SPACE-TIME SCAN STATISTIC BASED EARLY WARNING OF H1N1 INFLUENZA A IN SHENZHEN, CHINA

C. X. Cao [a, *], M. Xu [a, c], J. Q. Chen [b], Y. S. Wu [b], X. Xie [b]

[a] State Key Laboratory of Remote Sensing Science, Jointly Sponsored by the Institute of Remote Sensing Applications of Chinese Academy of Sciences and Beijing Normal University, Beijing 100101, P.R. China - cao413@irsa.ac.cn
[b] Shenzhen Center for Disease Control and Prevention, Shenzhen 518020, P.R. China - cdc@szcdc.net
[c] Graduate School of the Chinese Academy of Sciences, Beijing 100049, P.R. China - xumin@irsa.ac.cn

Commission Ⅷ, WG Ⅷ/2 - Health

KEY WORDS: H1N1 influenza A, Scan statistic, Space-Time cluster detection, GIS

ABSTRACT:

The paper has analyzed on the confirmed cases of H1N1 influenza in Shenzhen during May 26, 2009 to November 15, 2009 which were reported through direct network reporting system for infectious diseases. Combine the records of infectious cases with GIS. Geo-code the records of cases and select the home address of the sufferer as the basic unit of geo-coding. The retrospective space-time permutation scan statistic was employed to detect the space-time clusters of the H1N1 Influenza A in Shenzhen City. The result shows that the space-time clusters of H1N1 Influenza A in Shenzhen City are mainly in the northern areas which are border with Hong Kong during early in September.

1. INTRODUCTION

Since Mexico found the first confirmed case of influenza A (H1N1) on March 18th 2009, the epidemic has been rapidly spreading over the word in several months and would be broken out in many countries. China found its first H1N1 influenza A case on May 5th 2009, and until November 30th 2009 the whole country has confirmed 90,720 influenza A cases in all.

Geographical Information System (GIS) has a strong ability in the process of spatial related data. Most of the data in epidemiology research have closely relationship with geographical spatial information (Lin & Jiang, 2002). Therefore, combining the data of epidemiology with GIS would sufficiently employ the powerful cartography and spatial analysis ability of GIS to detect the space-time clustering and transmission pattern, support for the early warning of the break out of large-scale epidemics.

This paper has analyzed the H1N1 influenza A epidemic confirmed cases of Shenzhen city in China which were reported through the direct network reporting system for infectious diseases during from May 26th 2009 to November 15th 2009, and combined the confirmed cases with GIS. Geo-coding the records of confirmed cases according to the home addresses of the sufferers, which acquired the latitude and longitude of the sufferers through google earth tool. Lastly detect the temporal and spatial clustering of H1N1 influenza A epidemic in Shenzhen using the retrospective space-time permutation scan statistic.

2. STUDY AREA

Shenzhen city is located in the southern Guangdong province, China, Adjacent to Hong Kong by the south and border with Dongguan and Huizhou by the north, as shown in figure 1.

According to the data reported by National Bureau of Statistics, Shenzhen has an area of 1952.84 square kilometers. It has a population of 8.55 millions in all, including 6.75 millions of floating population.

Shenzhen is the first of special economic zone in China. Due to its special geographical position and the numerous floating populations, Shenzhen has been China's epidemic-prone area all along. From the break out of SARS in 2004 to the H1N1 influenza A epidemic, no matter from the view of scale or duration, the epidemic in Shenzhen habitat forefront of the whole country.

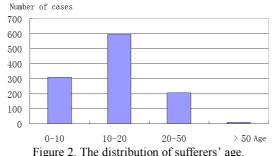Figure 1. The location of Shenzhen city.

3. DATA AND STATISTIC

The data of H1N1 influenza A cases are acquired from the direct network reporting system for infectious diseases which is the child system of China disease control and prevention information system. Each case has recorded the sufferer's

---

* Corresponding author. This is useful to know for communication with the appropriate person in cases with more than one author.

name, gender, birthday, vocation, home address, data of occurrence, address code.
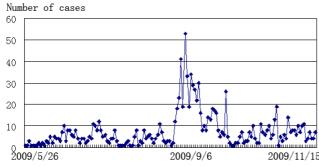
As the statistic, Shenzhen city had found 1,119 confirmed cases of H1N1 influenza A during May 26th 2009 to November 15th 2009, including 445 females and 674 males.The distribution of sufferers' sex has no obvious clustering. The ratio of male is a litter bigger than female, basically in the same level.
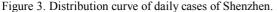
The paper has divided the sufferers' age into four ranks according to their birthdays: 0-10, 10-20, 20-50 and above 50. The statistic of sufferers' age is shown in figure 2.

