Spatial and Spatio-temporal Epidemiology 8 (2014) 35–45

Contents lists available at ScienceDirect

Spatial and Spatio-temporal Epidemiology

journal homepage: www.elsevier.com/locate/sste

# The spread process of epidemic influenza in the continental United States, 1968–2008

Bianca L. Malcolm *,[1]

*Columbia University Mailman School of Public Health, 722 West 168th Street, New York, NY 10032, USA*

ARTICLE INFO

*Article history:*
Received 10 July 2013
Revised 12 November 2013
Accepted 10 January 2014
Available online 22 January 2014

*Keywords:*
Influenza, Human
Mortality
Spatio-temporal analysis
Cluster analysis
Geographic mapping
Epidemic

ABSTRACT

Understanding the quantitative disease dynamics of influenza is important in developing strategies to control its spread. This research analyzed the dominant spread process of epidemic influenza in the continental United States over a 41-year period. Spatial autocorrelation and simple correlation were applied to pneumonia and influenza mortality to observe the effect of distance and population on the between-state transmission of seasonal influenza. Annual influenza epidemics exhibited distance-based spatial spread at the peak of activity, but did not undergo significant population-based spread at any point. Geographically-close states (<500 miles) showed higher correlations in the start, peak and end of annual epidemics compared with geographically-distant states. Additionally, significant local clustering was found in the Midwest, Ohio River Valley and Northeastern regions as well as Nevada and Utah throughout an influenza season. This research may be combined with others in order to determine the main epidemic pathways of seasonal influenza in the US.



## 1. Introduction

Recurrent epidemics of influenza occur annually during the winter season in temperate areas of the world, such as the United States. They are the result of genetic drift in which, in order to escape host immunity, the surface antigens of influenza viruses undergo small changes (Smith et al., 2004). These annual influenza epidemics cause considerable morbidity, mortality and economic burden (Simonsen et al., 1999). In the US alone, approximately 24,000 deaths a year can be attributed to influenza (CDC, 2010).

It is crucial to understand quantitatively how a disease spreads in modern society. The sudden appearance of the 2009 H1N1 pandemic increased interest in the design of efficient containment policies and demands for an accurate characterization of spatial and temporal epidemic influenza patterns (Mills et al., 2006; Colizza et al., 2007; Ferguson et al., 2005; Ferguson et al., 2006; Longini et al., 2005). One of the most important control strategies that arose out of the 2009 pandemic was the need to identify the main channels of transmission or "epidemic pathways" of seasonal influenza in the US. In fact, identification of these pathways is the first clue on how to control influenza's spread (Colizza et al., 2006).

While much is understood about the make-up and impact of seasonal influenza in the US, the spatial pattern of epidemic influenza has been less well characterized. Prior studies have analyzed the spread of influenza, developed unique models to describe and understand this disease, and explained the spatial distribution of influenza spread and of annual waves of infection in the US (Anderson and May, 1991; Baroyan et al., 1969; Bonabeau et al., 1998; Rvachev and Longini, 1985; Viboud et al., 2004). However, these studies have failed to detect the preferred channels of transmission for epidemic influenza in the US.

This study describes a method used to identify the dominant spreading process in the US. Spatial autocorrelation

* Tel.: +1 212 305 3927.
*E-mail addresses:* blm2115@columbia.edu, BMalcolm@cdc.gov
[1] Present address: 3311 Toledo Road, Rm. 7222, Hyattsville, MD 20782, USA. Tel.: +1 301 458 4467.

1877-5845/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.sste.2014.01.001

was combined with simple correlation to illustrate the disease dynamics of epidemic influenza. The objective of this study was to find robust transmission channels for epidemic spread by determining the dominant spread process of epidemic influenza in the continental United States over a 41-year period.

## 2. Methods

To examine the spatial structure of influenza epidemics in the continental United States between 1968 and 2008, monthly counts of pneumonia and influenza (P&I) mortality were obtained from the National Centers for Health Statistics (NCHS) and Centers for Disease Control and Prevention (CDC) Wonder Online Data System for Multiple Cause of Death (NCHS, 2012a; CDC, 2012). Population estimates used in the calculation of mortality rates were obtained from the Census Bureau (National Cancer Institute, 2012; NCHS, 2012b). All non-fetal deaths are included in this analysis.

