



Contents lists available at ScienceDirect

Journal of Acute Disease

journal homepage: www.jadweb.orgOriginal article <http://dx.doi.org/10.1016/j.joad.2015.06.006>

Effects of geographic scale on population factors in acute disease diffusion analysis

Poh-Chin Lai^{1*}, Chun Bong Chow², Ho Ting Wong¹, Kim Hung Kwong¹, Shao Haei Liu³, Wah Kun Tong², Wai Keung Cheung⁴, Wing Leung Wong⁵, Yat Wah Kwan⁶¹Department of Geography, The University of Hong Kong, Pokfulam Road, Hong Kong Special Administrative Region, China²Hospital Authority Infectious Disease Centre, Princess Margaret Hospital, Hong Kong Special Administrative Region, China³Hospital Authority, Kowloon, Hong Kong Special Administrative Region, China⁴School of Nursing, The University of Hong Kong, Pokfulam Road, Hong Kong Special Administrative Region, China⁵City University of Hong Kong, Hong Kong Special Administrative Region, China⁶Department of Paediatrics and Adolescent Medicine, Princess Margaret Hospital, Hong Kong Special Administrative Region, China

ARTICLE INFO

Article history:

Received 27 Feb 2015

Received in revised form 21 May 2015

Accepted 10 Jun 2015

Available online 29 Jul 2015

Keywords:

Infectious disease

H1N1

Population factors

Spatial analysis

Geographic aggregation

ABSTRACT

Objective: To explore socio-demographic data of the population as proxies for risk factors in disease transmission modeling at different geographic scales.**Methods:** Patient records of confirmed H1N1 influenza were analyzed at three geographic aggregation levels together with population census statistics.**Results:** The study confirmed that four population factors were related in different degrees to disease incidence, but the results varied according to spatial resolution. The degree of association actually decreased when data of a higher spatial resolution were used.**Conclusions:** We concluded that variables at suitable spatial resolution may be useful in improving the predictive powers of models for disease outbreaks.

1. Introduction

Infectious diseases have long been a major cause of deaths in many countries. However, this trend shifted from infectious to chronic non-communicable diseases after the Second World War when much of the developing world was undergoing modernization or transition to become more developed^[1]. Infectious diseases have regained prominence in recent decades and have caused significant social and economic impacts to the world^[2,3]. The most notable diseases were the severe acute

respiratory syndrome (SARS) in 2003 and the H1N1 influenza in 2009^[4,5]. SARS resulted in over 8400 infected individuals from March to July 2003 in more than 30 countries around the world^[6,7]. The spread of H1N1 influenza was even more rapid. It started in Mexico in March 2009 and spread to more than 11 countries around the world within a month. Locally in Hong Kong, H1N1 appeared in early May 2009 and over 30000 laboratory confirmed cases were reported within 6 months^[8].

Newly emerged infectious diseases have potentially far more damaging impacts because of rapid transmission. Researchers have thus suggested mathematical/epidemiological and simulation models to emulate disease surveillance or even predict disease spread^[9,10]. It has been argued that the performance of such predictive models could be improved with better heuristics (experience-based techniques for problem solving) as opposed to a pure data-driven approach. Research on identifying factors influencing the spread of a disease is thus necessary and becoming more important. For example, Merler and Ajelli

*Corresponding author: Lai Poh-Chin, Professor, Department of Geography, The University of Hong Kong, Pokfulam Road, Hong Kong Special Administrative Region, China.

Tel: +852 3917 2830

Fax: +852 2559 8994

E-mail: pclai@hku.hk

Peer review under responsibility of Hainan Medical College.

Foundation Project: Supported by the Research Fund for the Control of Infectious Diseases (Project 10090122) administered by the Food and Health Bureau and the Hong Kong SAR Government.

analyzed the effects of population heterogeneity and mobility on the spread of pandemic influenza in 37 European countries by considering the European population, as well as their air and rail movements^[11]. They concluded that various epidemiological parameters including basic reproduction number, cumulative attack rate and peak daily incidence rates depended heavily on socio-demographic factors, such as household size, percentage of worker population and percentage of student population. In examining the association between different socio-demographic variables and the number of SARS cases, Kwong and Lai found that the average number of rooms per household, net residential density and percentage of elderly population were significantly correlated with the number of SARS cases^[12]. Abara *et al.* explored infectious diseases from the perspective of environmental justice and contended that infectious diseases were inextricably linked with environmental change and social determinants^[13]. They further purported a holistic approach to consider socio-environmental determinants in future public health actions.

