

---

## Effect of environmental factors on the spatio-temporal patterns of influenza spread

---

K. M. L. CHARLAND<sup>1,2,3,4\*</sup>, D. L. BUCKERIDGE<sup>2</sup>, J. L. STURTEVANT<sup>1,5</sup>,  
F. MELTON<sup>6,7</sup>, B. Y. REIS<sup>1,3,4</sup>, K. D. MANDL<sup>1,3,4</sup> AND J. S. BROWNSTEIN<sup>1,3,4</sup>

<sup>1</sup> *Children's Hospital Informatics Program at the Harvard-MIT Division of Health Sciences and Technology, Boston, MA, USA*

<sup>2</sup> *McGill Clinical and Health Informatics (MCHI), McGill University, Montreal, Quebec, Canada*

<sup>3</sup> *Division of Emergency Medicine, Children's Hospital Boston, Boston, MA, USA*

<sup>4</sup> *Department of Pediatrics, Harvard Medical School, Boston, MA, USA*

<sup>5</sup> *Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA*

<sup>6</sup> *California State University, at Monterey Bay, Seaside, CA, USA*

<sup>7</sup> *NASA Ames Research Center, Moffett Field, CA, USA*

(Accepted 2 February 2009; first published online 19 March 2009)

### SUMMARY

Although spatio-temporal patterns of influenza spread often suggest that environmental factors play a role, their effect on the geographical variation in the timing of annual epidemics has not been assessed. We examined the effect of solar radiation, dew point, temperature and geographical position on the city-specific timing of epidemics in the USA. Using paediatric in-patient data from hospitals in 35 cities for each influenza season in the study period 2000–2005, we determined ‘epidemic timing’ by identifying the week of peak influenza activity. For each city we calculated averages of daily climate measurements for 1 October to 31 December. Bayesian hierarchical models were used to assess the strength of association between each variable and epidemic timing. Of the climate variables only solar radiation was significantly related to epidemic timing (95% CI –0.027 to –0.0032). Future studies may elucidate biological mechanisms intrinsically linked to solar radiation that contribute to epidemic timing in temperate regions.

**Key words:** Climate, geographical position, influenza, meteorological variables, solar radiation.

### INTRODUCTION

Despite the year-round circulation of the influenza virus in tropical regions [1–3], in temperate zones, annual influenza epidemics consistently emerge in the winter months. In the tropics, even though seasonal changes in temperature and day length are less pronounced and influenza epidemics are smaller,

influenza incidence still tends to peak during the rainy season [2, 4]. The coincident timing of influenza epidemics with climatic changes suggests that environmental factors may play a role in effecting changes in the virus, the host or the interplay between host and virus.

Several features of the winter experience in temperate regions, such as colder outdoor temperatures, drier indoor conditions, reduced exposure to solar radiation and indoor crowding have been proposed as determinants of epidemic timing [2, 5–7]. Recent

\* Author for correspondence: Dr K. M. L. Charland, 1 Autumn Street, Room 439, Boston, MA 02215, USA.  
(Email: Katia.Charland@childrens.harvard.edu)

animal studies have linked low temperatures and low relative humidity to increased rates of aerosol but not contact transmission of the influenza virus [5, 6, 8]. Using data from France, from 1971 to 2002, Viboud *et al.* also found a higher incidence of influenza during the cold phases of the El Niño/Southern Oscillation, relative to the warm phases [7]. These studies contribute to the growing body of evidence that point to the role of low temperatures and humidity in facilitating the airborne transmission of the influenza virus. To better understand the importance of relative humidity and temperature in determining the timing of influenza epidemics, Lowen *et al.* called for further studies based on surveillance and meteorological data at a fine spatial resolution [6].

In addition to temperature and relative humidity, solar radiation may also play a role, as varying exposure to sunlight is known to affect innate immunity and may also inactivate the airborne virus [2, 9, 10]. During autumn and winter, reduced exposure to sunlight affects immune function by altering the endogenous production of specific immunomodulators [2, 11–15]. For example, dermal production of vitamin D<sub>3</sub>, an important immune system regulator [16] requires direct skin exposure to ultraviolet B radiation. Thus, during the winter, levels of circulating vitamin D<sub>3</sub> diminish, rendering the inhabitants of temperate regions more susceptible to infection [2, 15]. However, another immune system regulator and antioxidant, melatonin, increases with decreased daylight exposure and thus may offer some protection against influenza in northern areas with fewer daylight hours [11–13]. Findings from a recent study by Finkelman *et al.* provide some support for a negative association between solar radiation and epidemic timing, at least at the global scale [17]. Analysing patterns of influenza epidemic timing among 19 temperate countries, from 1997 to 2005, they found that influenza A epidemics were significantly earlier in low-latitude countries than in high-latitude countries.

Other studies of influenza epidemic timing in the tropics and subtropics have revealed a relationship between climate and influenza spread [4, 18]. Chew *et al.* investigated the effects of climate on the broader class of viral respiratory infections and found a significant positive association between influenza B incidence and precipitation [18]. In a study describing the geographical variation in epidemic timing in Brazil, Alonso *et al.* believed that the observed spatio-temporal patterns of spread were better explained by climate (i.e. humidity,

temperature and precipitation) than by population demographics and travel [4].

In this paper we examine the effect of latitude, longitude, temperature, dew point and solar radiation on the city-specific timing of peak influenza activity throughout the USA, in order to better understand the environmental and geographic factors that promote influenza spread in temperate regions.

