

Spatial distribution and risk factors of influenza in Jiangsu province, China, based on geographical information system

Jia-Cheng Zhang¹, Wen-Dong Liu², Qi Liang², Jian-Li Hu², Jessie Norris³, Ying Wu², Chang-Jun Bao², Fen-Yang Tang², Peng Huang¹, Yang Zhao¹, Rong-Bin Yu¹, Ming-Hao Zhou², Hong-Bing Shen¹, Feng Chen¹, Zhi-Hang Peng¹

¹Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, People's Republic of China; ²Jiangsu Province Center for Disease Control and Prevention, Nanjing, People's Republic of China; ³National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, People's Republic of China

Abstract. Influenza poses a constant, heavy burden on society. Recent research has focused on ecological factors associated with influenza incidence and has also studied influenza with respect to its geographic spread at different scales. This research explores the temporal and spatial parameters of influenza and identifies factors influencing its transmission. A spatial autocorrelation analysis, a spatial-temporal cluster analysis and a spatial regression analysis of influenza rates, carried out in Jiangsu province from 2004 to 2011, found that influenza rates to be spatially dependent in 2004, 2005, 2006 and 2008. South-western districts consistently revealed hotspots of high-incidence influenza. The regression analysis indicates that railways, rivers and lakes are important predictive environmental variables for influenza risk. A better understanding of the epidemic pattern and ecological factors associated with pandemic influenza should benefit public health officials with respect to prevention and controlling measures during future epidemics.

Keywords: influenza, geographical information system, spatial autocorrelation analysis, spatial-temporal cluster analysis, People's Republic of China.

Introduction

Influenza is a highly contagious, acute respiratory disease that causes periodic seasonal epidemics and global pandemics (Hampson, 1999; Yoshinari et al., 2011). Despite advances in medication, annual influenza epidemics affect 10% to 20% of the global population resulting in about 3 to 5 million cases of severe illness and about 250,000 to 500,000 deaths worldwide (Lambert and Fauci, 2010; Amit et al., 2012). Influenza exerts a heavy burden on society, and significant resources are expended to prevent and mitigate influenza epidemics every year (Paul and Ran, 2011). These epidemics show marked seasonality in many countries including Jiangsu, a highly developed coastal province in the People's Republic of China (P.R. China). Jiangsu experiences a particular pattern

of influenza that starts around December, generally peaks in January or February, and thereafter gradually subsides, typically ending in April. This repeating trend suggests the influence of associated ecological factors of potential spatio-temporal importance. The identification of these factors would be helpful for establishing measures to lessen the social burden and prevent future pandemics, such as mitigation and improving monitoring methods.

Recent researches have focused on influenza's herd susceptibility, epidemiology and aetiology. Virus types (Denoeud et al., 2007), climate (Charland et al., 2009; Tang et al., 2010), regional immunity (Targonski et al., 2007), indoor crowding (Lofgren et al., 2007), and socioeconomic factors (Hens et al., 2009) were verified to be closely connected with the emergency and maintenance of the infection. In addition, it poses a major risk for chronic patients of all ages, especially the elderly in whom it causes severe morbidity associated with a higher death rate than in younger patients (Truscott et al., 2009; Hoen et al., 2011). The disease has also been studied with respect to geographic distribution at different scales (Amit et al., 2012). At the largest scale, Viboud et al. (2004) and Chowell et al. (2008) analysed the synchrony of influenza in the United States of America, France and Australia and

Corresponding author:
Zhi-Hang Peng
Department of Epidemiology and Biostatistics
School of Public Health, Nanjing Medical University
818 East Tianyuan Road, Jiangning district
Nanjing 211166, People's Republic of China
Tel. +86 25 8686-8436; Fax +86 25 8686-8499
E-mail: zhihangpeng@njme.edu.cn

found high synchrony between the United States of America and France but none between the northern and southern hemispheres. At the country scale, in Japan, Yoshinari et al. (2011) demonstrated that the influenza incidence in neighbourhoods is related to the family structure. In this study we focused on determining the temporal and spatial distribution of the spread of influenza epidemic in Jiangsu province from 2004 to 2011 and identified the factors influencing transmission using geographical information systems (GIS) with the aim of producing useful information for prevention and control measures during future epidemics.