In addition to traditional statistical approaches in estimating typical epidemiological parameters, researchers have also started applying spatial statistics and geographic information systems (GIS) to study patterns of disease clustering and dispersion^[14]. Lai *et al.* applied GIS to examine the spatial distribution of SARS patients in Hong Kong based on their residential addresses^[15]. They confirmed that the disease hot spots in urban areas did not occur randomly in space. Moreover, GIS and SaTScanTM were used by Lee and Wong to explore the initial diffusion pattern of H1N1 influenza from 1 May to 31 July 2009 in Hong Kong^[16]. Even though population density was found not significantly correlated with disease incidence, the study established that students played an important role in disseminating the disease. A follow-up study conducted by Lee and Wong, over a longer study period from 1 May to 30 September 2009, commented that the use of administrative districts in their earlier study in 2010 constrained their findings to reporting general patterns^[17]. To visualize disease diffusion, 500 m × 500 m cells instead of district boundaries were adopted to remove the border effect in the analysis of spatial clustering using measures like global and local Moran's I and the space-time permutation in SaTScanTM. Spatio-temporal clusters were found in the study, but Lee and Wong also cautioned the lack of analysis by demographic categories as one of the limitations^[17].

Generally speaking, researches have shown that spatial patterns can provide stimuli for formulating hypotheses of disease outbreaks^[18,19]. As human-to-human transmission of influenza is through close contacts, factors affecting the behavior of population including variation in demographic and environmental characteristics may affect the disease patterns^[6]. In this connection, this study aims at exploring socio-demographic data of the population as proxies for risk factors in disease transmission. If these factors could be established, they would be useful in developing predictive models for disease risk assessment.

2. Materials and methods

2.1. Research hypotheses

The main testable null hypotheses based on previous studies are as follows: 1) no relationship between young population

(under 15) and disease incidence; 2) no relationship between elderly population (65 and over) and disease incidence; 3) no relationship between cross-district/local movement (to work-place) and disease incidence and 4) no relationship between residential/population density and disease incidence^[11,12,16].

The alternative hypotheses H_a were expected a significant relationship between the test variable and disease incidence.

2.2. Data

A total of 548 patient records of confirmed H1N1 influenza from 1 May to 8 July 2009 were obtained from the Hospital Authority and the Department of Health. The records were anonymized using an arbitrary identifier and contained the following variables: age, gender, residential address, onset date of symptom and diagnostic conditions. In addition, population by-census statistics at the building group level in 2006 were downloaded from the website of the Census and Statistics Department to summarize the socio-demographic characteristics of urban environments in Hong Kong^[20]. These statistics included the following: percentage of elderly population (aged 65 and over), percentage of young population (age < 15), percentage of cross-district work population, percentage of local work population, net residential density and population density.

2.3. Method

2.3.1. Geocoding patient records

Residential addresses of the patients were geocoded using a GIS on address entries based on street or building names, which had been standardized and validated against the complete list of Hong Kong addresses compiled by the Rating and Valuation Department^[21]. Incomplete or unverifiable addresses were reported to the Hospital Authority for follow-up rectification before inclusion in the study sample. These geocoded point data were aggregated at two geographic levels: (i) 8761 grid cells of 200 m × 200 m covering only populated land areas of Hong Kong, excluding country parks as shown in Figure 1 and (ii) 18 District Council districts (Figure 2). Spatial aggregation at the grid cell level masked individual identity and enabled linkage with census data at the building group level. Spatial aggregation at the district level was for practical reason given that it was

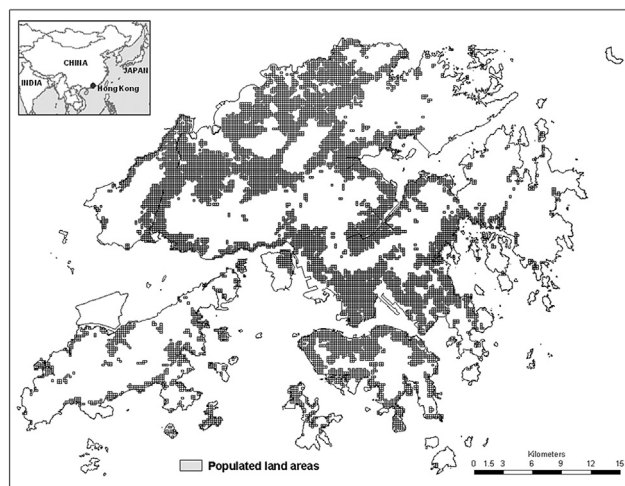


Figure 1. Populated land areas of Hong Kong (200 m × 200 m grid cells).