## METHODS

### Timing of influenza epidemics

We acquired hospitalization records for all respiratory illness for each of 37 paediatric hospitals participating in the Pediatric Health Information System (PHIS), for the study period 2000–2005 (Fig. 1). Situated in large metropolitan areas, PHIS member hospitals are amongst the largest and most advanced paediatric hospitals in the USA. Since only a subset of respiratory illness visits are a result of influenza infection, we included only those visits with principal diagnosis (ICD-9) codes most closely related to influenza. More specifically, we defined an influenza case as an in-patient visit with an ICD-9 code equal to one of 486·00, 487·00, 487·1 or 487·8, a subset of codes that are known to be specific markers of influenza infection [19]. Restriction to a few influenza-associated ICD-9 codes provides greater specificity in identifying in-patient visits due to influenza and not other co-circulating viruses, such as respiratory syncytial virus (RSV). RSV incidence has been significantly associated with climate [18, 20] and misclassifying RSV cases as influenza cases could impact the results of the analysis.

The study period comprised five influenza seasons where each influenza season was defined as the 40th week of the calendar year to the 39th week of the following year. For each city and each influenza season in the study period, we found the weekly totals of influenza cases and defined the timing of the epidemic as the week with the maximum number of influenza cases (i.e. the ‘city-specific peak week’). We also considered, as candidates for ‘city-specific peak week’, those weeks in which the number of visits were within 5% of the maximum number of visits. For about 70% of the 35 cities × 5 seasons = 175 epidemics, we were able to identify a single week that qualified as the ‘peak week’.

To guide the choice of city-specific peak week in the case of multiple candidate weeks we used the

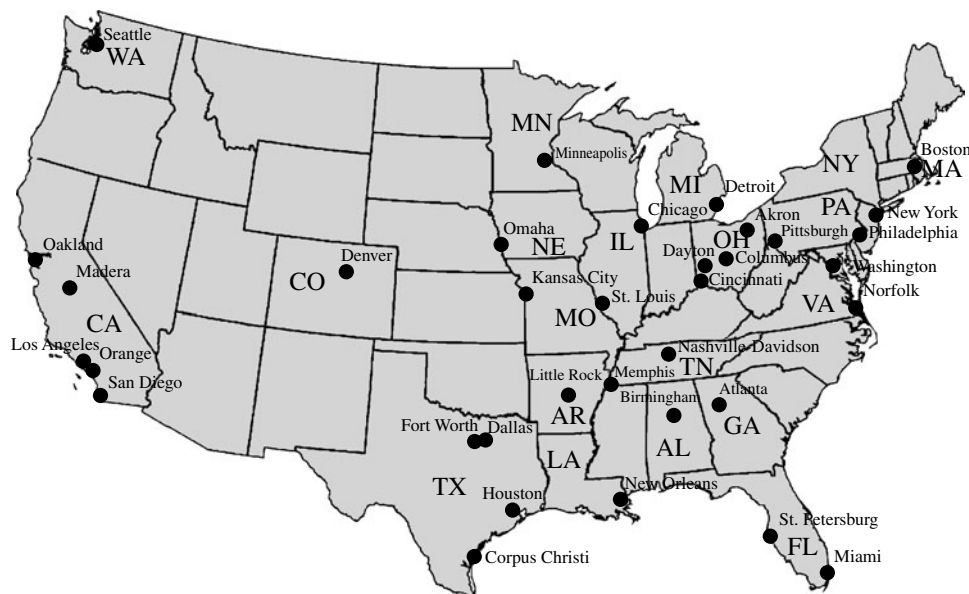


Fig. 1. Map of USA showing cities and states of Pediatric Health Information System hospitals.

regional WHO-NREVSS viral isolate data which are available at the CDC website [21]. Each year, WHO and NREVSS laboratories test a large number of specimens for the influenza virus. Weekly counts of positive specimens (viral isolates) are computed for each of nine regions that together comprise the 50 states of USA and the District of Columbia. Using these data, we determined the regional viral isolate peak week for each influenza season. In addition to the regional viral isolate peak weeks, we considered the city-specific RSV peak weeks, estimated from the series of weekly PHIS hospitalizations corresponding to the RSV-specific ICD-9 code 466.11. In selecting the city-specific influenza peak week our goal was to choose the candidate week that was preceded and followed by weeks of substantial influenza activity and the candidate week that corresponded more closely to the regional viral isolate peak week than the RSV peak week. First, if there were more than two candidates, we chose the two candidates that were closest to the WHO-NREVSS regional viral isolate peak week. The candidate that was finally chosen as the city-specific peak week satisfied at least two of the following three criteria: (1) the candidate week was closest to the regional viral isolate peak week, (2) the candidate week had the greatest number of visits in a 7-week period that included the 3 weeks preceding and following the candidate week and (3) the candidate week was furthest from the city-specific RSV peak week. The candidate week that

satisfied at least two of the three criteria was selected as ‘peak week’ for that city and influenza season.

#### Meteorological data and potentially confounding variables

For each city, daily average temperature and dew-point data were compiled using observations from the meteorological station closest to the hospital. These data were drawn from the National Oceanic and Atmospheric Administration (NOAA) National Climate Data Center (NCDC) Cooperative Summary of the Day TD3200 dataset. Daily short-wave solar radiation estimates were extracted from 8 km resolution gridded meteorological surfaces for the USA generated by the National Aeronautics and Space Administration (NASA) Ecological Forecasting Laboratory using the NASA Terrestrial Observation and Prediction System (TOPS). Incident short-wave solar radiation surfaces were generated by applying the DAYMET algorithm [22] which estimates incident solar radiation as a function of diurnal temperature range and sun-slope geometry. For each influenza season, we calculated city-specific averages of solar radiation, dew-point and average daily temperature measurements for the 3 months preceding the usual time of onset of epidemics (i.e. October–December). These daily climate variable averages were assessed as predictors of peak week in the statistical analyses.

