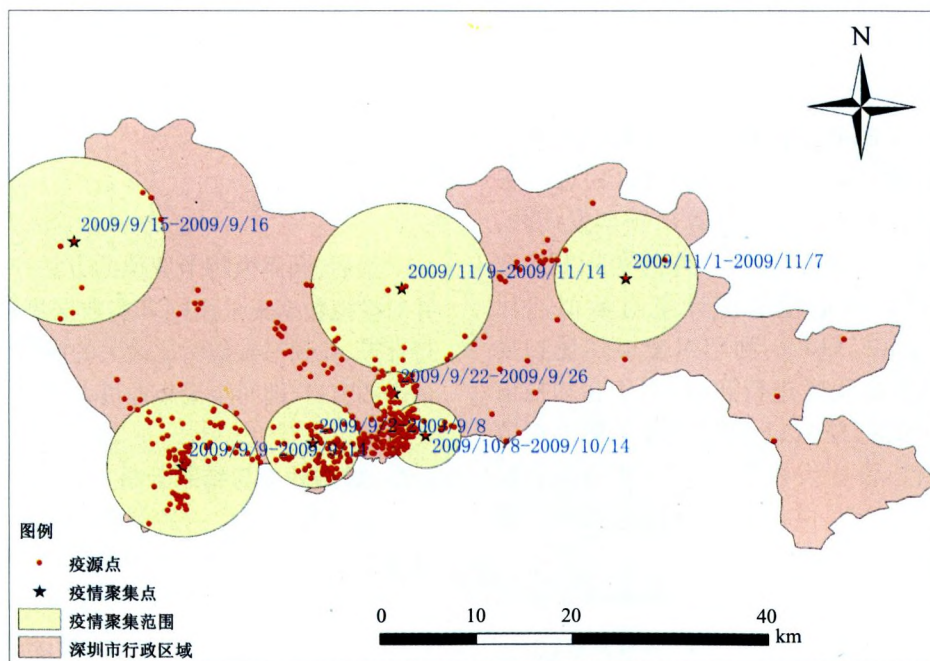


图 4 深圳市甲流感疫情聚集性时空分布图

Fig. 4 Map of H1N1 influenza A space-time clusters in Shenzhen.

从表 1 中可以看出,扫描统计量得到的前 2 个最强的聚集信号都是发生在 9 月上旬,这与前述统计的深圳市甲流感疫情日发病数统计结果是基本吻合的。也就是说深圳市甲流感疫情爆发主要是聚集在 9 月上旬学校刚开学的时候。从图 4 中

可以很直观地了解到深圳市在这期间的甲流感疫情的时空聚集情况,有些疫源点虽然在空间上表现出很强的聚集性,但是由于在时间上并没有聚集,因此,并未被检测为聚集区,这也是时空扫描统计量与单纯的空间扫描统计量的不同之处。聚

集性最强的几个区域都是在深圳市南边与香港接壤的位置,这固然是由于这些区域的人口密度大,人员被感染的可能性比较高,使得病例数在这些区域表现出较强的空间聚集性;从另一个侧面也

反映由于深圳与香港的人员往来非常频繁,使得深圳的疫情受香港疫情的影响。因此,对深圳市疫情的防控要重点布置在与香港往来的几个通关卡口岸处。

表 1 深圳市甲流疫情回顾性时空重排扫描结果

Tab.1 Cluster detection of H1N1 influenza A cases in Shenzhen using retrospective space-time permutation scan statistic.

信号	时间范围	疫情聚集中心点位置(°)	聚集半径(km)	实际发病数	预期发病数	p 值
1	2009/9/9 - 2009/9/14	22.5276 N, 113.9252 E	7.26	151	30.38	0.001
2	2009/9/2 - 2009/9/8	22.5498 N, 114.0465 E	4.66	199	53.44	0.001
3	2009/11/1 - 2009/11/7	22.7035 N, 114.3401E	6.77	54	3.42	0.001
4	2009/11/9 - 2009/11/14	22.6939 N, 114.1306 E	8.77	33	1.78	0.001
5	2009/9/22 - 2009/9/26	22.5962 N, 114.1234 E	2.21	28	2.31	0.001
6	2009/10/8 - 2009/10/14	22.5569 N, 114.1526 E	3.43	36	5.65	0.001
7	2009/9/15 - 2009/9/16	22.7375 N, 113.8243 E	8.69	12	0.29	0.001

4 结语

在流行病学领域, GIS 为与地理空间因素有关疾病的流行病学研究提供了一个新方法^[13-14]。在对流行病数据进行地理空间定位时,需要注意基本单元的选择,本文在对甲流感病例数据进行编码时,用患者的家庭住址作为基本(地理)单元,相对于通常的利用患者家庭或工作地址所在的行政区域来作为基本单元的方法而言,大大提高了空间聚集性探测的灵敏度。

另一方面,本文只是对深圳市从 5 月 26 日至 11 月 15 日的甲流疫情历史数据利用回顾性时空重排扫描统计量的方法进行了聚集性探测。而我国目前已建立的传染病疫情信息网络直报系统可以每天实时地提供各级疾控中心所上报的甲流感监测数据。因此,如果能够利用前瞻性时空重排扫描统计量逐日对实时数据进行聚集性探测,则可以实现对甲流感疫情的早期预警。

从本文聚集性探测的结果来看,深圳市甲流感疫情主要是聚集在与香港接壤的南部地区,根据调查,深圳市早期的甲流感病例也主要是从香港地区传入,因此,如果能够引入时空邻近度的概念^[14-16],利用香港与深圳之间的道路交通矢量图等计算早期病例与香港之间的时空邻近度,则有可能发现甲流感疫情的一些早期传播模式。

参考文献:

[1]林涛,姜庆五. 地理信息系统在流行病学领域的应用[J]. 海峡预防医学杂志,2002,8(4):35-36.