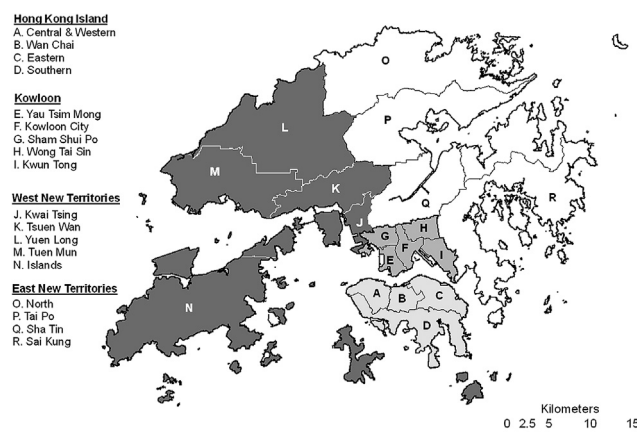


Figure 2. The 18 District Council districts of Hong Kong.

popular and the most commonly used administrative and census divisions for Hong Kong.

2.3.2. Data analysis

The data analysis involved descriptive statistics on demographic characteristics of the population and the spatio-temporal distribution of H1N1 cases. Reconstituted statistics from the 2006 population by-census at the grid cell and district levels were compared against the cumulative number of H1N1 cases at similar geographic scales and their Pearson's correlation coefficients recorded. All data analysis was conducted on ArcGIS 9.3 and SPSS 20. Statistical significance level was set at 0.05.

3. Results

Table 1 summarizes the demographic characteristics of patients of H1N1 influenza at the initial phases of the disease outbreak from 1 May to 8 July 2009.

It was apparent that most of the patients were young people below 25 (429; 78.28%) and only less than one percent of them was elderly aged 65 and above (3; 0.55%). The numbers of male and female patients were quite evenly distributed with a difference of a few percentage points (30; 5.48%).

Considering the spatial distribution of cumulative disease incidents, Kowloon was observed to incur the bulk of diseased cases (225; 41%) with the remaining occurrences equally spread between Hong Kong Island (161; 29%) and the New Territories (162; 29%) (**Table 2**). Looking at the time course of confirmed cases of H1N1 (**Table 3**), a sharp increase in the cumulative disease counts was observed in week 7 (136; 24.8%) peaking in

Table 1

Demographic characteristics of confirmed H1N1 cases.

Characteristics	Frequency	Percentage
Age groups	≤ 14	211
	15–24	218
	25–34	51
	35–44	37
	45–54	20
	55–64	8
	≥ 65	3
Gender	Male	289
	Female	259

Table 2

Spatial characteristics of confirmed H1N1 cases (1 May to 8 July 2009).

Districts	Counts	Percentage of total
Hong Kong Island	Central and Western District	24
	Wan Chai	38
	Eastern District	62
	Southern District	37
Kowloon	Yau Tsim Mong	32
	Kowloon City	44
	Sham Shui Po	29
	Wong Tai Sin	88
	Kwun Tong	32
	Islands District	10
New Territories	Kwai Tsing	18
	North District	10
	Sai Kung	23
	Sha Tin	32
	Tai Po	14
	Tsuen Wan	19
	Tuen Mun	17
	Yuen Long	19
Total	548	99

week 8 that accounted for nearly half of all cases within the study period (268; 49.0%). The outbreak seemed under control in week 9 (96; 17.5%) with a cumulative count of less than one percent of all cases (3; 0.5%) by week 10.

Table 4 presents the correlation coefficients (r) between selected population-related factors and disease incidence at three grid resolution (200 m × 200 m, 400 m × 400 m and 1000 m × 1000 m). It was noteworthy that a large proportion of the 8 761 grid cells (200 m × 200 m) did not have census data since relevant data were masked and they were too few numbers to report in those areas. **Table 4** lists the results for grid cells with census data in which four of the six population-related factors were found significantly correlated with disease incidence at different grid resolution: percentage of elderly population, percentage of cross district work population, net residential density and population density.

