

Ecological factors associated with pandemic influenza A (H1N1) hospitalization rates in California, USA: a geospatial analysis

Paul J. Maliszewski, Ran Wei

School of Geographical Sciences and Urban Planning, Arizona State University, P.O. Box 875302, Tempe, Arizona, USA

Abstract. The 2009 H1N1 influenza A virus subtype (H1N1) pandemic had a large impact in the United States of America (USA), causing an estimated 192,000 to 398,000 hospitalizations and 8,720 to 18,050 deaths between April 2009 and mid-March 2010. Recent research on the 2009 H1N1 pandemic has largely focused on individual, non-spatial demographic characterizations (e.g. age and race/ethnicity) associated with H1N1 hospitalizations. Broader ecological factors such as transportation use, land use and other socioeconomic factors are important aspects of influenza studies that have not been empirically examined. This research explores and identifies ecological factors associated with 2009 H1N1 pandemic hospitalization rates. We conducted a spatial regression analysis of county level hospitalization rates from 3 April to 15 September, 2009 obtained via the California Department of Public Health. Hospitalization rates were found to be spatially dependent. Public transportation usage rates and agricultural land use proportions were significant environmental factors positively related to hospitalization rates. Consistent with public health official's assumptions and existing evidence, county percentages of persons less than 18 years of age were positively associated with hospitalization. These findings help to clarify the limited consensus and dubious evidence on the role of broader ecological factors associated with pandemic influenza. A better understanding of the ecological risk factors associated with hospitalizations should also benefit public health officials with respect to their work aiming at improving emergency supply allocation and non-pharmaceutical intervention strategies in the context of an influenza pandemic.

Keywords: disaster mitigation and response, H1N1, influenza, pandemic, USA.

Introduction

Influenza pandemics are societal burdens, induce considerable costs, and are likely to occur again in the future (Germann et al., 2006). Undoubtedly, the 1918 influenza pandemic was among the worst health-related disasters in modern history, but the 2009 H1N1 influenza A virus subtype (H1N1) pandemic had also a large impact. In the United States of America (USA), between April 2009 and March 2010, it was responsible for an estimated 192,000 to 398,000 hospitalizations and 8,720 to 18,050 deaths (CDC, 2010). The effective management of pandemic influenza continues to constitute a major concern throughout the public health community (Balicer et al., 2006; Ferguson et al., 2006). A key issue facing

the public health community includes the development of strategies for reducing the spread and effects of pandemic influenza. The effective implementation of strategies for mitigating these issues necessitates an understanding of the relationship between the spread of influenza and its debilitating effects on humans as well as the identification of factors associated with morbidity resulting from influenza infection.

By predominantly exploring individual, non-spatial demographic characterizations, recent research on the 2009 H1N1 pandemic has overlooked underlying causes of hospitalization at the ecological level (see Jain et al., 2009; Louie et al., 2009). For instance, hospitalized people in California were primarily either of Hispanic origin or under 18 years of age (California Department of Public Health, 2009; Louie et al., 2009). Despite their significance, individual level studies are largely remedial in approach and discount broader, place-based policy relevant factors that are increasingly important amid emerging infectious diseases (Loeb, 2003; Greger, 2007). Although race/ethnicity and age are important factors associated with influenza hospitalizations, their occurrence may stem from other underlying mechanisms includ-

Corresponding author:
Paul J. Maliszewski
School of Geographical Sciences and Urban Planning
Arizona State University
Coor Hall, 975 S. Myrtle Avenue, Fifth Floor
P.O. Box 875302, Tempe, Arizona, USA
Tel. +1 850 345-1713; Fax +1 480 965-8313
Email: paul.maliszewski@asu.edu

ing environmental factors such as transportation and land use or be dependent on socioeconomic factors such as education and poverty (North et al., 1993; McLafferty, 2010). Additionally, spatial factors associated with influenza hospitalizations are significant, yet often disregarded, aspects of influenza studies that could provide efficient disaster mitigation and response strategies (Crichton et al., 2007a).

We conducted a spatial regression analysis to explore county level associations between 2009 pandemic H1N1 hospitalization rates in California and various transportation, agricultural, climatic, demographic and socioeconomic variables. The identification of associated ecological factors will increase our understanding of human illness during an influenza pandemic, having direct implications for informing disaster mitigation and response strategies. To explore these factors, we generated relevant hypotheses utilizing a risk-hazards framework (Turner et al., 2003) in which the hospitalization risk is a function of the degree of a population's *exposure* to pathogen carriers and the degree of a population's *susceptibility* to serious illness. Since Hispanics and persons less than 18 years of age were previously identified as susceptible to hospitalization, we tested the following hypotheses relevant to pathogen exposure.

1) *Public transportation usage rates are positively associated with hospitalization rates.* Transportation networks increase the spread of pathogens (DeHart, 2003; Khan et al., 2009) and populations exposed to public transportation are at an elevated risk of illness due to the higher probability of contact with pathogens (Freedman and Leder, 2005). Because influenza transmission occurs through touch, droplet and inhalation of aerosol particles, influenza is also transmissible within the confines of public transport cabins and passenger station platforms (Wein and Atkinson, 2009; Lindsley et al., 2010). However, little research has explored whether public transportation increases risk to illness from pandemic influenza. A sole article by Taylor and Pocock (1972) empirically examined the relationship between public transportation use and influenza illness and found a non-significant relationship between crowded public transportation use and absence from the workplace due to sickness.

2) *Agricultural land proportions are positively associated with hospitalization rates.* Indeed, 75% of the species of infectious organisms associated with emerging diseases, including influenza A, are transmissible between animals and humans, i.e. they are zoonotic (Cleaveland et al., 2001). Initial zoonotic infections in humans result from intimate interfaces

between humans and domesticated animals (Slingenbergh et al., 2004; Pearce-Duvel, 2006). Although influenza transmission from animals to humans is rare, a novel influenza A virus can adapt to human-to-human spread through antigenic shift, potentially causing a pandemic (Nicholson et al., 2003). Once a novel influenza virus becomes readily transmissible among humans, agricultural communities are expected to be important players in pandemics depending on the dynamic social and transportation networks that affect the spread of a novel influenza virus. This may be especially true for areas containing migrant and seasonal farm-workers, who tend to be highly mobile and vulnerable to disease (Hansen and Donohoe, 2003). Therefore, it may be possible that agricultural communities associated with an agricultural community containing the initial zoonotic transmission are at a higher risk of infection.