Materials and methods

Data collection and study area

Jiangsu province is located in the eastern part of P.R. China, comprising 106 districts with an area of 102.6 km² between longitudes 116.60° E to 121.67° E and latitudes 31.01° N to 34.89° N. According to the 2010 census, the population was approximately 78.66 million. Influenza data for the period 2004 to 2011 were obtained from the Jiangsu Center for Disease Control and Prevention, which also approved the protocol and the way the database was established for this surveillance study of influenza in the province.

Since the spread of influenza epidemics are complex and possibly multi-factorial, with all plausible determinants not necessarily causal, the determinants assessed in this study were chosen taking the potential relationship between influenza and environment into account. We investigated candidate prognostic variables possibly related to influenza from the natural and artificial ecosystems: annual mean temperature, humidity, altitude, lakes, rivers, communication (localization of railways/highways) and the normalized difference vegetation index (NDVI) for spring, summer, autumn and winter.

Spatial autocorrelation analysis

Global spatial autocorrelation of the eight years of incidence was estimated by Moran's *I* statistic. This measure was used to estimate the correlation among neighbouring observations, and to find the patterns and levels of spatial clustering among neighbouring districts. Moran's *I* value ranges from -1 to 1 and when observations closer in space have more similar attributes than areas situated far apart, a Moran's *I* will take a positive value, suggesting a clustered, spa-

tial structure. The Moran's *I* statistic is calculated by the formula:

$$I = \frac{n \sum_{i=3}^n \sum_{j=3}^n \omega_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\left(\sum_{i=3}^n \sum_{j=3}^n \omega_{ij} \right) \sum_{i=3}^n (x_i - \bar{x})^2} \quad (\text{equation 1})$$

where *n* is the total number of districts; w_{ij} the element in the spatial weight matrix corresponding to the observation pair *i, j*; x_i and x_j observations for areas *i* and *j*, and \bar{x} the mean value.

Local Getis analysis was used to test the statistical significance of local clusters and determine the spatial extent of these clusters, which is useful for identifying individual members of local clusters by determining the spatial dependence and relative magnitude between an observation and neighbouring observations (Getis et al., 2003). The local G-statistic can be written as follows:

$$G_i(d) = \frac{\sum_{i=3}^n \sum_{j=3}^n \omega_{ij}(d) \chi_j}{\sum_{i=3}^n \sum_{j=3; j \neq i}^n \chi_j} \quad (\text{equation 2})$$

The local G-statistic includes the value in the calculation at *i*. Assuming that $G_i^*(d)$ is approximately normally distributed, the output of $G_i^*(d)$ can be calculated as a standard normal variant with an associated probability from the z-score distribution. Clusters at the 95% significance level from a two-tailed normal distribution indicate a very high probability of presence of spatial clusters, but only positively significant clusters are mapped.

Spatio-temporal cluster analysis

Scan statistics was carried out with SaTScan (Kulldorff et al., 2007) to detect spatial and temporal clusters of influenza cases. This approach can be pictured as moving a cylinder, whose base is the scanning window, while the height corresponds to time over the study region (Wu et al., 2011). It is centered on the centroid of each county and whenever a case is found, the software calculates the likelihood ratio in comparison with the area outside the window to find the most likely disease cluster. In this case, the area within the circular window varied in size from 0 to 50% of the total population. Possible clusters were tested within the variable window around the centroid of each county group and the null hypothesis that the relative risk (RR) of influenza was the same between county groups, or collection of county groups, and the

remaining county groups. A P-value <0.05 was considered statistically significant.

Statistical analysis

The most important prognostic variables were selected as follows. First, all variables were individually tested for an association with the outcome status of sub-districts by univariable logistic analysis, while taking into account the spatial correlation. Variables that were statistically significant at P <0.05 were included for further analysis. Second, stepwise, multivariate logistic regression analysis was performed to identify independent risk factors for influenza, using a standard backward likelihood ratio method (Fang et al., 2012). Statistical analyses were performed with SPSS version 17.0 for Windows.