Table 4 shows the results only for grid cells containing confirmed H1N1 cases. Population density emerged as the only factor showing a consistent relationship with disease incidence at different spatial resolution. **Table 4** also shows improved performance in terms of the effect size of the correlation coefficients with decreased spatial resolution for all instances.

Table 3

Temporal characteristics of confirmed H1N1 cases (1 May to 8 July 2009).

Week of confirmation	Counts	Percentage of total
Week 1 (1st May–7th May)	1	0.2
Week 2 (8th May–14th May)	0	0.0
Week 3 (15th May–21st May)	0	0.0
Week 4 (22nd May–28th May)	9	1.6
Week 5 (29th May–4th June)	11	2.0
Week 6 (5th June–11th June)	24	4.4
Week 7 (12th June–18th June)	136	24.8
Week 8 (19th June–25th June)	268	49.0
Week 9 (26th June–2nd July)	96	17.5
Week 10 (3rd July–9th July)	3	0.5
Total	548	100.0

Table 4

Correlation between disease incidence (1 May to 8 July 2009) and demographic factors.

Variables	All cells with census data			Cells with H1N1 cases only		
	200 m × 200 m	400 m × 400 m	1 000 m × 1 000 m	200 m × 200 m	400 m × 400 m	1 000 m × 1 000 m
Number of cells	2 937	1 226	380	354	285	163
Elderly population (aged 65 and over) (%)	0.067**	0.155**	0.278**	0.035	0.164**	0.308**
Young population (aged <15) (%)	-0.010	-0.044	-0.121*	0.074	0.027	-0.121
Cross district work population (%)	0.060**	0.082**	0.118*	0.078	0.128*	0.129
Local work population (%)	-0.029	-0.043	-0.054	-0.001	-0.015	-0.035
Net residential density	0.062**	0.138**	0.208**	0.134*	0.015	0.139
Population density	0.245**	0.290**	0.373**	0.225**	0.233**	0.264**

*: Correlation was significant at 0.05 level.

**: Correlation was significant at 0.01 level. 200 m × 200 m, 400 m × 400 m and 1 000 m × 1 000 m are three grid resolutions.

4. Discussion

Our results of the demographic characteristics of H1N1 cases were consistent with findings of other studies reporting that disease burden of respiratory diseases was higher in the younger age groups^[22,23]. This observation suggests that schools where teenagers frequently assembled are likely an environmental risk factor for spreading H1N1 influenza. It also suggests indirectly that school closure could be an effective measure of social distancing to prevent disease spread. The absence of significant gender differences was also consistent with findings of Lee and Wong^[16].

In reference to the four testable hypotheses of this research that explores possible association between disease incidence and census variables of the population, our finding on hypothesis 4 was different from that of Lee and Wong who reported no association between population density and the occurrences of H1N1 influenza^[16]. Their study examined the association between population density and disease incidence based on roughly 400 district council constituency areas (DCCAs) from 1 May to 31 July 2009. Our study, on the contrary, showed population density to be statistically and significantly correlated with disease incidence at various grid cell resolution for all cells with census data ($r = 0.245, 0.290, 0.373$; $P < 0.01$) as well as only cells with confirmed H1N1 ($r = 0.225, 0.233, 0.264$; $P < 0.01$). Indeed, the effect size of the correlation coefficients was also the largest among the selected variables in the analysis.

We believe that the inconsistency could be a result of technical artifacts: (i) spatial resolution; (ii) exclusion criterion; and (iii) difference in the study period. Our study employing grid cells of uniform sizes appeared to be more sensitive in illuminating relationships than some 400 DCCAs. Indeed, a subsequent study by Lee and Wong^[17] concurred that spatial analysis could be affected by data resolution. They indicated that the 18 Administrative Districts of Hong Kong were less appropriate for disease diffusion study than the finer resolution of 500 m × 500 m grid cells although they did not repeat the analysis on the relationship between population density and the number of case reports. Our study demonstrated successfully that the grid cell approach was able to extract the relationship hidden by larger aggregated spatial units. Our study also showed that excluding country parks (all cells with census data in Table 4) and excluding non-diseased grid cells (cells with H1N1 cases only in Table 4) offered additional discriminating

powers to isolate salient factors in disease relationships. However, a higher spatial resolution does not necessarily mean improved associative relationships. By considering smaller spatial units of 200 m or 400 m, we introduced inevitably more data scattering and diminished health effects because of insufficient explanatory powers especially when disease cases in the earlier phases of an infection outbreak were not many. Finally, our study considered data in the early phases of the 2009 H1N1 outbreak (1 May–8 July 2009) whereas Lee and Wong^[17] extended their data analysis to the end of July 2009. Their data analysis could be confounded by more rigorous intervention and control measures at the later stages of the outbreak.