3) *Hospitalization rates are spatially dependent.* Pathogen spillover is a common occurrence in emerging diseases (Power and Mitchell, 2004). H1N1 influenza is highly communicable and proximal county populations are likely to interact, producing unfavorable externalities. Moreover, closer geographical observations will tend to have more similar attributes than geographical observations farther apart (Tobler, 1970). Thus, the effects of H1N1 may be predisposed to spatial coalescence.

Materials and methods

Study area

The state of California was chosen as the study area as it is home to several prominent urban areas, a diverse natural and cultural climate, and one of the world's most fecund agricultural areas. Both in terms

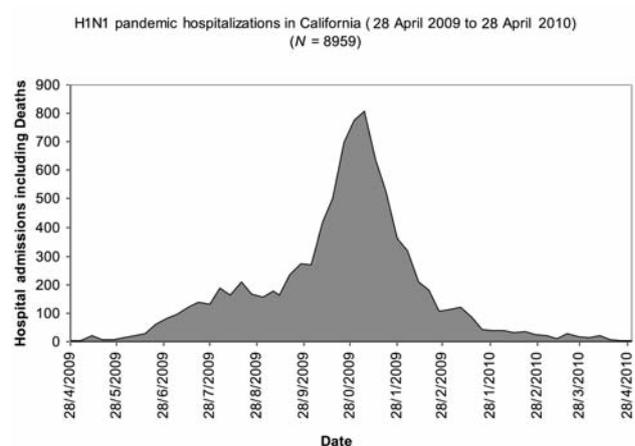


Fig. 1. H1N1 hospitalizations in California (28 April 2009 to 28 April 2010).

of official hospitalizations and confirmed deaths, California was also among those US states, most affected by the 2009 H1N1 pandemic. The first known case occurred in southern California in late March 2009 (Brown, 2009). By 6 March 2010, H1N1 had spread throughout most of the state accounting for 8,837 officially reported hospitalizations and 553 confirmed deaths (California Department of Public Health, 2009). The 2009 H1N1 pandemic in California retained the conspicuous epidemic bell shaped curve often characteristic of pandemics (Fig. 1). In terms of hospitalizations, the pandemic gained momentum throughout the summer of 2009, peaked in early November 2009, and promptly tapered off in the following months.

Unit of observation

Starting on 3 April 2009 the CDPH regularly published a database of reported hospitalizations result-

ing from H1N1 infection at the county level. Observations at the individual level are ideal for avoiding data aggregation issues; however, individual level hospitalization data were not available. Units of observation at the postal (ZIP) code level would also be problematic as many of the data we consider, such as public transportation usage rates, are only available at the county level. Furthermore, given the relatively small number of hospitalizations (2,010) within the observed time period, smaller units of analysis would increase the number of observations with zero hospitalizations raising issues of normality, islands in the spatial data and patient residence-hospital assignments. Of the 58 counties in California, 14 reported zero hospitalizations due to H1N1 within the time period studied (3 April - 15 September 2009). Most of the counties with no reports of hospitalizations were found on the sparsely populated northern and north-eastern edge of California. Because of this, we tested for departure from normality.