In addition to the climate variables, we assessed the strength of association between three socio-demographic variables and peak week. The first was population size, a variable that was found to be a potential determinant of epidemic timing in the USA for data aggregated at the state level [23]. We also assessed the effect of population density on timing because increased crowding may promote faster spread of an epidemic and result in an earlier influenza peak. The last variable we considered was the percentage of the population aged <18 years. Studies on the age-specific timing of influenza activity have found that paediatric groups lead adult groups in timing of influenza activity [24, 25]. Increased rates of transmission may occur in areas that have a greater proportion of the population aged <18 years [26]. Data for socio-demographic variables were obtained from the United States Census 2000 [27].

### Missing data

Of the 37 hospitals, two were excluded from the analysis. For one of the hospitals (in Milwaukee, WI) we had insufficient meteorological data due to numerous data gaps. The second hospital (in Palo Alto, CA) had only 606 cases across the five influenza seasons, resulting in a time series without discernible influenza peaks. In addition, two cities, St Louis (MO) and Atlanta (GA) were missing solar radiation values for the first influenza season, so these were imputed with solar radiation measurements from nearby cities Kansas City (MO) and Birmingham (AL)/Nashville (TN), respectively. Two cities, New York City and Oakland were missing in-patient counts for the first two influenza seasons and their missing peak weeks were automatically imputed in the Bayesian analysis.

### Statistical analysis

Our goal was to assess the strength of the association between the city-specific climate variable averages and the city-specific peak weeks over the five influenza seasons of the study period. The correlations in measurements from the same city and measurements from the same influenza season preclude a regression analysis based on data pooled over all cities and influenza seasons. While an overall impression of the strength of association may have been attained from separate analyses by season, the range in the climate variable measurements (averages) was limited for some seasons so accuracy in parameter estimation

may have been compromised by using this approach. In addition, greater statistical power may be achieved with a statistical analysis that utilizes all the data at once. Thus, we used multilevel models to determine the individual contribution of each of the environmental factors on epidemic timing.

We applied a Bayesian approach, as inference for generalized mixed models can be particularly sensitive to the point estimates of the variance parameters when there is insufficient information in the data given the complexity of the model. Bayesian hierarchical models average over the uncertainty of the parameters of the model rather than using point estimates [28] thus, in the case of sparse data where the point estimates of variance parameters are imprecise, the Bayesian approach should give more reliable results. We also had to consider that latitude, temperature, solar radiation and dew point are correlated variables. Due to our relatively small sample size and the collinearity between the variables, we assessed the relationship between each variable with peak week without adjusting for the other covariates.

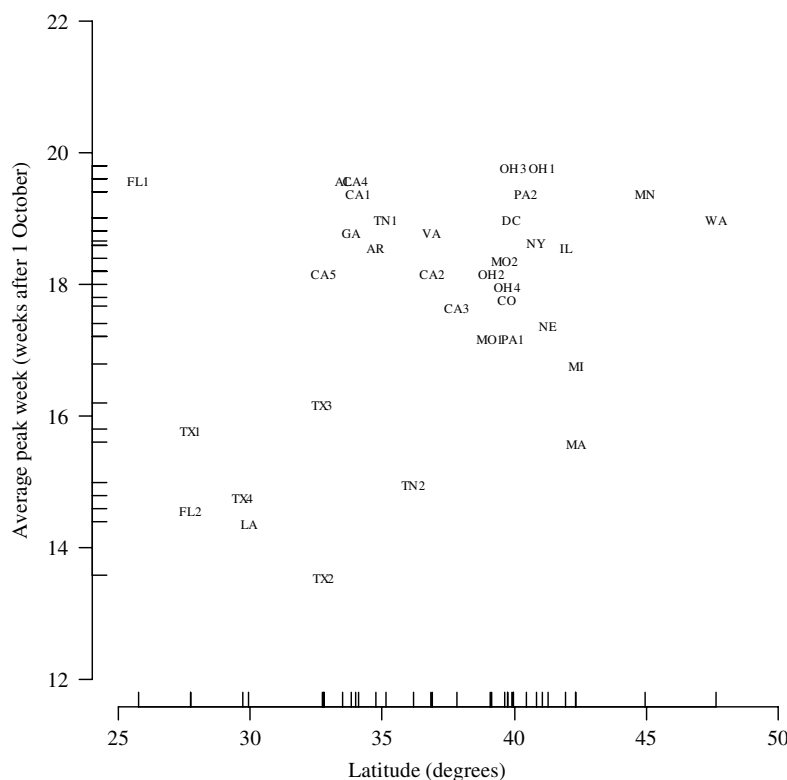
The Bayesian model we used included city and influenza season random effects to account for the correlations in multiple measurements from the same city and influenza season. Let  $y_i$  represent the peak week of the epidemic and  $x_i$  the independent variable (i.e. climate variable, latitude or longitude). Further, let the categorical variables  $j[i]$  and  $k[i]$  represent the city and the influenza season, respectively, in which measurement  $i$  was made. There were 35 cities and five influenza seasons giving a total of  $35 \times 5 = 175$  peak week, solar radiation, temperature and dew-point measurements. The formulation of the model is as follows:

