Original 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 Chow[2], Ho Ting Wong[1], Kim Hung Kwong[1], Shao Haei Liu[3], Wah Kun Tong[2], Wai Keung Cheung[4], Wing Leung Wong[5], Yat Wah Kwan[6]

[1]*Department of Geography, The University of Hong Kong, Pokfulam Road, Hong Kong Special Administrative Region, China*

[2]*Hospital Authority Infectious Disease Centre, Princess Margaret Hospital, Hong Kong Special Administrative Region, China*

[3]*Hospital Authority, Kowloon, Hong Kong Special Administrative Region, China*

[4]*School of Nursing, The University of Hong Kong, Pokfulam Road, Hong Kong Special Administrative Region, China*

[5]*City University of Hong Kong, Hong Kong Special Administrative Region, China*

[6]*Department of Paediatrics and Adolescent Medicine, Princess Margaret Hospital, Hong Kong Special Administrative Region, China*

ARTICLE INFO

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 8 400 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 30 000 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

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 SaTScan™ 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 SaTScan™. 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 workplace) 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 complied 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) 8 761 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
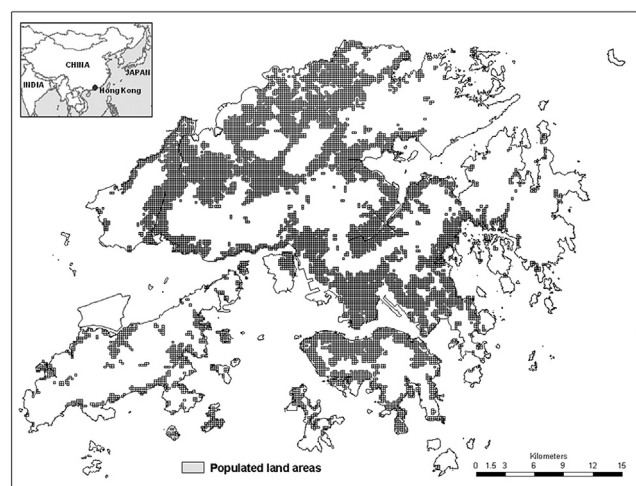

Figure 1. Populated land areas of Hong Kong (200 m × 200 m grid cells).