Results

Spatial autocorrelation analysis

The global autocorrelation statistics results for the annualised incidence of influenza for the time period covered by the study are summarised in Table 1. Moran's *I* was found to be significant in 2004, 2005,

Table 1. Spatial autocorrelation analysis for annual incidence of influenza in Jiangsu province, P.R. China.

Year	Moran's <i>I</i>	Z-score	P-value
2004	0.380	3.969	<0.001
2005	0.801	7.628	<0.001
2006	0.242	2.475	0.013
2007	0.095	1.055	0.292
2008	0.408	4.213	<0.001
2009	-0.082	-0.613	0.540
2010	0.001	0.154	0.878
2011	-0.159	-1.395	0.163

2006 and 2008 at the significance level of 0.05 indicating spatial heterogeneity for these years. Moran's *I* was not significant for the years 2007, 2009, 2010 and 2011.

The z-score outcomes, as calculated by equation 2, were categorised as clusters or non-clusters at the 5% significance level. The local Getis maps of the distribution of influenza cases for Jiangsu province are shown in Fig. 1 with Table 2 listing all counties showing a statistically significant autocorrelation (P <0.05). These counties were all spatially autocorre-

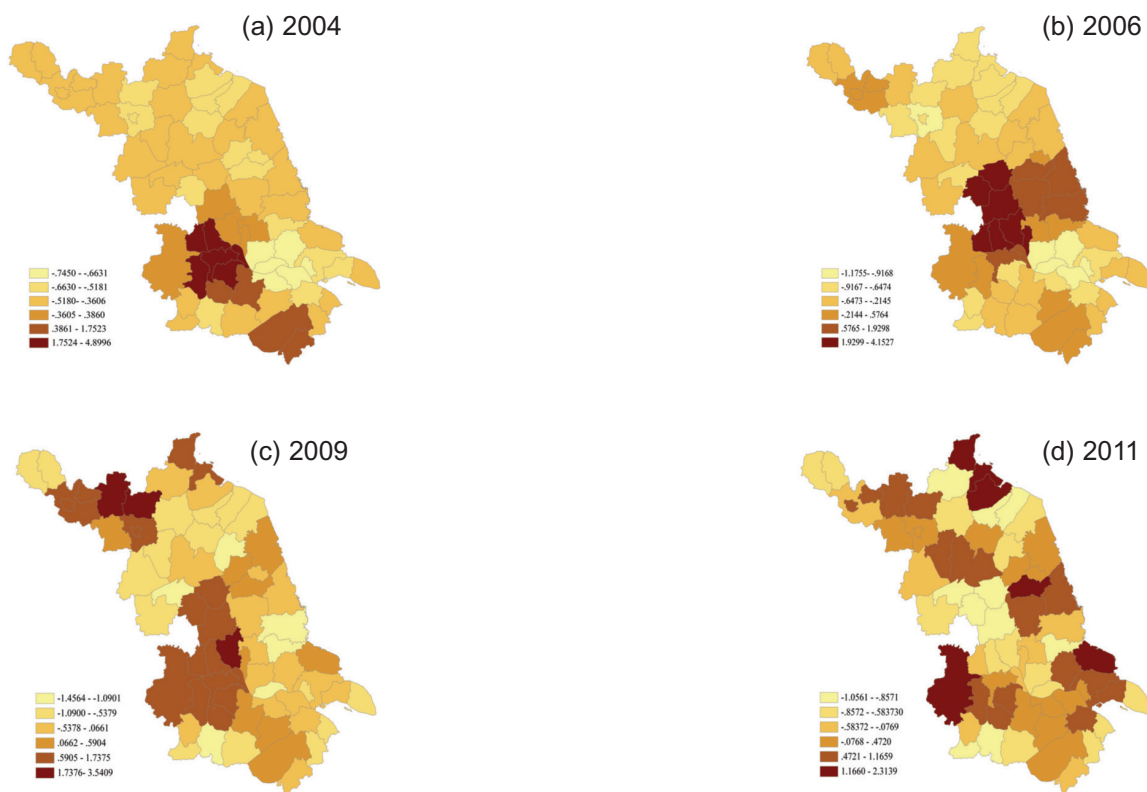


Fig. 1. Spatial clusters of annual incidence of influenza (hotspots) in Jiangsu province, P.R. China, for the years 2004 (a), 2006 (b), 2009 (c) and 2011 (d).