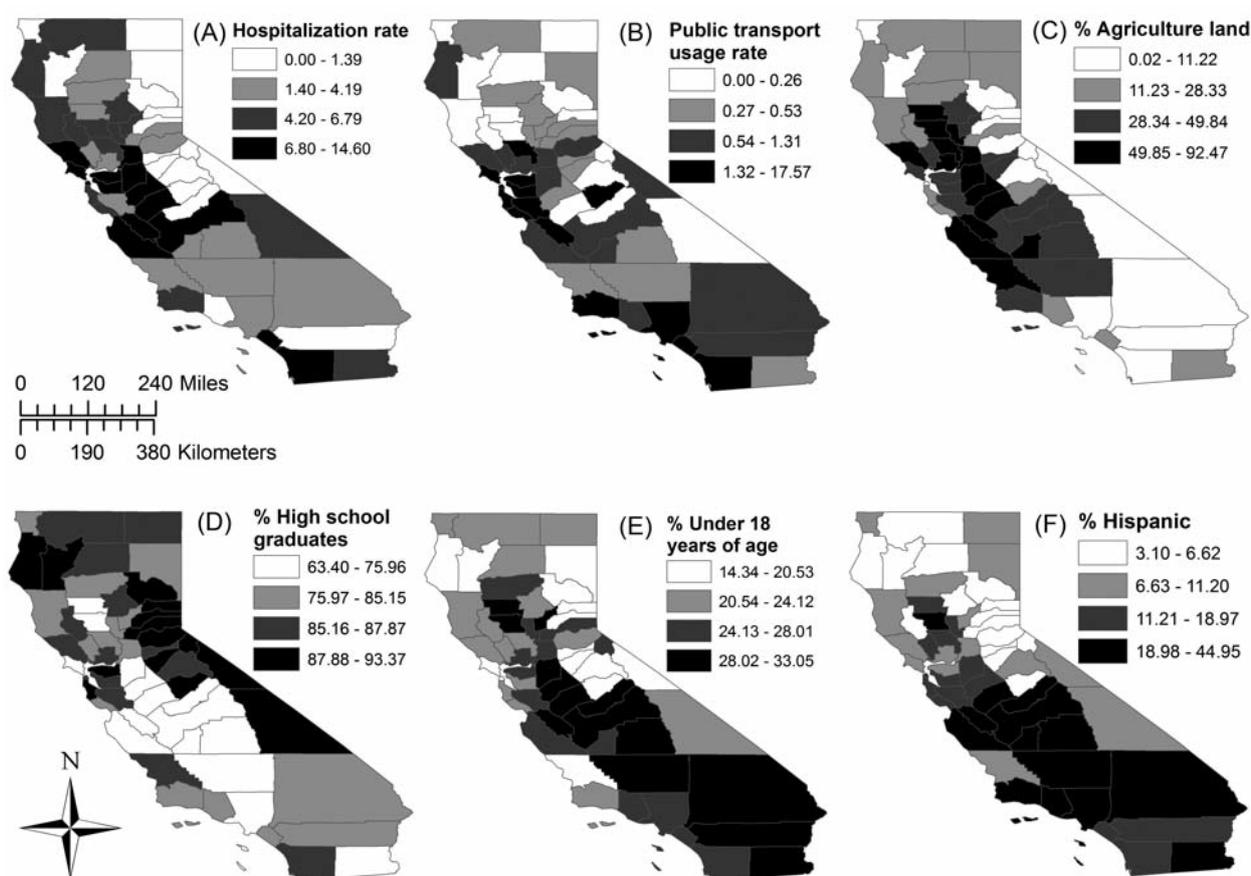


Fig. 2. Spatial distributions of select variables in observed California counties ($n = 58$). (A) Raw hospitalization rates classified in quartiles, 3 April – 15 September 2009 (source of hospitalizations: California Department of Public Health, 2009). (B) Public transportation usage rates classified in quartiles (Source: U.S. Census Bureau, 2009). (C) Agricultural land proportions classified in quartiles (source of agricultural land area data: Department of Agriculture, 2007). (D) Percentage of persons aged 25 years and above with high school diplomas (source: U.S. Census Bureau, 2009). (E) Percentage of persons under 18 years of age (Source: U.S. Census Bureau, 2009). (F) Percentage of persons Hispanic (source: U.S. Census Bureau, 2009).

Dependent variable

Proportions of hospitalizations resulting from illness due to H1N1 influenza to the at-risk population represent the dependent variable. Since the numbers of hospitalizations resulting from influenza B were few, and hospitalizations were not reported for influenza C (California Department of Public Health, 2009), we did not analyse the impact of co-circulating viruses but only the reported counts of H1N1 hospitalizations. The rates were constructed by dividing each county's number of reported H1N1 hospitalizations from 3 April to 15 September 2009 by each county's total 2009 population estimates available from the US Census Bureau (U.S. Census Bureau, 2000). In order to circumvent any effect of nationwide vaccination that began in October 2009 (CDC, 2009) hospitalizations reported after 15 September 2009 were excluded. Hospitalizations include fatal and non-fatal cases. Fig. 2A shows the spatial distribution of raw hospitalization rates across California. As is evident in Fig. 2A, counties in Central California tended to have higher hospitalization rates than counties in northern and southern California.

Independent variables

We explored environmental and socioeconomic independent variables described in Table 1. We nominally grouped the variables into transportation, agricultural, climate, age, race/ethnicity, sex and socioeconomic categories. The age, race/ethnicity, sex and socioeconomic variables were obtained via the American Community Survey (ACS) 3-year estimates (2006-2008). Select socioeconomic and demographic variables relevant to our hypotheses and known characteristics of hospitalization including population proportions of high school graduates (aged 25 years), population proportions of persons under the age of 18 years, and population proportions of Hispanics are shown in Fig. 2 (D-F).

For the transportation related variables, we analysed public transportation usage rates, population density, and county distance to the closest county with an international airport. Public transportation usage estimates were gathered from the ACS 5-year estimates (2005-2009) and defined as the percentage of population using public transportation (bus, rail or ferryboat, but excluding taxi) as their primary journey-to-work mode. Population density was constructed by dividing the US Census Bureau's 2009 county population estimates by each county's area in square miles. County distance to the nearest airport was calculated by meas-

uring Euclidean distances between each county's centroid (geometric center) to the centroid of its nearest county with an international airport (small, medium or large hub). Fig. 2B is a quartile map showing public transportation usage rates. The urban areas tend to have the highest public transportation usage rates and the San Francisco Bay Area is dominant in terms of usage rates among California's urban areas.

For agricultural variables, we tested agricultural land proportions, total number of pigs, pig density and population proportions of migrant workers for each observed county obtained via the 2007 Census of Agriculture. Agricultural land proportions were constructed by dividing agricultural land area by the total area in square miles of each county. Pig density was constructed by dividing the total number of pigs in each county by each county area in square miles. Fig. 2C shows the proportions of agricultural area in the observed counties in California. As is evident in Fig. 2C, counties with the highest proportions of agricultural land tend to be found in central California where the agrarian economy dominates.

Since many influenza studies have found climatic factors to be related to influenza transmission (see Tang et al., 2010), it would be reasonable to consider climatic effects in human influenza transmission rates. Therefore, climatic variables were included using county mean temperatures and mean dew point temperatures ($^{\circ}\text{C}$) constructed by averaging each county's weather station records between the months of April and September 2009 obtained via the Prism Climate Group.

Data analysis

Moran's I , a measure of spatial autocorrelation (Moran, 1950), was applied for exploratory spatial analysis. When neighbouring observations in space have more similar attributes than observations farther away, a Moran's I will take a high (positive) value, suggesting a clustered spatial structure. A Moran's I value will be low (negative) when neighbouring observations tend to have disparate values, indicating a regular or dispersed spatial pattern. Alternatively, if a Moran's I value is neither significantly positive nor negative, the spatial distribution of observations is not significantly different from a randomly generated spatial distribution of values. Moran's I was formulated as:

$$I = \left(\frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \right) \left(\frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \right),$$

Table 1. Variable names, descriptions, and statistics.