Figure 2. The distribution of sufferers' age.

As the result of statistic in figure 3, the age of sufferers is mainly clustered in the rank of 10-20, more than half of the total cases. That's because most of the people that aged in the rank of 10-20 are students. The H1N1 influenza A is a respiratory disease, the virus spreads primarily droplets through a human-to-human transmission, as well as the direct or indirect contact by mouth, nose, eye and so on. The students long-playing stay in the closed, poorly ventilated classrooms. Their activities of reading, talking and playing will all increase the probability of transmission of H1N1 influenza A virus. However, the people that aged over 50 are mainly elders who have been retired at home. They have litter chance to contact with outside crowd, which led their low infection ratio of H1N1 influenza A. These also reflect that the main transmission approach of H1N1 influenza A is human-to-human infection.

In order to analyze the develop trends of the epidemic in Shenzhen, the paper has statistic daily confirmed cases of H1N1 influenza A during May 26th 2009 to November 15th 2009. The distribution curve is shown in figure 3.

Figure 3. Distribution curve of daily cases of Shenzhen.

Since the first confirmed case of H1N1 influenza A was found in Shenzhen on May 26th 2009, 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. After that it decreased in a fluctuant way. 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. In the followed two months, the government took active measures to vaccinate

H1N1 influenza A vaccine for the students, which led to the decrease of cases. Therefore, as more and more people vaccinate the vaccine, the epidemic will be gradually decline and ultimately be effectively controlled.

4. DETECTION OF SPACE-TIME CLUSTERING

As the experts' estimation, the spread of H1N1 influenza A epidemic in China will be mainly in the pattern of collective infection and community outbreak in quite a long period. Therefore, to detection the clustering of H1N1 influenza A in both spatial and temporal dimension will be very useful to discover the space-time transmission pattern and be effective to the early warning of the epidemic, which can support the epidemic prevention and control work of the government.

The domestic experts always detected clusters of epidemic only in the temporal dimension in the early period (Sai et al., 2004). This method of pure time clustering detection has a large limitation (Yin, 2007): Suppose that a particular epidemic outbreak in a small region of the city, if simply use a temporal clustering detection method to statistic daily cases of the city, even if the number of cases in the region issued in a sharp increase, it is very possible that the quantity of cases in the region is not enough to raise the curve of daily cases in whole city, which caused that the clustering can't be recognized. Therefore, only if combined with the geographic information data, can the clustering of the region be detected using a time-space statistic method.

4.1 Space-time permutation scan statistic

Scan statistic (Naus, 1965) is a widely used method of epidemic clustering detection, which can detect increases of cases in the local temporal and (or) spatial dimension and examine whether this increase is due to random variation. It can not only detect the existence of clustering in a certain region, but also precisely position the clustering (Fang et al., 2005; Jones et al., 2006; Yin et al., 2009).

Space-time scan statistic is the extension of the space scan statistic (Kulldorf 1997), which has introduced the temporal dimension based on the initial space scan statistic, thus it can simultaneity detect the clusters in temporal and spatial dimension. Accordingly, the circular scan window has changed into cylindrical, in which cylindrical bottom corresponds to a certain geographical area, and highness corresponds to a certain period. For each scan window, the expected number of cases is calculated according to the observed data and population data. Then, use the observed and expected number of cases in and outside the scan window to construct the test statistics of LLR (Log Likelihood Ratio) to evaluate the abnormal level in the scan window. And select the largest window of LLR as the most abnormal window. Lastly, evaluate the statistical significance of the window using Monte Carlo method (Kulldorf, 1998). As the size and position of the cylindrical scan window constantly changes, the space-time scan can deeply analyze the time, location and scale of the epidemic and help to early warn of the outbreaks.

For the paper has defined the patient's home address as the basic geographical unit, it is difficult to acquire the accurate population data of the basic unit as the method which only use the administrative boundary as the basic geographical unit. Thus, the paper use a space-time permutation scan to conduct the clustering analysis, which is different to the Poisson distribution based scan statistic that it doesn't need the population data of the basic unit.