Table 2. Spatial clusters with significant higher incidence of influenza in Jiangsu province, P.R. China.

Year	Number	County
2004	3	Danyang, Zhenjiang, Yizheng
2005	11	Danyang, Zhenjiang, Yangzhou, Taizhou, Jiangyan, Taixing, Yizheng, Jiangdu, Jingjiang, Yangzhong, Siyang
2006	4	Yangzhou, Gaoyou, Jiangdu, Baoying
2007	1	Taicang
2008	4	Yangzhou, Taizhou, Gaoyou, Jiangdu
2009	3	Xinyi, Pizhou, Jiangdu
2010	3	Nanjing, Zhenjiang, Jurong
2011	1	Lianyungang

lated and can thus be termed hotspots for influenza in Jiangsu province. By comparing the spatial clusters of the local G statistic for the incidence of influenza from 2004 to 2011, we found that the hotspots of high incidence rates were first found in the south-western districts, such as Zhenjiang and Danyang. We also found that areas with high incidence had a tendency towards dispersed distribution. By the year 2011, all spatial clusters of influenza were highly dispersed across the whole province.

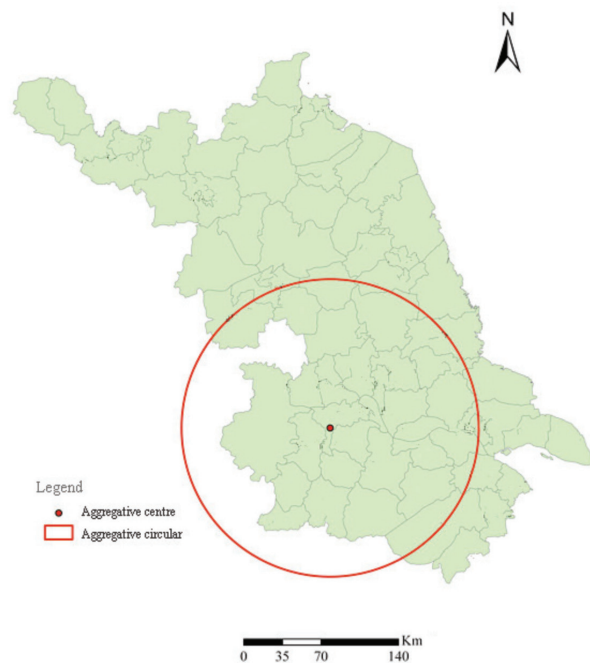


Fig. 2. Spatial-temporal distributions of influenza cases in Jiangsu province, P.R. China.

Spatio-temporal cluster analysis

By applying spatio-temporal cluster analysis for the study period in Jiangsu, it was found that influenza was non-randomly distributed, both in space and time. A most likely cluster was identified by using the maximum spatial cluster size and the maximum temporal cluster size of 50% of the total population (Fig. 2). The most likely cluster comprised 31 counties, including Zhenjiang, Danyang, Jurong, Jintan, Yangzhou, Yangzhong, Yizheng, Taizhou, Jiangdu, Changzhou, Lishui, Nanjing, Taixing, Liyang, Jingjiang, Yixing, Jiangyan, Gaochun, Gaoyou, Jiangyin, Wuxi, Jinhu, Rugao, Xinghua, Haian, Zhangjiagang, Baoying, Xuyi, Dongtai, Nantong and Suzhou in the south-western parts of Jiangsu province. It had a 134.5 km radius with the centre in Zhenjiang and the time frame was 2006-2010. This cluster had an overall RR of 5.61 and an incidence rate of 5.2/100,000. The excess risk within a non-random pattern of disease distribution was significant.