Variable	Description	Mean	SD	Min	Max
Dependent					
HR	Hospitalizations per 100,000 people (3 April 2009 - 15 September 2009)	4.64	3.96	0.00	14.61
Transportation					
PUBTRAN	% of population using public transport (bus, rail or ferryboat (excluding taxi)) as primary journey-to-work mode	1.22	2.46	0.00	17.57
AIRPORT	Distance to closest international airport (in miles)	66.42	60.08	0.00	233.61
DENSITY	Population per square mile	645.56	2273.22	1.58	16874.46
Agricultural					
SWINE	Number of pigs	1412.60	4574.31	0.00	29017.00
SWINEDEN	Pigs per square mile	0.66	2.75	0.00	18.80
AG	Agricultural area/total area (in square miles)	32.90	25.23	0.02	92.47
MIG	Migrant worker percentage of population	0.001	0.001	0.000	0.003
Climate					
TEMP	Mean temperature in °C (April-September 2009)	20.29	2.81	15.40	29.75
DEW	Mean dewpoint temperature in °C (April-September 2009)	6.57	2.54	0.29	11.88
Age					
<18	% of population under 18 years of age	24.38	4.53	14.34	33.05
>65	% of population over 65 years of age	12.40	3.30	7.70	26.50
Race/ethnicity					
WHITE	% of population Caucasian	82.80	9.50	56.40	92.80
ASIAN	% of population Asian	7.80	7.60	0.80	31.30
NATIVE	% of population Native American	2.10	2.00	0.50	12.00
BLACK	% of population African American	4.10	3.50	0.40	15.40
HISPANIC	% of population Hispanic	13.70	9.10	3.10	44.90
Sex					
FEMALE	% of population female	49.80	1.40	42.30	52.00
Socioeconomic					
HS	% of persons aged >25 years with a high school diploma	81.94	7.85	63.40	93.37
POV	% of population under the poverty line	14.02	4.42	6.20	22.10
MHI	Median household income (in dollars)	57055	14453	34726	88101

SD = standard deviation.

where w_{ij} is a binary spatial weight element which = 1, if the counties i and j are considered neighbours, and = 0 if not, and $i \neq j$. The total number of observations is represented by n , while x_i is the associated value at county i , and the mean value of \bar{x} .

The first spatial weights matrix was generated with neighbours defined on the basis of the first order of Queen contiguity, i.e. any two counties were considered neighbours if a common point on their boundaries is shared. The second matrix was generated with neighbours defined on the basis of k -nearest neighbours such that only the four nearest counties to county i were considered neighbours. The third matrix was generated with neighbours defined on the basis of a minimum Euclidean distance threshold such that each county had at least one neighbor and every county's centroid within a distance radius of 92.92 miles of county i 's centroid

was considered a neighbour. To test for spatial autocorrelation among observed county hospitalization rates, a special Moran's I statistic for empirical Bayes smoothed rates of hospitalizations was used (see Assunção and Reis, 1999). Since counties with small populations and low (positive) hospitalization counts can have relatively high hospitalization rates, empirical Bayes smoothing was used to deflate the weight of counties with relatively small populations to avoid misleading information due to extremely high rates from the instability of variance in hospitalization rates (Cressie, 1992; Bailey and Gatrell, 1995; Anselin et al., 2004).

Regression analysis

A standard linear regression model was constructed including independent variables we reasonably

hypothesized to be related to H1N1 hospitalization rates in California. All variables in the initial inclusion are shown in Table 1. The standard linear regression model is formally expressed in matrix notation as:

$$y = X\beta + \varepsilon,$$

where y is a vector of observations on the dependent variable, X a matrix of observations on the independent variables, β a vector of regression coefficients to be estimated, and ε a vector of error terms. The unknown regression coefficients were estimated with ordinary least squares (OLS) following the backward stepwise regression procedure for model reduction and insalient variable elimination. Given the relatively small number of observations ($n = 58$) and potential issues of multi-collinearity attributable to the large number of independent variables considered, we continued the backwards stepwise procedure until we ended up with a small number of salient variables resulting in an acceptable multi-collinearity condition number (MCN).

OLS assumes the dependent variable and the errors to be independent and identically distributed sequences of random variables, meaning each random variable's observations are independent of one another and follow the same probability distribution. Due to viral shedding, hospitalization rates are likely to be dependent across space possibly resulting in spatially clustered data. OLS estimation with a spatially clustered response variable may violate statistical assumptions of independence. To conduct a proper regression analysis of ecological level hospitalization rates, it is potentially reasonable to account for spatial autocorrelation and abate spatial interdependency issues by controlling for spatial effects (Anselin, 1988, 2003). Spatial dependence can generally take two forms. First, spatial dependence can result from underlying spatial interaction processes in the form of externalities. Second, spatial dependence can result from misspecification in the form of omitted variables, incorrect functional specification, or measurement error. Consequently, we ran spatial diagnostic tests to determine whether spatial lag dependence or spatial error dependence should be controlled for in the model before coefficient estimation. The spatial diagnostic tests were based on a Moran's I analysis of the OLS residuals and Lagrange multiplier methods detailed in Anselin et al. (1996) using a binary spatial weights matrix with neighbours defined on the basis of the first order of Queen contiguity. Depending on whether spatial dependence among county hospital-

ization rates is resultant of spatial externalities or spatial measurement error, we constructed a spatial externality (lag) model and a spatial error model to incorporate and assess the effect of spatial dependence in the form of either spatial externalities or spatial measurement error. The spatial lag model is formally expressed as:

$$y = \rho Wy + X\beta + \varepsilon.$$

The difference between the spatial lag model and the standard linear model is the incorporation of a spatially lagged dependent variable on the right hand side accounting for the average hospitalization rate of observation i 's neighbours (Anselin, 1988). As such, the spatial lag model assumes that a county's hospitalization rate is affected by its neighbouring counties' hospitalization rates via viral shedding externalities. In the case that this assumption is correct while ignoring the proper specification, coefficient estimates will be biased. The spatially lagged variable was constructed through W , a row-standardized spatial weights matrix and ρ is the spatial lag coefficient to be estimated.