Hypotheses 2 and 3, respectively about disease incidence with the elderly population or cross-district work population exhibited significant relationships but their effect sizes were relatively small. Nevertheless, the positive relationships with disease incidence appeared logical. Elderly individuals were expected to have lower resistance to illnesses while the mobile cross-district workers were expected to have higher probabilities in contracting infectious diseases. The insignificant relationships between disease incidence and younger population (ages < 25) for hypothesis 1 were likely confounded by better hygiene and control measures at the school level. Even though the younger population made up the largest proportions of H1N1 cases as reported in Table 1, the majority of infection occurred among school-age children attending the same school that did not seem to exhibit significant spatial association.

This study examines population related risk factors at the initial phases of the H1N1 influenza outbreak in 2009. It confirms that four population factors (percentage of young population < 15; percentage of elderly population ≥ 65; percentage of cross-district work population; and population density) are related in varying degrees to disease incidence pending spatial resolution. The findings have practical implications given that these statistics are readily available from the Census and Statistics Department. The established relationships inform public health officials the target population groups (the young and the elderly) and pertinent locations (employment centers and densely populated areas) to strategize intervention measures or tighten public health practices. Moreover, the variables may be useful in improving the predictive powers of models for disease outbreaks.

The process of disaggregating map units to the finer grid cell level has proven effective in two aspects. Firstly, the gridded data are easy to manipulate in an automated setting. Secondly, the grid format seems to ameliorate the Modifiable Areal Unit

Problem^[24]. Even though grid cells may be preferred, aggregation by preset areal units (such as census subdivisions and administrative districts) has remained essential because they represent standard geographic units to collect census data or areas of jurisdiction to implement broad policies, laws and regulations. The grid format may be useful for analysis but it must be reconstituted into some preset administrative units to draw reference to area-based socio-economic measures. Our study showed that results would vary in accordance with spatial resolution. Errors and false alarms could be prevented or minimized by choosing the proper data resolution.

One limitation in relying on risk factors based on local census is the inability to account for a sudden upsurge in disease occurrences arising from external sources. Nelson indicated that the SARS outbreak in 2003 originated from a visitor as opposed to a local resident^[6]. Although the four factors could still be useful in simulating disease transmission, additional variables such as the occupancy rate of hotels within an area would be useful in estimating the potential of imported cases.

On the other hand, data loss due to extremely unclear or unavailability of locational information may reduce the reliability of research findings. In an emergency disease outbreak, it is possible that health practitioners tend to concentrate on clinical works that can be done for saving the patients, and set the task of collecting relevant information for formulating public health policies as the second priority. In many cases, records with poor locational information have to be discarded. If the number of records discarded is huge, the sample collected could become a biased sample resulting in potentially misleading observations and conclusions.

Conflict of interest statement

The authors report no conflict of interest.

Acknowledgments

We are grateful to the following government departments of the Hong Kong Special Administrative Region for data access: Census and Statistics Department, Hospital Authority, Lands Department, and Planning Department. This paper is the result of research collaboration between Princess Margaret Hospital, Hospital Authority, and Department of Geography at the University of Hong Kong. The project is funded by the Research Fund for the Control of Infectious Diseases (Project 10090122) administered by the Food and Health Bureau and the Hong Kong SAR Government.