Statistical analysis

Results of the univariable logistic analyses are shown in Table 3. Four of the 11 variables yielded P-values <0.05. The univariable analyses revealed that the districts with railway, national highways, rivers and lakes were more likely to have outbreaks reported. Consequently, three risk factors were selected by multivariate logistic regression: railways, rivers and lakes (Table 4).

Discussion

Our study provides a complete overview of the spatial and temporal characteristics of influenza in Jiangsu province with the help of GIS, a widely applied approach not only capable of accurately visualising spatial distribution accurately and revealing spatial potential clusters, but also useful for exploratory analysis (Sakai et al., 2004; Aron et al., 2006; Kitron et al., 2006; Zhou et al., 2007; Ding et al., 2011). Since spatial autocorrelation measures the degree at which near and distant situations are related, spatio-temporal cluster analysis identifies aggregation time, centre and radius, this study combined these approaches to get a comprehensive understanding of the influenza distribution pattern. Moreover, by accounting for the difference in incidence rates between inside and outside the scanning window, spatio-temporal cluster lowers the possibility of pre-selection bias that

Table 3. Univariable logistic analysis of potential ecological determinants in Jiangsu province, P.R. China.

Determinant	Curve-fitting equation	R ²	F ^a	P-value
Temperature	$y = -0.017x^3 + 18.864x - 238.645$	0.009	0.289	0.750
Humidity	$y = 22.835x^2 - 123.240x + 171.087$	0.031	1.034	0.362
Height	$y = -0.025x^2 + 0.460x + 3.745$	0.040	0.878	0.457
Railway	$y = -1.55E-0.05x^3 - 0.014x + 6.908$	0.986	187.253	<0.001
National highway	$y = 5.13E-0.05x^3 - 0.002x^2 - 0.001x + 7.389$	0.996	489.609	<0.001
River	$y = 0.001x^3 - 0.015x^2 + 0.078x + 5.348$	0.883	15.030	0.003
Lake	$y = -0.008x^2 + 0.112x + 5.974$	0.940	31.233	<0.001
NDVI ^b spring	$y = 120.456x^3 - 178.566x^2 + 74.615x - 5.012$	0.050	1.102	0.355
NDVI ^b summer	$y = -335.404x^3 + 427.492x^2 - 177.132x + 25.821$	0.013	0.226	0.850
NDVI ^b autumn	$y = -128.838x^3 + 142.223x^2 - 56.711x + 9.489$	0.013	0.269	0.848
NDVI ^b winter	$y = 226.395x^3 - 212.905x^2 + 50.808x - 0.790$	0.042	0.913	0.440

^aGoodness of fit; ^bnormalized difference vegetation index.

occurs when range and centre of the study area are chosen (Wei et al., 2011). General spatial autocorrelation indicated that influenza incidence was aggregated at the provincial level, and local spatial autocorrelation analysis found that south-western districts were consistently hotspots of high incidence though they dispersed over time. In addition, the spatio-temporal cluster analysis identified a most likely cluster in the south-western region of Jiangsu province in the 2006-2010 period encompassing 31 counties. This result is consistent with the local spatial autocorrelation analysis on the whole and should assist health departments in their attempts to develop an improved prevention strategy in these districts.

The analysis illustrates that a positive, spatial correlation exists between the railways and the influenza incidence. As is known, travel has become an important issue in the prevention of infectious diseases. Travel by railways increase the spread of pathogens and renders populations close to public transportation systems vulnerable to infections. An explanation for the localization of the hotspots found may be the strong growth of the economy in the south-western parts of Jiangsu, which goes hand in hand with establishment of public transportation systems. With the same development in other regions, new influenza

hotspots can be counted on there as well including the potential of spreading all over the province. The observation of the influence of public transportation on influenza presence is in accordance with other recent findings (Fang et al., 2008; Paul and Ran, 2011). However, in multivariate studies, no correlation was observed with national highways. One explanation may be that highways may not influence influenza transmission directly but could have reduced influenza transmission due to the vicinity of railways.