Spatial dependence could also be in the form of spatial measurement error rather than underlying spatial interaction processes. Thus, we performed Lagrange multiplier tests to indicate whether spatial autocorrelation was more effectively removed with a spatial lag or spatial error model. Depending on whether the Lagrange multiplier tests indicated the proper model specification to be a spatial error model, in which spatial error dependence is a result of spatial autocorrelation in measurement errors, we also constructed a spatial error model formally expressed as:

$$y = X\beta + u,$$

where $u = \lambda Wu + \varepsilon$ is a vector of spatially autocorrelated error terms and λ is a coefficient to be estimated. The incorporation of spatial dependence through spatial weighting in the spatial lag and spatial error models are similarly constructed; however, the spatial error model includes a spatially lagged vector of error terms which is effectively treated as noise instead of a spatially lagged dependent variable treated as an explanatory variable. An issue in estimating spatial regression models' coefficients is the spatially lagged variables are endogenous and thus require special estimation techniques such as maximum likelihood or instrumental variables (Anselin, 2002). We estimated each spatial model's coefficients with the maximum

likelihood estimation method through GeoDa™ (Anselin et al., 2006). Akaike information criterion (AIC) was used to determine the model with the best fit.

Results

Descriptive statistics

Descriptive statistics of each variable are shown in Table 1. A total of 2,010 official H1N1 influenza hospitalizations were reported between 3 April and 15 September 2009 in California. The average hospitalization rate was 4.64 hospitalizations (SD = 3.96) per 100,000 people. The maximum hospitalization rate was 14.61 hospitalizations per 100,000 people observed in San Benito county. No outliers existed in the distribution of hospitalization rates.

Moran's I results

For each variable, Moran's *I* values based on three different spatial weights matrices were calculated. Permutated significance levels were based on 9,999 randomized spatial distributions of hospitalization rates. Moran's *I* was positive and significant for the empirical Bayes smoothed hospitalization rates for each spatial weights matrix: 0.312 ($P < 0.001$) for Queen-based weights, 0.336 ($P < 0.001$) for four nearest neighbours, and 0.132 ($P = 0.012$) for distance-based weights. Thus, it is highly improbable the observed spatial distribution of hospitalization rates was formed by a randomly generated spatial process. All variables showed significant positive spatial clustering except for total number of pigs and pig density whose Moran's *I* values appeared no different from a randomly generated spatial distribution. Moran's *I* coefficients based on the first order of Queen contiguity and 4-nearest neighbours tended to be higher compared with the distance-based spatial weights.

Regression results

The final model's estimated coefficients are shown in Table 2. OLS estimations for the reduced model yielded an *R*-squared of 0.432 and an AIC of -503.005. Significantly positive variable coefficients include public transportation usage rates, agricultural land proportions, and population proportions of persons under 18 years of age. The MCN was 15.243 signifying low multi-collinearity. The Jarque-Bera test for non-normality was not significant ($P = 0.181$), indi-

cating insignificant residual skewing (Jarque and Bera, 1987). The Breusch-Pagan test for heteroskedasticity was also not significant ($P = 0.454$), demonstrating insignificant non-constant variance among the residuals (Breusch and Pagan, 1979).

A Moran's *I* for the OLS residuals was calculated to identify initial evidence of spatial dependence. Moran's *I* with spatial weights defined on the basis of the first order of Queen contiguity for the OLS residuals was positive (0.154) and significant ($P = 0.029$). The Lagrange multiplier test for spatial lag dependence was significant for the final model ($P = 0.025$). The Lagrange multiplier test for spatial error dependence was not significant ($P = 0.088$). Thus, the diagnostic tests indicated the proper model specification to be the spatial lag model instead of the spatial error model (Anselin, 2005). Based on the Lagrange multiplier tests and the fact that the Moran's *I* was positive and significant for the OLS residuals, spatial dependence should be accounted for in the model prior to estimation; otherwise, the statistical assumption of independence among observations would be in violation and produce inconsistent and potentially biased estimations ensuing in contentious results. Furthermore, accounting for spatial dependence will shed light on the degree of hospitalization rate externalities.

The resulting estimations of each spatial model's coefficients with spatial weights based on Queen contiguity are shown in Table 2. For the spatial lag model, the maximum likelihood estimation yielded an AIC of -506.409, which is an improvement from the non-spatial model. In terms of the independent variables, each of the non-spatial variable's coefficient signs remained the same and significant. The coefficient for spatial lag dependence (LAG) was positive (0.369) and significant ($P = 0.006$). For the spatial error model, the maximum likelihood estimation yielded an AIC of -506.229, which was also an improvement from the non-spatial model but did not outperform the spatial lag model. Each variable's coefficient signs remained the same and significant. The spatial error coefficient (ERROR) was positive (0.342) and significant ($P = 0.032$).

Overall, in comparing the spatial models to the non-spatial model, the spatial lag model reduced the final model's parameter inconsistency and bias in estimation (Anselin, 1988) and had a smaller AIC criterion value indicating an overall improvement in the reduced model's fit to the data. A map of both the predicted hospitalization rates and predicted errors are provided in Fig. 3.

Table 2. Regression results.

Variable	Non-spatial model (n = 58)		Spatial lag model (n = 58)		Spatial error model (n = 58)	
	Coefficient	P	Coefficient	P	Coefficient	P
CONSTANT	-5.572 (2.453)	0.068	-5.239 (2.223)	0.018	-4.088 (2.468)	0.098
LAG	-	-	0.369 (0.135)	0.006	-	-
ERROR	-	-	-	-	0.342 (0.160)	0.032
PUBTRAN	0.756 (0.172)	<0.0001	0.598 (0.163)	0.0002	0.660 (0.175)	0.0001
AG	0.059 (0.018)	0.002	0.046 (0.017)	0.008	0.054 (0.020)	0.006
18	0.261 (0.106)	0.017	0.240 (0.096)	0.013	0.249 (0.106)	0.019
R ²	0.432	-	-	-	-	-
Pseudo R ²	-	-	0.500	-	0.478	-
AIC	-503.005	-	-506.409	-	-506.229	-
MCN	15.243	-	-	-	-	-
Jarque-Bera	3.414	0.181	-	-	-	-
Breusch-Pagan	2.623	0.454	1.244	0.742	1.425	0.700
Residual Moran's <i>I</i>	0.154	0.029	-	-	-	-
LM (lag)	5.021	0.025	-	-	-	-
LM (error)	2.908	0.088	-	-	-	-
Likelihood ratio	-	-	5.404	0.020	3.223	0.073

Dashes indicate values are either not applicable or not available. Standard errors are in parentheses. Abbreviations: Akaike information criterion (AIC); Lagrange multiplier (LM); multicollinearity condition number (MCN).