$$y_i \sim N(a.city_{j[i]} + a.season_{k[i]} + \beta \cdot x_i, \sigma_y^2), \quad i = 1, \dots, 175$$

$$a.city_j \sim N(\mu_{a.city}, \sigma_{a.city}^2), \quad j = 1, \dots, 35$$

$$a.season_k \sim N(\mu_{a.season}, \sigma_{a.season}^2), \quad k = 1, \dots, 5,$$

where  $N(\mu, \sigma^2)$  represents the normal distribution with mean  $\mu$  and variance  $\sigma^2$ . The random effects for city and season are denoted by  $a.city$  and  $a.season$ , respectively. The city random effects are subject to the identifiability constraint  $\sum_{j=1}^{35} a.city_j = 0$ . We chose vague priors for the standard deviations and means of the random effects, Uniform(0, 100) and Normal(0,  $10^6$ ) distributions, respectively. All analyses were conducted using R and WinBUGS 1.4 software [31, 32]. WinBUGS software is particularly convenient



**Fig. 2.** Peak week averaged over the five influenza seasons vs. latitude of city (hospital). AL, Alabama (Birmingham); AR, Arkansas (Little Rock); CA, California (1, Los Angeles; 2, Madera; 3, Oakland; 4, Orange; 5, San Diego); CO, Colorado (Denver); DC, District of Columbia (Washington, DC); FL, Florida (1, Miami; 2, St Petersburg); GA, Georgia (Atlanta); IL, Illinois (Chicago); LA, Louisiana (New Orleans); MA, Massachusetts (Boston); MN, Minnesota (Minneapolis/St Paul); MO, Missouri (1, Kansas City; 2, St Louis); NE, Nebraska (Omaha); NY, New York (New York City); OH, Ohio (1, Akron; 2, Cincinnati; 3, Columbus; 4, Dayton); PA, Pennsylvania (1, Philadelphia; 2, Pittsburgh); TN, Tennessee (1, Memphis; 2, Nashville); TX, Texas (1, Dallas; 2, Corpus Christi; 3, Fort Worth; 4, Houston); VA, Virginia (Norfolk); WA, Washington (Seattle).

because it automatically performs multiple imputations for missing outcome data [32].

The potential scale reduction factor  $\hat{R}$  was used to assess convergence of the Markov Chain Monte Carlo (MCMC) sequence [33] and test runs demonstrated that with 30 000 iterations,  $\hat{R} < 1.01$  for all models, indicating that convergence was achieved. Thus, three chains of length 30 000 with different initial values were generated, 15 000 were discarded as burn-in iterations and the sequence was thinned, so that only every fifth simulated draw was retained leaving a posterior sample of 9000 for parameter estimation.

## RESULTS

From 1 October 2000 to 30 September 2005, there were a total of 63 334 PHIS cases that met the case definition. City-specific five season totals ranged from 734 (Omaha, NE) to 4844 (Houston, TX) with a mean of 1828 cases per city. The mean of solar radiation

averages for October–December was  $185.0 \text{ W/m}^2$  and ranged from  $40.13$  to  $316.9 \text{ W/m}^2$  over the five influenza seasons. The mean of the temperature and dew-point averages were  $11.0 \text{ }^\circ\text{C}$  (range  $-0.43$  to  $24.3$ ) and  $5.3 \text{ }^\circ\text{C}$  (range  $-6.3$  to  $9.7$ ), respectively. Latitude ranged from  $25.7^\circ$  to  $47.7^\circ$  and longitude ranged from  $-122.3^\circ$  to  $-71.1^\circ$ .

Figure 2 shows peak week averaged over the five influenza seasons plotted against latitude. In the multilevel modelling of peak week on geographical coordinates, we observed a positive association with latitude ( $0.13$  weeks per degree, 95% CI  $0.013$ – $0.24$ ) but not with longitude ( $0.0115$  weeks per degree, 95% CI  $-0.033$  to  $0.056$ ).

In Figure 3, we present the cities with the five earliest peak weeks in each season. Cities in Texas tended to be earlier in all but the 2004–2005 season. Epidemics in Texas were estimated to be about 3 weeks earlier than in other cities ( $-3.16$  weeks, 95% CI  $-4.69$  to  $-1.64$ ). However, relative to the Texas



cities, the epidemics in cities at similar southern latitudes in California tended to have later peaks. In fact, cities in California were never amongst the earliest (Fig. 3). In Figure 4, we show the series of hospital inpatient visits for four cities, Orange (CA), Cincinnati (OH), Houston (TX) and New Orleans (LA).

The results of the multilevel modelling, assessing the association of each climate variable with the peak week of influenza visits, are shown in Table 1. Solar radiation is negatively associated with peak week ( $-0.015$  weeks per  $W/m^2$ , 95% CI  $-0.027$  to  $-0.0032$ ). Figure 5 displays the graph of peak week vs. solar radiation averaged over the five influenza seasons. We did not find a similar significant relationship between peak week and temperature ( $-0.098$  weeks/ $^{\circ}C$ , 95% CI  $-0.21$  to  $0.021$ ) and peak week and dew point ( $-0.085$  weeks/ $^{\circ}C$ , 95% CI  $-0.20$  to  $0.031$ ). The socio-demographic variables were not associated with the week of peak influenza activity (Table 1).