Suppose C_{zd} is the observed number of cases in the region z in d days, then the total number of observed cases C is

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

For each region and day, the experted number of cases is calculated conditioning on the observed cases. The proportion of all cases μ_{zd} that occurred in region z times the total number of cases during day d is

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

The expected number of cases μ_A in the scan cylindrical window A is the summation of these expectations within the cylinder:

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

Use the Generalized Likelihood Ratio (GLR) to measure the abnormal level of the number of cases in the scan windows A.

$$GLR = \left\{ \frac{C_A}{\mu_A} \right\}^{C_A} \left\{ \frac{C - C_A}{C - \mu_A} \right\}^{C - C_A} \qquad (4)$$

Lastly, the most likely cluster is calculated for each simulated dataset in exactly the same way as for the real data. Statistical significance is evaluated using Monte Carlo hypothesis testing. Find the most abnormal window and calculated the p value.

4.2 Data process and result analysis

For the acquired cases are formatted by Excel table, which have no spatial information, it is required to geo-code them first. Recently, most of the references selected the administrative boundary as the basic geographical unit and assumed the center of the administrative region as the scan focus (Fang et al., 2005; Kulldorf, 2005; Yin *et al*., 2009). This method of geo-code has a limitation: when the outbreak only involves a very small region in the administrative boundary, but the whole incidence of disease is in a high level, then it is difficult to detect the clusters. In addition, it has a poor sensitivity. Therefore, the paper selected the patient's home address as the basic geographical unit and used the google earth tool to find the longitude and latitude of each patient which was used to geo-code the case.

When geo-coding the cases, the paper deleted 136 records which have no detail record of home address of the home is not in Shenzhen city. The rest of 983 records are geo-coded and

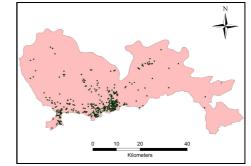put into the database. Figure 4 is the spatial distribution of the cases.

Figure 4. The spatial distribution of H1N1 influenza A cases in Shenzhen.

The space-time permutation scan statistic can divided into two types: retrospective analyses and prospective analysis. The paper chooses the retrospective space-time permutation scan statistic to detect the clusters, for our aim is to detect the spread pattern of H1N1 influenza A in Shenzhen. The incubation period of H1N1 influenza A is 1 to 7 days, so the paper chooses the longest incubation period of 7 days as the maximum temporal cluster size. The maximum spatial cluster size is defined 10 kilometers, approximately equal to the north-south distance of central Shenzhen city. The result of detection is shown in table 1. Simultaneity, the paper created the layer of epidemic clusters, which use the clustering center as the buffer focus and the clustering radius as buffer radius. The layer of epidemic clusters are used to overlaid on the layer of administrative bounder and spatial distribution of H1N1 influenza A cases. As shown in figure 5.

It can be seen from table 1 that the top two signals of the clusters are both occurred in early September, which is consistent with the statistic of daily cases analysis in section 3.4. In a word, the outbreak of H1N1 influenza A in Shenzhen is mainly in early September when the students went back to schools. From figure 6, we can get that the strongest signals of aggregation areas are all contiguous to Hong Kong, which is of course due to that their high population density. So people in these areas have a relative high ratio of being infected, caused the strong spatial clustering of H1N1 influenza A in these areas. In addition, 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 prevention and control of H1N1 influenza A in Shenzhen should focus on the customs clearances between the two regions.

Table 1. Clusters of cases detection in Shenzhen using retrospective space-time permutation scan statistic.

Signal	Data range	Location of clusters	Clustering radius（km）	Actual cases	Expected cases	*P* value
1	2009/9/9 - 2009/9/14	22.527592 N, 113.925186 E	7.26	151	30.38	0.001
2	2009/9/2 - 2009/9/8	22.549756 N, 114.046522 E	4.66	199	53.44	0.001
3	2009/11/1 - 2009/11/7	22.703542 N, 114.340072 E	6.77	54	3.42	0.001
4	2009/11/9 - 2009/11/14	22.693922 N, 114.130558 E	8.77	33	1.78	0.001
5	2009/9/22 - 2009/9/26	22.596211 N, 114.123389 E	2.21	28	2.31	0.001

| 6 | 2009/10/8 - 2009/10/14 | 22.556886 N, 114.152575 E | 3.43 | 36 | 5.65 | 0.001 |
| 7 | 2009/9/15 - 2009/9/16 | 22.737461 N, 113.824278 E | 8.69 | 12 | 0.29 | 0.001 |

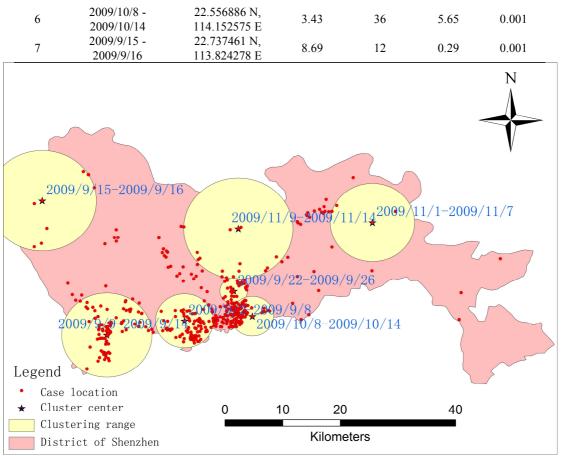Figure 5. Map of H1N1 influenza A space-time clusters in Shenzhen.