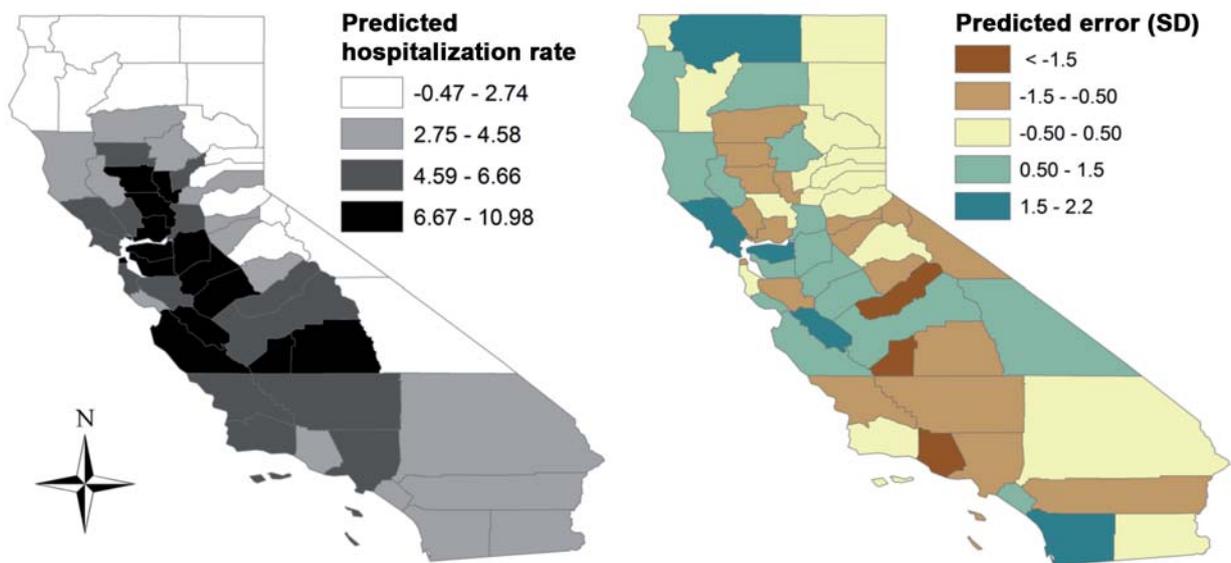


Fig. 3. Map of predicted hospitalization rates (left) and predicted errors (right), as expressed in standard deviation (SD), for the spatial lag model. Predicted hospitalization rates are classified in quartiles.

Discussion

The results support our initial hypotheses. First, public transportation usage rates were positively and significantly associated with hospitalization rates. This was expected since transportation networks increase the spread of pathogens and render populations exposed to public transportation systems vulnerable to infection. Second, agricultural land proportions were positively and significantly related to hospitalization rates. Even though there is a clear correlation between agricultural areas and hospitalization rates, it is fallacious to assume that agricultural communities are at a higher risk of exposure and infection because of higher levels of intimate contact with domesticated livestock. In all likelihood, a single human-to-non-human animal zoonotic event (or a few events at most) occurred to give rise to the 2009 pandemic. As such, the association is likely due to dynamic social or transportation networks among agricultural communities rather than agriculture itself. That is, many agricultural communities will be associated with other agricultural communities, and their interconnected networks affect the spread of disease. Counties with higher proportions of agricultural land are likely to have greater contact levels between agricultural workers and linkages with agricultural workers in Mexico. For example, California's agricultural communities may have propagated the H1N1 pandemic provided the likely trade between migrant farm workers and livestock within and outside California counties and with Mexico, the origin of the 2009 H1N1 pandemic (Greger, 2007; Aras et al., 2009). This, in combination with agricultural communities being more vulnerable to disease in general, may explain why agricultural areas tend to have higher rates of hospitalization. Third, the effect of geographical spillovers of hospitalization rates should be recognized. The spatial lag coefficient was positive and significant indicating hospitalization rates were dependent upon neighbouring county hospitalization rates. Through this finding, however, apparent contagion is impossible to distinguish from real contagion. Finally, owing to weaker immune systems, extreme-age persons are highly susceptible to becoming seriously ill from influenza infection (Viboud et al., 2004; Crichton et al., 2007b). We controlled for this assumption and showed that persons less than 18 years of age are more susceptible to becoming severely ill from novel influenza A (H1N1) at the county level.

The estimations have also yielded other results worth noting. First, each of the socioeconomic variables including population proportions of high school

graduates (aged >25 years), percentage of persons living in poverty, and median household income were not significantly related to hospitalization rates. Second, mean county temperatures and dew point temperatures were not significantly related to hospitalization rates, which is inconsistent with other studies claiming that host immune responses are adversely sensitive to cold and dry climates (Lowen et al. 2007; Tang et al., 2010). Finally, all race/ethnicity variables were not significantly related to county hospitalization rates, which is inconsistent with the majority of hospitalizations being Hispanic at the individual level.

In terms of policy relevance, public policy makers should be cautious to reduce public transportation usage rates in the event of an influenza pandemic since reductions would have adverse economic consequences (Inglesby et al., 2006; Blendon et al., 2008). Reducing the amount of agricultural land is also unreasonable for economic reasons. Ideally, public policy should target counties with higher public transportation usage rates and agricultural areas for disaster planning and response such as pharmaceutical and non-pharmaceutical strategies to prevent disease spread and ultimately lower hospitalization rates from pandemic situations. Plausible pharmaceutical strategies include the use of anti-viral drugs, and perhaps vaccines, among exposed populations. Plausible non-pharmaceutical policy implementations include requiring the use of face masks and/or hand sanitizing gel prior to boarding a crowded public transport cabin or reducing exposure to endemic animals or requiring cleaner work environments. Immunization via vaccination remains a challenge for the public health community since it is likely that new vaccines will not be available by the time a novel influenza pandemic gains momentum. When available, shocks from allocations of anti-viral drugs and vaccines to "hot-spots" are potentially effective strategies to combat hospitalizations in the early stages of a pandemic.