References

- [1] Gordis L. *Epidemiology*. Philadelphia: Elsevier/Saunders; 2009.
- [2] Danziger R. The social impact of HIV/AIDS in developing countries. *Soc Sci Med* 1994; **39**(7): 905-17.
- [3] Meltzer MI, Cox NJ, Fukuda K. The economic impact of pandemic influenza in the United States: priorities for intervention. *Emerg Infect Dis* 1999; **5**(5): 659-71.
- [4] Centers for Disease Control and Prevention. Severe acute respiratory syndrome (SARS). DeKalb County: Centers for Disease Control and Prevention; 2012. [Online] Available from: <http://www.cdc.gov/sars/index.html> [Accessed on 1st August, 2012]
- [5] Centers for Disease Control and Prevention. Information on swine influenza/variant influenza viruses. DeKalb County: Centers for Disease Control and Prevention; 2012. [Online] Available from: <http://www.cdc.gov/flu/swineflu/index.html> [Accessed on 1st August, 2012]
- [6] Nelson KE. Emerging and new infectious diseases. In: Nelson KE, Williams CFM, editors. *Infectious disease epidemiology: theory and practice*. Sudbury: Jones and Bartlett Publishers; 2007, p. 407-504.
- [7] World Health Organization. Severe acute respiratory syndrome (SARS): status of the outbreak and lessons for the immediate future. Geneva: World Health Organization; 2003. [Online] Available from: http://www.who.int/csr/media/sars_wha.pdf [Accessed 1st August, 2012]
- [8] Au A. Epidemiological characteristics of the first wave of Human Swine Influenza (HSI) in Hong Kong. *Commun Dis Watch* 2009; **6**(23): 89-90.
- [9] Woolhouse M. How to make predictions about future infectious disease risks. *Philos Trans R Soc Lond B Biol Sci* 2011; **366**(1573): 2045-54.
- [10] Wu JT, Cowling BJ. The use of mathematical models to inform influenza pandemic preparedness and response. *Exp Biol Med* 2011; **236**(8): 955-61.
- [11] Merler S, Ajelli M. The role of population heterogeneity and human mobility in the spread of pandemic influenza. *Proc Biol Sci* 2010; **277**: 557-65.
- [12] Kwong KH, Lai PC. Spatial components in disease modeling. In: Taniar D, Gervasi O, Murgante B, Pardede E, Apduhan BO, editors. *Computational science and its applications – ICCSA 2010*. Berlin: Springer Berlin Heidelberg; 2010, p. 389-400.
- [13] Abara W, Wilson SM, Burwell K. Environmental justice and infectious disease: gaps, issues, and research needs. *Environ Justice* 2012; **5**(1): 9-20.
- [14] Cowling BJ, Lau MSY, Ho LM, Chuang SK, Tsang T, Liu SH, et al. The effective reproduction number of pandemic influenza in Hong Kong: prospective estimation. *Epidemiology* 2010; **21**: 842-6.
- [15] Lai PC, Wong CM, Hedley AJ, Lo SV, Leung PY, Kong J, et al. Understanding the spatial clustering of severe acute respiratory syndrome (SARS) in Hong Kong. *Environ Health Perspect* 2004; **112**(15): 1550-6.
- [16] Lee SS, Wong NS. Characterizing the initial diffusion pattern of pandemic (H1N1) 2009 using surveillance data. *PLoS Curr* 2010; **2**: RRN1151.
- [17] Lee SS, Wong NS. The clustering and transmission dynamics of pandemic influenza A (H1N1) 2009 cases in Hong Kong. *J Infect* 2011; **63**: 274-80.
- [18] So FM, Lai PC. Spatial epidemiology of asthma in Hong Kong. In: Lai PC, Mak ASH, editors. *GIS for health and the environment: development in the Asia Pacific Region*. Berlin: Springer-Verlag; 2007, p. 154-69.
- [19] McKee KT, Shields TM, Jenkins PR, Zenilman JM, Glass GE. Application of a geographic information system to the tracking and control of an outbreak of shigellosis. *Clin Infect Dis* 2000; **31**(3): 728-33.
- [20] Census and Statistics Department. 2006 Population by-census. Hong Kong: Census and Statistics Department; 2006. [Online] Available from: <http://www.byccensus2006.gov.hk/en/index.html> [Accessed on 1st August, 2012]
- [21] Rating and Valuation Department. *Names of buildings. vol. 2. The new territories*. Hong Kong: Rating and Valuation Department; 2011. [Online] Available from: http://www.rvd.gov.hk/en/doc/urban_201204.pdf and http://www.rvd.gov.hk/en/doc/nt_201204.pdf [Accessed on 1st August, 2012]
- [22] Schanzer D, Vachon J, Pelletier L. Age-specific differences in influenza a epidemic curves: do children drive the spread of influenza epidemics? *Am J Epidemiol* 2011; **174**(1): 109-17.
- [23] Silvennoinen H, Peltola V, Vainionpää R, Ruuskanen O, Heikkinen T. Incidence of influenza-related hospitalizations in different age groups of children in Finland: a 16-year study. *Pediatr Infect Dis J* 2011; **30**(2): e24-8.
- [24] Rushton G. Improving the geographical basis of health surveillance using GIS. In: Gatrell AG, Loytonen M, editors. *GIS and health*. London: Taylor & Francis; 1998.