To assess the sensitivity of the results to the choice of prior, we re-ran the analyses with three additional sets of priors for the variance parameters but the change in the results was minimal. Only average temperature and average dew point appeared marginally significant when using a gamma prior for the precisions. We also assessed sensitivity of the results to the time period of averaging for the climate variables. Rather than averaging weekly measurements over 1 October to 31 December, we re-ran the analyses using two alternatives: averages over 29 October to 31 December and averages over 26 November to 31 December. These changes did not alter the significance for any of the variables.

## DISCUSSION

In this study we have shown that, for the five influenza seasons between 2000 and 2005, cities with higher average solar radiation values over the time period October–December tended to have earlier influenza epidemics. Similarly, we found that latitude was positively associated to peak week but longitudinal position was not. Neither temperature nor dew point was significantly related to the week of peak influenza activity.

In previous studies, patterns of influenza spread have been associated with latitude, on the global scale, and longitude, at a finer geographic resolution in Japan and Europe [17, 34, 35]. Although we found

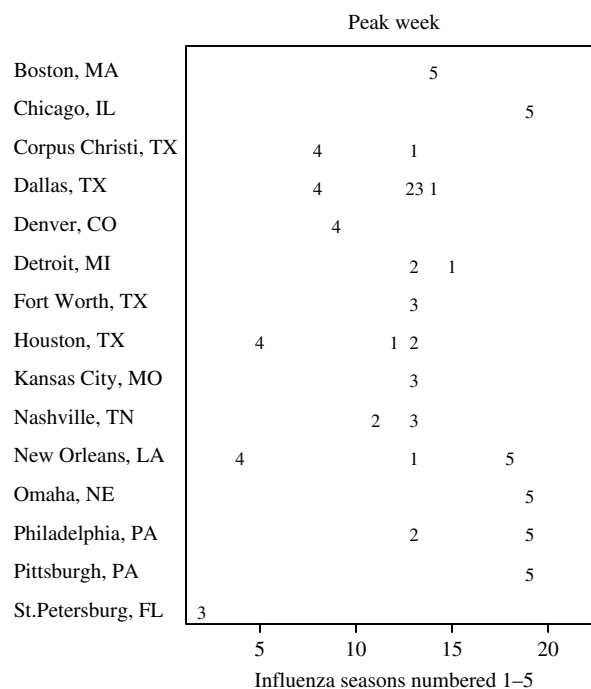


Fig. 3. Earliest five peak weeks for each influenza season.

a relationship between latitude and timing of peak influenza activity, this association may be explained, to some extent, by our finding of a negative association between solar radiation and peak timing. Although our objective was to determine whether the spatial variation in the timing of peak influenza activity could be related to the geographical variation in solar radiation, temperature and dew point, we note that these results do not support many of the current hypotheses that attempt to explain the winter timing of influenza epidemics. For example, if the timing of influenza epidemics were strongly related to indoor crowding, drier indoor conditions, colder outdoor temperatures, ultraviolet inactivation of the virus or reduced levels of circulating vitamin  $D_3$ , we would expect to see earlier epidemics in the northern cities of the USA. Even at the global scale, Finkelman *et al.* observed that countries closer to the equator had earlier epidemics than countries that were further from the equator, a finding that also goes against commonly held beliefs regarding the determinants of the winter timing of annual influenza epidemics [17].

A study by Viboud *et al.* revealed a higher incidence of influenza during the cold phases of the El Niño/Southern Oscillation, relative to the warm phases [7]. Other evidence in support of the role of temperature and relative humidity comes from animal studies in which lower temperatures and humidity were linked

Table 1. Relationship between climate, geographical and socio-demographic variables and timing of epidemic

Variable	Regression coefficient	S.D.	95% credible interval
Solar radiation (W/m <sup>2</sup> )	-0.015	0.0060	-0.027 to -0.0032 -0.031 to -0.00004 (99% CI)
Latitude	0.13	0.057	0.013 to 0.24
Longitude	0.0115	0.023	-0.033 to 0.056
Dew point (°C)	-0.085	0.059	-0.20 to 0.031
Average daily temp. (°C)	-0.098	0.060	-0.21 to 0.021
% of population < 18 yr	-0.074	0.15	-0.36 to 0.21
Population size (per 10 000)	0.005	0.024	-0.041 to 0.052
Population density (per 100 000 × square mile of area)	0.82	0.71	-0.58 to 2.28