The regression analysis detected that also water resources (rivers and lakes) are related to influenza incidence. Water-bodies are commonly an integral part of the transport system, but they are also a waterfowl habitat where domestic and wild waterfowl tend to concentrate that can carry avian influenza virus. Birds and other animals may transport the virus on their feathers or fur to a water source, where the virus can remain infective for several days at temperatures, after coming into contact with an infected animal or contaminated surface on a farm. The spread of the virus becomes easier when the body of water is stagnant and adjacent to a farm or community, e.g. a lake. People may unconsciously take the virus to a body of water from contaminated surfaces or infected birds. This result also supports the results of previous ecological

Table 4. Results of the multivariable analysis for ecological determinants in Jiangsu province, P.R. China.

Ecological determinant	Coefficient	SE ^a	CI ^b	P-value	
Railway	-0.196	0.245	2.043	2.735	<0.001
River	0.555	0.176	-1.654	-0.69	0.003
Lake	0.183	0.028	0.070	0.179	<0.001

^aStandard error; ^b95% confidence interval.

studies (Marius and Dirk, 2012; Syed et al., 2012).

In this study, the mean county temperatures and humidity were not significantly related to the incidence of influenza, which is inconsistent with other studies claiming that influenza viruses are adversely sensitive to cold and dry climates (Urashima et al., 2003; Radina et al., 2010).

Conclusion

The analysis of the spatial distribution and potential risk factors reveal that outbreaks of influenza spread together with railways and national highways, which places Jiangsu province at high risk for dissemination of future influenza epidemics. It is important to strengthen intervention efforts concentrating on the south-western and northern regions targeting the influenza hotspots there.

Acknowledgements

This work was supported in part by Natural Science Foundation Key Project of Jiangsu Province, China (BK2010079), "National S&T Major Project Foundation of China" (2011ZX10004-902), Jiangsu Province Health Development Project with Science and Education (NO.ZX201109), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

References

- Amit H, Oren B, Guy K, Rami Y, Uri R, Lewi, 2012. Modeling and statistic analysis of the spatio-temporal patterns of seasonal influenza in Israel. *PLoS One* 10, e45107.
- Aron JL, 2006. Barriers to use of geospatial data for adaptation to climate change and variability: case studies in public health. *Geospat Health* 1, 11-16.
- Charland KM, Buckeridge DL, Sturtevant JL, Melton F, Reis BY, Mandl KD, Brownstein JS, 2009. Effect of environmental factors on the spatio-temporal patterns of influenza spread. *Epidemiol Infect* 137, 1377-1387.
- Chowell G, Miller M, Viboud C, 2008. Seasonal influenza in the United States, France, and Australia: transmission and prospects for control. *Epidemiol Infect* 136, 852-864.
- Denoeud L, Turbelin C, Ansart S, Valleron AJ, Flahault A, Carrat F, 2007. Predicting pneumonia and influenza mortality from morbidity data. *PLoS One* 2, e464.
- Ding KQ, Bao CJ, Hu JL, 2011. Spatial analysis of tuberculosis epidemic in Jiangsu province based on geographic information system. *Chin J Public Health* 27, 1097-1099.
- Fang LQ, Sake J, Song L, Caspar W, Gong P, Xu B, Yan L, Yang H, Jan HR, Cao WC, 2008. Environmental factors contributing to the spread of H5N1 avian influenza in mainland China. *PLoS One* 3, e2268.
- Fang LQ, Wang LP, Sake J, Liang S, Tong SL, Li YL, 2012. Distribution and risk factors of 2009 pandemic influenza A (H1N1) in mainland China. *Am J Epidemiol* 175, 890-897.
- Getis A, Morrison AC, Gray K, Scott TW, 2003. Characteristics of the spatial pattern of the dengue vector, *Aedes aegypti*, in Iquitos, Peru. *Am J Trop Med Hyg* 69, 494-505.
- Hampson AW, 1999. Epidemiological data on influenza in Asian countries. *Vaccine* 17, S19-S23.
- Hens N, Goeyvaerts N, Aerts M, Shkedy Z, Van DP, Beutels P, 2009. Mining social mixing patterns for infectious disease models based on a two-day population survey in Belgium. *BMC Infect Dis* 9, 5.
- Hoen AG, Buckeridge DL, Charland KML, Mandl KD, Quach C, 2011. Effect of expanded US recommendations for seasonal influenza vaccination: comparison of two pediatric emergency departments in the United States and Canada. *CMAJ* 183, E1025-E1032.
- Kitron U, Clennon JA, Cecere MC, Gürtler RE, King CH, Vazquez-Prokopec G, 2006. Upscale or downscale: applications of fine scale remotely sensed data to Chagas disease in Argentina and schistosomiasis in Kenya. *Geospat Health* 1, 49-58.
- Kulldorff M, Mostashari F, Duczmal L, Yih WK, Kleinman K, Platt R, 2007. Multivariate scan statistics for disease surveillance. *Stat Med* 26, 1824-1833.
- Lambert LC, Fauci AS, 2010. Influenza vaccines for the future. *N Engl J Med* 363, 2036-2044.
- Lofgren E, Fefferman NH, Naumov YN, Gorski J, Naumova EN, 2007. Influenza seasonality: underlying causes and modeling theories. *J Virol* 81, 5429-5436.
- Marius G, Dirk UP, 2012. Risk factor modelling of the spatio-temporal patterns of highly pathogenic avian influenza (HPAIV) H5N1: a review. *Spat Spatiotemporal Epidemiol* 3, 173-183.
- Paul JM, Ran W, 2011. Ecological factors associated with pandemic influenza A (H1N1) hospitalization rates in California, USA: a geospatial analysis. *Geospat Health* 6, 95-105.
- Radina PS, Farida A, Rachard KK, 2010. Modeling and predicting seasonal influenza transmission in warm regions using climatological parameters. *PLoS One* 5, e9450.
- Sakai T, Suzuki H, Sasaki A, Saito R, Tanabe N, Taniguchi K, 2004. Geographic and temporal trends in influenzalike illness, Japan, 1992-1999. *Emerg Infect Dis* 10, 1822-1826.
- Syed S, Annette KE, Paritosh KB, Jens PC, Abu S, Nils T, 2012. Ecological determinants of highly pathogenic avian influenza (H5N1) outbreaks in Bangladesh. *PLoS One* 7, e33938.
- Tang JW, Lai FY, Wong F, Hon KL, 2010. Incidence of common respiratory viral infections related to climate factors in hospitalized children in Hong Kong. *Epidemiol Infect* 138, 226-235.