5. CONCLUSION

With the powerful capability spatial data analysis, GIS has provided a new approach for the studies of geo-spatial related epidemic in the field of epidemiology (Cao *et al.*, 2003). We should pay attention to the selecting of basic geographical unit when geo-coding the epidemiological data. The paper uses home addresses of the patients' as the basic geographic unit, which greatly improves the sensitivity of detecting spatial clustering compared to the simple use of administrative region as the basic unit.

The paper only employed approach of retrospective space-time permutation scan statistics to detect the clustering of H1N1 influenza A in Shenzhen during May 26th 2009 to November 15th 2009. However, China has already established the direct network reporting system for infectious diseases which can provide real-time surveillance data of H1N1 influenza A. So we can realize the early warning of the epidemic if using a prospective space-time permutation scan statistics to detect the data day-to-day.

It shows in the detecting result that the northern Shenzhen areas which border with Hong Kong are the main clustering region of H1N1 influenza A. The survey that the early cases are all introduced from Hong Kong has support it. Therefore, we may discover the early transmission mode of H1N1 influenza A in Shenzhen if we calculate the spatial and temporal proximity (Li *et al.*, 2007) between Shenzhen and Hong Kong using the vector maps of traffic and so on.

ACKNOWLEDGEMENTS

REFERENCES

Cao, C. X., Li, X. W., Yan, X., et al. 2003. Geo-spatial Information and Analysis of SARS Spread Trend. *Journal of Remote Sensing*. **7**(4), pp. 241-244.

Fang, L. Q., Cao, C. X., Chen, G. S., et al. Studies on the spatial distribution and environmental factors of highly pathogenic avian influenza in Mainland China, using geographic information system technology. *Chinese Journal of Epidemiology*, **26**(11), pp. 839-842.

Jones, R. C., Liberatore, M., Fernandez, J. R., et al. 2006. Use of a prospective space-time scan statistic to prioritize shigellosis case investigations in an urban jurisdiction. *Public Health Reports*, **121**, pp. 133-139.

Kulldorff, M. A., 1997. Spatial Scan Statistic. Communications in Statistics. *Theory and Methods*, **26**, pp. 1481-1496.

Kulldorff, M. A., Athas, W. F., Feuer, E. J., et al. 1998. Evaluating Cluster Alarms: A Space-Time Scan Statistic and Brain Cancer in Los Alams, New Mexico. *Public Health Briefs*, **88**(9), pp. 1377-1380.

Kulldorff, M. A., Heffernan, R., Hartman. J., et al. 2005. A Space–Time Permutation Scan Statistic for Disease Outbreak Detection. *PLoS Medicine*, **2**(3), pp. 216-224.

Li, X. W., Cao, C. X., Chang, C. Y., 2007. The First Law of Geography and Spatial-Temporal Proximity. *Chinese Journal of Nature*, **29**(2), pp. 69-72.

Lin, T., Jiang, Q. W., 2002. Application of Geographical Information System to Epidemiological Study. *Strait Journal of Preventive Medicine*. **8**(4), pp. 35-36.

Naus, J., 1965. The distribution of the size of maximum cluster of points on the line. *J Am Stat Assoc*, **60**, pp. 532-538.

Sai, X. Y., Zhang, Z. Y., Xu, D. Z., et al. 2004. Application of "time series analysis" in the prediction of schistosomiasis prevalence in areas of "breaking dikes or opening sluice for waterstore" in Dongting Lake areas, China. *Chinese Journal of Epidemiology*, **25**(10), pp. 863-866.

Yin, F., 2007. A Study on the Applicability of Space-time Scan Statistic in Early Warning of Infectious Disease. Doctor Dissertation, Chendu: SiChuan University.

Yin, F., Li, X. S., Feng, Z. J., et al. 2009. Real-time surveillance and early warning of infectious disease based on the national notifiable infectious disease reporting system and space-time detection. *Modern Preventive Medicine*. **36**(12), pp. 2204-2207.