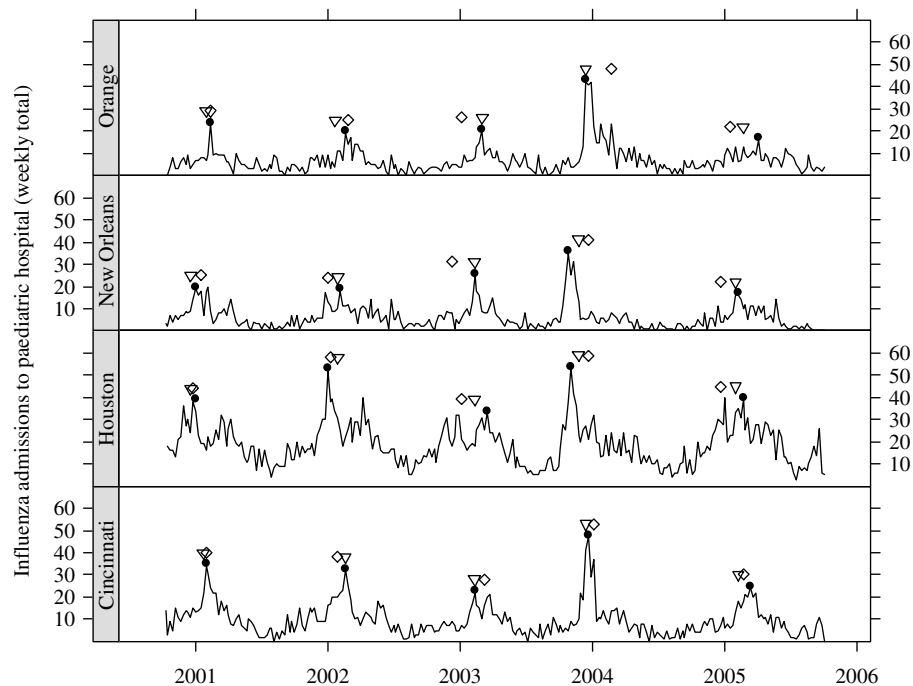


Fig. 4. Series of weekly hospital counts for four cities. Peak week for: influenza (●), regional viral isolate (▽), respiratory syncytial virus (◇).

to higher rates of aerosol (but not contact) transmission of the influenza virus [5, 6, 8]. Although our study focused on the spatio-temporal variation of epidemic timing (i.e. not incidence) within an influenza season, we did not find a significant relationship with temperature or dew point and the observed significant relationships with solar radiation and latitude would imply that, if there were a significant relationship, we would find higher temperatures and dew point associated with earlier epidemics.

While the overall spatio-temporal trends of peak influenza activity give evidence of the role of solar radiation and latitude as predictors of epidemic timing, the specific patterns in the geographical variation of peak influenza activity deserve further attention. Viboud *et al.* used state-specific counts of excess mortality due to pneumonia and influenza to examine the timing of influenza epidemics for the 48 contiguous states of the USA (and the District of Columbia) for the study period 1972 to 2002 and found that

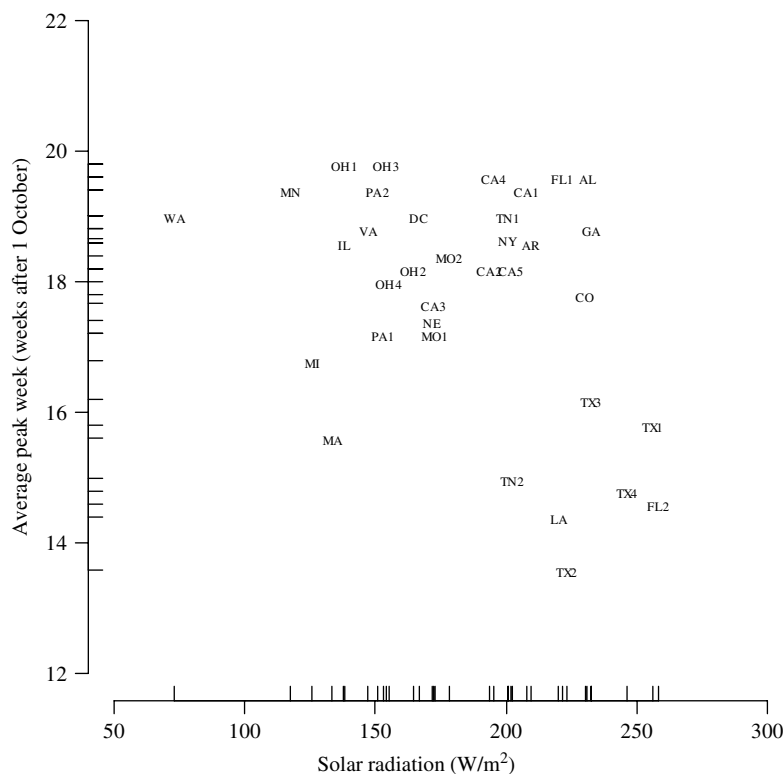


Fig. 5. Peak week vs. solar radiation averaged over the five influenza seasons. (See Fig. 2 legend for abbreviations.)

influenza epidemics most often started in the state of California and the more populous states had a greater tendency towards synchrony in the timing of epidemics [23]. In our study, we found an association with latitude but this observed association may have been partly driven by the early epidemics in a few south-eastern cities (e.g. cities in eastern Texas were earlier by an average of 3 weeks).

Unlike other studies, in which air travel, connections to Asia and Australia and population size have been offered as determinants of influenza spread [23, 36], we did not find an association between city population size or density and epidemic timing and the cities with the greatest share of overseas visitors, namely New York City and Los Angeles [37], did not have earlier epidemics. While it is true that these conclusions are based on only 5 years of data, lifestyle and air travel have changed in the last 30 years and, as a result, the important determinants of influenza spread may also have changed somewhat during that period.

Although the results concerning the relationship between solar radiation and epidemic timing are robust, given that observations from the same city and observations from the same year are correlated, this study was limited by a relatively small sample size of 35 cities, each with just five seasons of in-patient

visit data. Furthermore, we had a greater number of cities in the more humid eastern region of the country and those cities in western USA were mostly situated in the south. The inclusion of a greater number of cities in the northern extremes of latitude and in the more arid western regions may have revealed a more pronounced effect of latitude and may have enabled the detection of a significant association with temperature and dew point. In addition, a longer period of study may have permitted an examination of the interaction between the environmental factors and the influenza season's predominant strain.