[2]詹发先,杨连第,张治东. 甲型 H1N1 流感的预防与控制初探[J]. 公共卫生与预防医学,2009,20(3):1-5.

[3]赛晓勇,张治英,徐德忠,等. 不同时间序列分析法洞庭湖区血吸虫病发病预测中的比较[J]. 中华流行病学杂志,2004,25(10):863-866.

[4]殷菲. 时-空扫描统计量在传染病早期预警中的应用研究[D]. 成都:四川大学博士学位论文. 2007.

[5]殷菲,冯子健,李晓松,等. 基于前瞻性时空重排扫描统计量的传染病早期预警系统[J]. 卫生研究,2007,36(4):455-458.

[6]Naus J. The Distribution of the Size of Maximum Cluster of Points on the Line [J]. J. Am. Stat. Assoc., 1965, 60: 532-538.

[7]方立群,曹春香,陈国胜,等. 地理信息系统应用于中国大陆高致病性禽流感的空间分布及环境因素分析[J]. 中华流行病学杂志,2005, 26(11):839-842.

[8]Jones R C, Liberatore M, Fernandez J R, et al. Use of a Prospective Space-time Scan Statistic to Prioritize Shigellosis Case Investigations in an Urban Jurisdiction[J]. Public Health Reports, 2006, 121:133-139.

[9]殷菲,李晓松,冯子健,等. 基于网络直报系统和时空聚集性探测的传染病模拟实时监测与预警[J]. 现代预防医学,2009,36(12):2204-2207.

[10]Kulldorff M A. Spatial Scan Statistic. Communications in Statistics[J]. Theory and Methods, 1997, 26: 1481-1496.

- [11] Kulldorff M A, Athas W F, Feuer E J, et al. Evaluating Cluster Alarms: A Space - Time Scan Statistic and Brain Cancer in Los Alamos [J], New Mexico. Public Health Briefs, 1998, 88(9): 1377 - 1380.
- [12] Kulldorff M A, Heffernan R, Hartman J, et al. A Space - Time Permutation Scan Statistic for Disease Outbreak Detection [J]. PLoS Medicine, 2005, 2(3): 216 - 224.
- [13] 曹春香, 李小文, 闫琨. 地理空间信息与 SARS 疫情走势 [J], 遥感学报, 2003, 7(4): 241 - 244.
- [14] 钟少波. GIS 和遥感应用于传染病流行病学研究——以乙肝和高致病性禽流感为例 [D]. 北京: 中国科学院研究生院博士学位论文. 2006.
- [15] 李小文, 曹春香, 常超一. 地理学第一定律与时空邻近度的提出 [J]. 自然杂志, 2007, 29(2): 69 - 72.
- [16] 杨华, 李小文, 施宏. SARS 沿交通线的“飞点”传播模型 [J]. 遥感学报, 2003, 7(4): 251 - 255.

Space-time Cluster Detection of Pandemic H1N1 Influenza A Using GIS

XU Min^{1,3}, CAO Chunxiang¹, CHENG Jinquan², WU Yongshen², XIE Xu², LI Xiaowen¹

(1. State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Applications, CAS, Beijing 100101, China;

2. Shenzhen Center for Disease Control and Prevention, Shenzhen 518020, China;

3. Graduate University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: The influenza A (H1N1) pandemic was root in the gene mutation of swine, so it is also called “swine influenza”. Since the introduction of the first case of H1N1 in China, the epidemic had been rapidly spreading over most provinces of the country in several months and would present a situation of large broken out. The paper analyzed the pandemic H1N1 influenza A in Shenzhen, China during May 26, 2009 to November 15, 2009. The data of cases rooted in the direct network reporting system for infectious diseases. First, the statistic cases were analyzed according to gender, age and vocation of the sufferers, which shows that the age of sufferers is mainly clustered in the rank of 10 - 20, more than half of the total cases, and the vocation which has the most infectors is student. Then, temporal distribution in the unit of daily cases was analyzed. The morbidity had been in a relative low level in the beginning 3 months. It suddenly increased rapidly in September and reached the peak on September 6th 2009. That's because early September is the time that students got back to school, the mutual infect among students led to the rapid increase of H1N1 influenza A cases. Finally, the data of cases was processed by GIS. The sufferers' home addresses were selected as the basic units of geo-coding and google earth tool was used to find the longitude and latitude of each sufferer. The retrospective space-time permutation scan statistic was employed to detect the space-time clusters. The longest incubation period of 7 days was chosen as the maximum temporal cluster size and the maximum spatial cluster size is defined 10 kilometers, approximately equal to the north-south distance of central Shenzhen city. The result shows that the space-time clusters of H1N1 influenza A in Shenzhen City are mainly in the northern areas which border with Hong Kong during early September. The come-and-go between Shenzhen and Hong Kong are very frequent, which caused that the epidemic in Hong Kong affect that in Shenzhen a lot. Thus, the government should enhance the control of virus in the customs between Shenzhen and Hong Kong.

Key words: H1N1 influenza A; scan statistic; space-time cluster detection; GIS