We recognize several limitations in this study. For example, the total number of official hospitalizations may not reflect the actual total amount of cases since H1N1 infection does not necessitate hospitalization (California Department of Public Health, 2009). Thus, the true total number of incidences of illness may be higher due to unreported or unconfirmed cases. In addition, given that there are issues of directing associations of risk exposure to individuals from grouped counts of individuals, judgments or conclusions about individuals from counties as units of observation is inappropriate (Washio et al., 2008). We utilized the best data available given our hypotheses and intent.

Accordingly, this is an ecological study and results must be considered as such. Additionally, we exclude hospitalizations beyond the point of nationwide vaccine distribution conducted in October 2009; thus, results pertain to the early half of the pandemic. Lastly, as with any study involving correlations, we do not imply that correlation equals causation.

In conclusion, influenza pandemics cannot be prevented, but they can be mitigated. An improved understanding of the environmental, social and economic factors dictating influenza survival and transmissibility is necessary for mitigating an emerging pandemic and its impairing effects. The use of spatial regression analysis enhances the understanding of the H1N1 pandemic in California and illustrates how disaster mitigation and response strategies for strengthening the well-being of at-risk populations can be improved.

By bringing the spatial, ecological risk-hazards framework to the forefront of pandemic H1N1 influenza studies, the scope of influenza pandemic studies has been broadened. Key ecological factors associated with H1N1 influenza hospitalization rates in California have been identified including support for the role of public transportation, agricultural areas and spatial dependence in pandemic influenza. These findings have direct implications for the efficacy of pharmaceutical and non-pharmaceutical intervention procedures for pandemic mitigation.

Acknowledgements

The authors thank the editor and anonymous reviewers for informative feedback that helped improve the paper. The authors acknowledge the California Department of Public Health, U.S. Census Bureau, Prism Climate Group and the Census of Agriculture for providing the data. The authors also thank Luc Anselin, Barbara Clare, Christopher Goranson, Frank Hardisty, Mark Horner, Lindsey Interlante, Kevin Konty, Michael Kuby, Elizabeth Mack, Nicholas Malizia, Paul Padegimas and Bohumil Svoma for helpful discussions.

References

- Anselin L, 1988. *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, pp. 284.
- Anselin L, 2002. Under the hood: issues in the specification and interpretation of spatial regression models. *Agr Econ* 27, 257-267.
- Anselin L, 2003. Spatial externalities, spatial multipliers, and spatial econometrics. *Int Regional Sci Rev* 26, 153-166.
- Anselin L, 2005. *Exploring spatial data with GeoDa: a workbook*. Urbana-Champaign, IL: Spatial Analysis Laboratory, Department of Geography, University of Illinois, USA.
- Anselin L, Bera AK, Florax R, Yoon MJ, 1996. Simple diagnostic tests for spatial dependence. *Reg Sci Urban Econ* 26, 77-104.
- Anselin L, Kim YW, Syabri I, 2004. Web-based analytical tools for the exploration of spatial data. *J Geogr Syst* 6, 197-218.
- Anselin L, Syabri I, Kho Y, 2006. GeoDa: an introduction to spatial data analysis. *Geogr Anal* 38, 5-22.
- Aras S, Aiyar A, Amedee AM, Gallaher WR, 2009. Molecular character of influenza A/H1N1 2009: implications for spread and control. *Indian J Microbiol* 49, 339-347.
- Assunção RM, Reis EA, 1999. A new proposal to adjust Moran's *I* for population density. *Stat Med* 18, 2147-2162.
- Bailey TC, Gatrell AC, 1995. *Interactive Spatial Data Analysis*. Longman, pp. 413.
- Balicer RD, Omer SB, Barnett DJ, Everly GS, 2006. Local public health workers' perceptions toward responding to an influenza pandemic. *BMC Public Health* 6, 99.
- Blendon RJ, Koonin LM, Benson JM, Cetron MS, Pollard WE, Mitchell EW, Weldon KJ, Herrmann MJ, 2008. Public response to community mitigation measures for pandemic influenza. *Emerg Infect Dis* 14, 778-786.
- Breusch T, Pagan A, 1979. A simple test for heteroscedasticity and random coefficient variation. *Econometrica* 47, 1287-1294.
- Brown D, 2009. New strain of swine flu investigated. *The Washington Post*, April 22.
- California Department of Public Health, 2009. Provisional number of pandemic (H1N1) 2009 hospitalizations, ICU and fatal cases in California, by local health jurisdiction. Retrieved from http://ww2.cdph.ca.gov/data/statistics/Pages/H1N1_FluDataTables.aspx (accessed on March 2010).
- CDC, 2009. General information about 2009 H1N1 vaccines. Centers for Disease Control and Prevention; retrieved from <http://www.cdc.gov/h1n1flu/vaccination/general.html> (accessed on December 2009).
- CDC, 2010. CDC estimates of 2009 H1N1 influenza cases, hospitalizations and deaths in the United States, April 2009 – March 13, 2010. Centers for Disease Control and Prevention; retrieved from http://www.cdc.gov/h1n1flu/estimates/April_March_13.html (accessed on July 2010).
- Cleaveland S, Laurenson MK, Taylor LH, 2001. Diseases of humans and their domestic mammals, pathogen characteristics, host range and the risk of emergence. *Philos T Roy Soc B* 356, 991-999.
- Cressie N, 1992. Smoothing maps using empirical Bayes predictors. *Geogr Anal* 24, 75-95.
- Crichton EJ, Elliott SJ, Moineddin R, Kanaroglou P, Upshur R, 2007a. A spatial analysis of the determinants of pneumonia and influenza hospitalizations in Ontario (1992-2001). *Soc Sci Med* 64, 1636-1650.
- Crichton EJ, Elliott SJ, Moineddin R, Kanaroglou P, Upshur