In our analyses, we did not adjust for multiple comparisons, but our main study objective involved an assessment of the association of only five variables with epidemic timing, thus the expected number of false positives is  $<1$ . Moreover, solar radiation was significant at the 0.01 level and with a threshold for significance set at 0.01 for each test, the significance level for the family of five tests is  $<0.05$  (i.e.  $5 \times 0.01$ ).

Epidemiological studies of influenza are often limited by a broad case definition and the lack of viral confirmation for cases. Influenza-like illness encompasses many respiratory illnesses, including RSV and influenza. Children (i.e. aged  $<18$  years), in



particular, have different clinical presentations which may lead to misclassification. To avoid misclassification and confusing RSV peak weeks for influenza peak weeks, we included only in-patient visits with ICD-9 codes belonging to a subset of codes that have been shown to be specific for influenza [19]. In addition, when there were several candidates for peak week, we used the CDC regional viral isolate peaks and city-specific RSV peaks (estimated from the PHIS data) to guide the choice of peak that would best represent the true influenza peak. In children aged <5 years, RSV accounts for a greater proportion of respiratory infection than does influenza. However, the contribution of RSV and influenza is about equal in children aged 5–14 years [38]. Thus, it is plausible that restricting the case definition to children aged  $\geq 5$  years will lead to more accurate identification of influenza cases. Unfortunately, only about one third of the cases fitting our influenza case definition were aged  $\geq 5$  years. Consequently, the series of weekly counts of cases in this age group was particularly noisy for hospitals with smaller total counts and epidemic peaks were not easily identified. Rather than basing our analyses on children aged  $\geq 5$  years, we validated a subset of our peak-week measurements identified from the all-ages series with those from the  $\geq 5$  years series. To do this, we determined the peak weeks for each season for eight hospitals with larger total counts and using a multilevel model with a dichotomous independent variable (i.e. ‘all-ages’ = 0, ‘ $\geq 5$  years’ = 1) we did not find a significant difference between the 8 hospital  $\times$  5 season = 40 peak-week measurements from the two datasets, giving greater reassurance that we are indeed identifying influenza and not RSV peak weeks. Even so, healthcare utilization data encompasses a complex behavioural component and as these data were not collected for the specific purpose of this research, the results are subject to the standard limitation of analysing administrative data.

We chose to average meteorological variable measurements over the 3-month period of October–December and considered two other time periods for averaging in a sensitivity analysis but averages over shorter periods may be revealing. Further, future research should investigate the effects of other environmental factors, such as precipitation and diurnal temperature range, as previous studies have found evidence of their effect on influenza spread [18, 39]. Although solar radiation was found to be significantly related to peak timing, the relationship between these

variables can be further refined. In particular, future research could provide more detailed insight into possible threshold effects of these meteorological variables on the peak timing of epidemics.

Peak week of the epidemic was used as the marker of epidemic timing in our study because it is one of the most easily identified features of the distribution of incident cases in a given season. Although we used the week corresponding to the crude peak, when there was not a single, well-defined peak, estimates of the peak week from time-series analytical methods or estimates of the mean week may have better captured the week corresponding to the centre of the epidemic. Onset and end of the epidemic may be more difficult to discern, particularly for data at a fine geographical resolution where the series of counts of influenza cases are noisy. If more stable series of counts can be formed, for example, by combining in-patient data from multiple hospitals in a city, identification of onset and end of the epidemic may be more accurately estimated and markers of spread based on these may be considered. For example, a composite measure of magnitude and duration may reveal interesting patterns in influenza spread.

In this study, using data at a fine spatial resolution, we observed a significant positive relationship between latitude and epidemic timing, indicating that, in the USA, cities at lower latitudes tend to have earlier epidemics relative to cities at higher latitudes. This complements the findings of Finkelman *et al.* that revealed a positive relationship between latitude and epidemic timing on the global scale [17]. In addition, the significant inverse relationship between solar radiation and epidemic timing potentially offers a biological basis for the observed significance of latitudinal position. Future studies may help to elucidate the possible biological mechanisms underlying solar radiation that may contribute to the spatio-temporal patterns of influenza spread in temperate regions.

## ACKNOWLEDGMENTS

This work was supported by R21AI073591-01 from the National Institute of Allergy and Infectious Diseases, R21LM009263-01 from the National Library of Medicine, National Institutes of Health, and PAN-83152 from the Canadian Institutes of Health Research. The authors are grateful to Dr Marc Lipsitch for helpful suggestions and comments. Thanks are also due to Dionne Graham for help in acquiring the hospitalization data.

## DECLARATION OF INTEREST

None.