- Targonski PV, Jacobson RM, Poland GA, 2007. Immunosenescence: role and measurement in influenza vaccine response among the elderly. *Vaccine* 25, 3066-3069.
- Truscott J, Fraser C, Hinsley W, Cauchemez S, Donnelly C, Ghani A, Ferguson N, Meeyai A, 2009. Quantifying the transmissibility of human influenza and its seasonal variation in temperate regions. *PLoS Curr* 1, RRN1125.
- Urashima M, Shindo N, Okabe N, 2003. A seasonal model to simulate influenza oscillation in Tokyo. *Jpn J Infect Dis* 56, 43-47.
- Viboud C, Boelle PY, Pakdaman K, Carrat F, Valleron AJ, Flahault A, 2004. Influenza epidemics in the United States, France, and Australia, 1972-1997. *Emerg Infect Dis* 10, 32-39.
- Wei et al., 2011.
- Wu W, Guo J, Guan P, Sun Y, Zhou B, 2011. Clusters of spatial, temporal, and space-time distribution of hemorrhagic fever with renal syndrome in Liaoning province, northeastern China. *BMC Infect Dis* 11, 229.
- Yoshinari K, Reiko S, Yoshiki T, Yasuhiko O, Yugo S, Asami S, Taeko O, Hiroshi S, 2011. Geodemographics profiling of influenza A and B virus infections in community neighborhoods in Japan. *BMC Infect Dis* 11, 36.
- Zhou XN, Lv S, Yang GJ, Kristensen TK, Bergquist NR, Utzinger J, Malone JB, 2007. Spatial epidemiology in zoonotic parasitic diseases: insights gained at the 1st International Symposium on Geospatial Health in Lijiang, China, 2007. *Parasit Vectors* 2,10.