- REG, 2007b. An exploratory spatial analysis of pneumonia and influenza hospitalizations in Ontario by age and gender. *Epidemiol Infect* 135, 253-261.
- DeHart RL, 2003. Health issues of air travel. *Annu Rev Publ Health* 24, 133-151.
- Department of Agriculture, 2007. Census report; U.S. Department of Agriculture. Retrieved from http://www.agcensus.usda.gov/Publications/2007/Full_Report/index.asp (accessed on December 2009).
- Ferguson NM, Cummings DA, Fraser C, Cajka JC, Cooley PC, Burke DS, 2006. Strategies for mitigating an influenza pandemic. *Nature* 442, 448-452.
- Freedman DO, Leder K, 2005. Influenza: changing approaches to prevention and treatment in travelers. *J Travel Med* 12, 36-44.
- Germann TC, Kadau K, Longini IM, Macken CA, 2006. Mitigation strategies for pandemic influenza in the United States. *Proc Natl Acad Sci USA* 103, 5935-5940.
- Greger M, 2007. The human/animal interface: emergence and resurgence of zoonotic infectious diseases. *Crit Rev Microbiol* 33, 243-299.
- Hansen E, Donohoe M, 2003. Health issues of migrant and seasonal farm workers. *J Health Care Poor U* 14, 153-164.
- Inglesby TV, Nuzzo JB, O'Toole T, Henderson DA, 2006. Disease mitigation measures in the control of pandemic influenza. *Biosecur Bioterror* 4, 366-275.
- Jain S, Kamimoto L, Bramley AM, Schmitz AM, Benoit SR, Louie J, Sugarman DE, Druckenmiller JK, Ritger KA, Chugh R, Jasuja S, Deutscher M, Chen S, Walker JD, Duchin JS, Lett S, Soliva S, Wells EV, Swerdlow D, Uyeki TM, Fiore AE, Olsen SJ, Fry AM, Bridges CB, Finelli L, 2009. Hospitalized patients with 2009 H1N1 influenza in the United States, April-June, 2009. *New Engl J Med* 261, 1935-1944.
- Jarque CM, Bera AK, 1987. A test for normality of observations and regression residuals. *Int Stat Rev* 55, 163-172.
- Khan K, Arino J, Hu W, Raposo P, Sears J, Calderon F, Heidebrecht C, Macdonald M, Liauw J, Chan A, Gardam M, 2009. Spread of a novel influenza A (H1N1) virus via global airline transportation. *New Engl J Med* 361, 212-214.
- Lindsay WG, Blachere FM, Davis KA, Pearce TA, Fisher MA, Khakoo R, Davis SM, Rogers ME, Thewlis RE, Posada JA, Redrow JB, Celik IB, Chen BT, Beezhold DH, 2010. Distribution of airborne influenza virus and respiratory syncytial virus in an urgent care medical clinic. *Clin Infect Dis* 50, 693-698.
- Loeb M, 2003. Community-acquired pneumonia in older people: the need for a broader perspective. *J Am Geriatr Soc* 51, 539-543.
- Louie JK, Acosta M, Winter K, Jean C, Gavali S, Schechter R, Vugia D, Harriman K, Matyas B, Glaser CA, Samuel MC, Rosenberg J, Talarico J, Hatch D, 2009. Factors associated with death or hospitalization due to pandemic 2009 influenza A (H1N1) infection in California. *JAMA* 302, 1896-1902.
- Lowen AC, Mubareka S, Steel J, Palese P, 2007. Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathog* 3, 1470-1476.
- McLafferty S, 2010. Placing pandemics: geographical dimensions of vulnerability and spread. *Eurasian Geogr Econ* 51, 143-161.
- Moran PAP, 1950. Notes on continuous stochastic phenomena. *Biometrika* 37, 17-33.
- Nicholson KG, Wood JM, Zambon M, 2003. Influenza. *Lancet* 362, 1733-1745.
- North F, Syme SL, Feeney A, Head J, Shipley MJ, Marmot MG, 1993. Explaining socioeconomic differences in sickness absence: the Whitehall II study. *BMJ* 306, 361-366.
- Pearce-Duvett JMC, 2006. The origin of human pathogens: evaluating the role of agricultural and domestic animals in the evolution of human disease. *Biol Rev* 81, 369-382.
- Power AG, Mitchell CE, 2004. Pathogen spillover in disease epidemics. *Am Nat* 164, S79-S89.
- Slingenbergh J, Gilbert M, de Balogh K, Wint W, 2004. Ecological sources of zoonotic diseases. *Rev Sci Tech OIE* 23, 467-484.
- Tang JW, Lai FYL, Wong F, Hon KLE, 2010. Incidence of common respiratory viral infections related to climate factors in hospitalized children in Hong Kong. *Epidemiol Infect* 138, 226-235.
- Taylor PJ, Pocock SJ, 1972. Commuter travel and sickness absence of London office workers. *Br J Prev Soc Med* 22, 165-172.
- Tobler W, 1970. A computer model simulating urban growth in the Detroit region. *Econ Geogr* 46, 234-240.
- Turner BL, Kasperson RE, Matson PA, McCarthy JJ, Corell RW, Christensen L, Eckley N, Kasperson JX, Luers A, Martello ML, Polsky C, Pulsipher A, Schiller A, 2003. A framework for vulnerability analysis in sustainability science. *Proc Natl Acad Sci USA* 100, 8074-8079.
- U.S. Census Bureau, 2000. Statistical abstract of the United States. Retrieved from <http://www.census.gov/prod/www/statistical-abstract-us.html> (accessed on October 2009).
- U.S. Census Bureau, 2009. 2006-2008 American community survey 3-year estimates. Retrieved from <http://www.census.gov/acs/www/Products/> (accessed on October 2009).
- Viboud C, Boelle PY, Cauchemez S, Lavenue A, Valleron AJ, Flahault A, Carrat F, 2004. Risk factors of influenza transmission in households. *Br J Gen Pract* 54, 684-689.
- Washio M, Oura A, Mori M, 2008. Ecological studies of influenza infection and the effect of vaccination: their advantages and limitations. *Vaccine* 26, 6470-6472.
- Wein LM, Atkinson MP, 2009. Assessing infection control measures for pandemic influenza. *Risk Anal* 29, 949-962.