## REFERENCES

- Russell CA, *et al.* The global circulation of seasonal influenza A (H3N2) viruses. *Science* 2008; **320**: 340–346.
- Cannell JJ, *et al.* Epidemic influenza and vitamin D. *Epidemiology and Infection* 2006; **134**: 1129–1140.
- Rambaut A, *et al.* The genomic and epidemiological dynamics of human influenza A virus. *Nature* 2008; **453**: 416–419.
- Alonso WJ, *et al.* Seasonality of influenza in Brazil: a traveling wave from the Amazon to the subtropics. *American Journal of Epidemiology* 2007; **165**: 1434–1442.
- Polozov IV, *et al.* Progressive ordering with decreasing temperature of phospholipids of influenza virus. *Nature Chemical Biology* 2007; **4**: 248–255.
- Lowen AC, *et al.* Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathogens* 2007; **3**: e151.
- Viboud C, *et al.* Association of influenza epidemics with global climate variability. *European Journal of Epidemiology* 2004; **19**: 1055–1059.
- Lowen AC, *et al.* High temperature (30 °C) blocks aerosol but not contact transmission of influenza virus. *Journal of Virology* 2008; **82**: 5650–5652.
- Sagripanti J-L, Lytle CD. Inactivation of influenza virus by solar radiation. *Photochemistry and Photobiology* 2007; **83**: 1278–1282.
- Jensen MM. Inactivation of airborne viruses by ultraviolet irradiation. *Applied Microbiology* 1964; **12**: 418–420.
- Srinivasan V, *et al.* Melatonin, environmental light and breast cancer. *Breast Cancer Research and Treatment* 2008; **108**: 339–350.
- Carrillo-Vico A, *et al.* A review of the multiple actions of melatonin on the immune system. *Endocrine* 2005; **27**: 189–200.
- Vondrašová D, Hájek I, Illnerová H. Exposure to long summer days affects the human melatonin and cortisol rhythms. *Brain Research* 1997; **759**: 166–170.
- Maxwell JD. Seasonal variation in vitamin D. *Proceedings of the Nutrition Society* 1994; **53**: 533–543.
- Hope-Simpson RE. The role of season in the epidemiology of influenza. *Journal of Hygiene* 1981; **86**: 35–47.
- Cantorna MT, *et al.* Vitamin D status, 1,25-dihydroxyvitamin D<sub>3</sub>, and the immune system. *American Journal of Clinical Nutrition* 2004; **80** (Suppl.): 1717S–1720S.
- Finkelman BS, *et al.* Global patterns in seasonal activity of influenza A/H3N2, A/H1N1, and B from 1997 to 2005: viral coexistence and latitudinal gradients. *PLoS One* 2007; **12**: e1296.
- Chew FT, *et al.* Seasonal trends of viral respiratory tract infections in the tropics. *Epidemiology and Infection* 1998; **121**: 121–128.
- Marsden-Haug N, *et al.* Code-based syndromic surveillance for influenza-like illness by International Classification of Diseases, Ninth Revision. *Emerging Infectious Disease* 2007; **13**: 207–216.
- Yusuf S, *et al.* The relationship of meteorological conditions to the epidemic activity of respiratory syncytial virus. *Epidemiology and Infection* 2007; **135**: 1077–1090.
- CDC (Centers for Disease Control and Prevention). Flu activity and surveillance. (<http://www.cdc.gov/flu/weekly/fluactivity.htm#IAPMSurveillance>). Accessed 2 May 2008.
- Thornton PE, Running SW, White MA. Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology* 1997; **190**: 214–251.
- Viboud C, *et al.* Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* 2006; **312**: 447–451.
- Olson DR, *et al.* Monitoring the impact of influenza by age: emergency department fever and respiratory complaint surveillance in New York City. *PLoS Medicine* 2007; **4**: 1349–1361.
- Brownstein JS, Kleinman KP, Mandl KD. Identifying pediatric age groups for influenza vaccination using a real-time regional surveillance system. *American Journal of Epidemiology* 2005; **162**: 686–693.
- Brownstein JS, Mandl KD. Pediatric population size is associated with geographic patterns of acute respiratory infections among adults. *Annals Emergency Medicine* 2008; **52**: 63–68.
- United States Census 2000 (<http://factfinder.census.gov/home/saff/main.html?lang=en>). Accessed 2 May 2008.
- Gelman A, Hill J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press, 2007.
- Glantz SA, Slinker BK. *Applied Regression & Analysis of Variance*, 2nd edn. New York: McGraw-Hill, 2001.
- Tu YK, *et al.* Problems of correlations between explanatory variables in multiple regression analyses in the dental literature. *British Dental Journal* 2005; **199**: 457–461.
- R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2005 (<http://www.R-project.org>).
- Spiegelhalter DJ, *et al.* WinBUGS: Bayesian Inference Using Gibbs Sampling Manual, version 1.4, 2002. London: Imperial College. Cambridge, UK: MRC Biostatistics Unit (<http://www.mrc-bsu.cam.ac.uk/winbugs>).
- Gelman A, *et al.* *Bayesian Data Analysis*, 2nd edn. Boca Raton: Chapman & Hall/CRC, 2004.
- Paget J, *et al.* Influenza activity in Europe during eight seasons (1999–2007): an evaluation of the indicators used to measure activity and an assessment of the timing, length and course of peak activity (spread) across Europe. *BMC Infectious Diseases* 2007; **7**: 141–147.

35. Sakai T, *et al.* Geographic and temporal trends in influenza-like illness, Japan, 1992–1999. *Emerging Infectious Disease* 2004; **10**: 1822–1826.
36. Brownstein JS, Wolfe CJ, Mandl KD. Empirical evidence for the effect of airline travel on inter-regional influenza spread in the United States. *PLoS Medicine* 2006; **3**: e401.
37. United States Department of Transportation. Overseas visitors to the United States: top 20 destination states and territories, 2001 ([http://www.bts.gov/publications/state\\_transportation\\_statistics/pennsylvania/html/table\\_04\\_05.html](http://www.bts.gov/publications/state_transportation_statistics/pennsylvania/html/table_04_05.html)). Accessed 16 May 2008.
38. Fleming DM, *et al.* Respiratory illness associated with influenza and respiratory syncytial virus infection. *Archives of Disease in Childhood* 2005; **90**: 741–746.
39. Liu CM, *et al.* Temperature drops and the onset of severe avian influenza A H5N1 virus outbreaks. *PLoS One* 2007; **2**: e191.