P&I mortality is treated as a proxy for influenza activity in these analyses. Pneumonia is a frequent complication of influenza and it is often difficult to distinguish clinically between the two. Furthermore, testing for influenza is rarely performed during an autopsy, therefore few deaths are classified as caused by influenza. Even though influenza causes only a small fraction of deaths attributed to pneumonia, combining these two causes of death has been previously shown to be a reliable proxy for studying the timing and amplitude of influenza activity (CDC, 2010; Simonsen et al., 1999; Thompson et al., 2009; Viboud et al., 2004; 2006). The proportion of influenza-attributable deaths in P&I mortality data is unimportant to measure the timing of seasonal influenza because of the seasonal pattern of pneumonia deaths, which see a significant increase in the winter months because of influenza's acerbation of existing morbidities despite pneumonia being recorded year-round (CDC, 2010).

### 2.1. Definition of influenza seasons

The period between September and August of the following year comprised an influenza "epidemic period". A seasonal wave was then identified within this time period in order to identify the critical time points of influenza seasons (e.g. start, peak, end) because P&I deaths are recorded year-round in the US, which makes assessment of the timing of a season difficult. Careful definitions of exposure allow for better assessment of the impact of timing on outcomes such as deaths (Sandoval et al., 2008).

An annual seasonal influenza wave in the US was determined by first reworking the aggregate monthly P&I mortality rates into standard normal scores. A monthly normal score in excess of 0.5 standard deviations above the zero mean, with an additional "lead in" calendar month at the beginning denoted the official beginning of an influenza season. Two consecutive drops in the monthly normal score below 0.5 standard deviations below the zero mean, or if the end of the epidemic period was reached, with a "lead out" month added to the end, signaled the end of the influenza season (Smallman-Raynor and Cliff, 1998). This process was used to identify annual seasonal influenza waves for the US as a whole as well as for each of the 48 contiguous states (and the District of Columbia) in the dataset for each of the 40 influenza seasons between 1968 and 2008.

Of 492 months in the dataset, 268 were identified as a part of an epidemic influenza season. Within these epidemic periods, three phases were denoted: build-up, peak and fadeout. The build-up phase refers to the period from the start of an epidemic to its peak. An epidemic's peak is when its P&I mortality rate is at its highest. The fadeout phase is the period after the epidemic peak to the end of an epidemic.

### 2.2. Spread processes of influenza mortality

A technique of autocorrelation on graphs was used to determine the nature of the processes that underlie the spread patterns of epidemic influenza in the US. To perform spatial autocorrelation analysis, the area in which spread occurs (i.e. United States) was treated as a graph consisting of a set of nodes, states, and the links between them. The links were chosen to create a graph which corresponds with the hypothetical diffusion process. Analyses examined three main types of diffusion process for infectious disease: contagious, hierarchical and mixed (Smallman-Raynor and Cliff, 1998; 2001). This analysis produced three graphs. In order to distinguish these graphs, the following BW statistic was used:

$$BW = \frac{1}{2}\sum_{n=1}^{n}\sum_{n=1}^{n} w_{ij}(x_i - x_j)^2$$

$$i \neq j$$

where $w_{ij} = 1$ if a link existed between the vertices $i$ and $j$ on the graph in question and $w_{ij} = 0$ otherwise.

A contagious process represents a highly localized diffusion process and implies that influenza mortality moves in a wave-like pattern from its center of introduction to its closest neighbor. A contagious graph represents the simplest network configuration that links all geographical units and minimizes the straight-line distance between them. For each month of the study period, monthly P&I mortality rate for geographic units $i$ and $j$, $\omega_{ij}$, was set to 1 in a matrix $W$ if geographical units $i$ and $j$ were linked in the configuration, and $\omega_{ij} = 0$ otherwise. The effect of distance on the correlation of P&I mortality between geographic units decreased as distance between two geographic units increased.

A hierarchical, or population-based, process would imply that influenza mortality in the US moves progressively from more populous states to less populous states. In a hierarchical graph, $\omega_{ij} = 1$ if geographic unit $j$ was the next larger or the next smaller unit in population size to unit $i$ and $\omega_{ij} = 0$ otherwise. A mixed process would imply that influenza mortality in the US follows a spread pattern that contains both contagious and hierarchical components. In the mixed contagious-hierarchical graph, $\omega_{ij} = 1$ if unit $j$ was the geographically nearest unit that was either larger or smaller in population size than unit $i$, and $\omega_{ij} = 0$ otherwise.