SPATIAL AND TEMPORAL DYNAMICS OF INFLUENZA

by

James H. Stark

BS, American International College, 1996

Sc.M, Johns Hopkins School of Public Health, 2000

Sc.M, Harvard School of Public Health, 2004

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2011

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This dissertation was presented

by

James H. Stark

It was defended on

March 23, 2011

and approved by

Dissertation Advisor:
Stephen Wisniewski, PhD, Professor, Department of Epidemiology
Graduate School of Public Health, University of Pittsburgh

Committee Member:
Derek Cummings, PhD, Assistant Professor, Department of Epidemiology
Bloomberg School of Public Health, The Johns Hopkins University

Committee Member:
Bard Ermentrout, PhD, Professor, Department of Mathematics
School of Arts and Sciences, University of Pittsburgh

Committee Member:
Ravi Sharma, PhD, Assistant Professor, Department of Behavioral and Community Health Sciences, Graduate School of Public Health, University of Pittsburgh

Committee Member:
Stephen Ostroff, MD, Director, Bureau of Epidemiology
Pennsylvania Department of Health

Committee Member:
Samuel Stebbins, MD, MPH, Assistant Professor, Department of Epidemiology
Graduate School of Public Health, University of Pittsburgh

SPATIAL AND TEMPORAL DYNAMICS OF INFLUENZA

James H. Stark, PhD

University of Pittsburgh, 2011

Despite the significant amount of research conducted on the epidemiology of seasonal influenza, the patterns in the annual oscillations of influenza epidemics have not been fully described or understood. Furthermore, the current understanding of the intrinsic properties of influenza epidemics is limited by the geographic scales used to evaluate the data. Analyses conducted at larger spatial scales may potentially conceal local trends in disease structure which may reveal the effect of population structure or environmental factors on disease spread. By using influenza incidence data from the Commonwealth of Pennsylvania and United States influenza mortality data, this dissertation characterizes seasonal influenza epidemics, evaluates factors that drive local influenza epidemics, and provides an initial assessment in how administrative borders influence surveillance for local and regional influenza epidemics.

Evidence of spatial heterogeneity existed in the distribution of influenza epidemics for Pennsylvania counties resulting in a cluster of elevated incidence in the South Central region of the state that persisted during the entire study period (2003-2009). Lower monthly precipitation levels during the influenza season (OR = 0.52, P = 0.0319), fewer residents over age 64 (OR = 0.27, P = 0.01) and fewer residents with more than a high school education (OR = 0.76, P = 0.0148) were significantly associated with membership in this cluster. In addition, significant synchrony in the timing of epidemics existed across the entire state and decayed with distance (regional correlation $\rho = 62\%$). Synchrony as a function of population size displayed evidence of

hierarchical spread with more synchronized epidemics occurring among the most populated counties. A gravity model describing movement between two populations was the best predictor of influenza spread suggesting that non-routine and leisure travel drive local epidemics. Within the United States, clusters of epidemic synchronization existed, most notably in densely populated regions where connectivity is stronger.

Observation of county and state epidemic clusters highlights the importance and necessity of correctly identifying the ontologic unit of epidemicity for influenza and other diseases. Recognition of the appropriate geographic unit to implement effective surveillance and prevention methods can strengthen the public health response and minimize inefficient mechanisms.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

1.1 OVERVIEW

Influenza is a contagious respiratory pathogen responsible for annual seasonal epidemics in temperate climates resulting in significant morbidity and mortality. Epidemics in the northern hemisphere typically occur between October and April. The extent of antigenic variation of the virus, protective immunity in the population, and virus virulence influences the relative size and impact of the epidemic each year. Many characteristics of the transmission dynamics within households have been elucidated; however, less is known about the transmission dynamics at larger spatial scales and over longer time periods. Improvement in surveillance systems has fostered a new period of spatiotemporal analyses enabling researchers to reveal hierarchical transmission patterns. Ultimately, recognition of such trends could direct limited human and economic resources to improve control strategies aimed at minimizing transmission through targeted vaccinations, directed hygienic advertisements and informed surveillance.

1.2 VIROLOGY

Influenza viruses consist of eight segmented genomes and belong to the family *Orthomyxoviridae* [1]. Of the three influenza viruses A, B and C, only influenza A and B are

known to cause substantial seasonal morbidity and mortality [2]. Further classification of influenza viruses into subtypes is based on the antigenic and genetic differences between the two surface glycoproteins, hemagglutinin (HA) and neuraminidase (NA). The greatest glycoprotein diversity occurs with the Influenza A virus which has 16 HA and 9 NA identifiable subtypes. In contrast, influenza B has only one recognizable hemagglutinin and neuraminidase [3]. Hemagglutinin serves as major surface antigen responsible for binding virions to the host cell and thus facilitating entry [4]. Vaccines target this membrane protein. The membrane protein, neuraminidase, is responsible for cleaving the progeny virions from the host cell receptors enabling virus spread and is the focus of antiviral drugs and vaccines [3].

Seasonal epidemics and pandemics are maintained by antigenic variation associated with the two surface glycoproteins, HA and NA. Changes in the antigenic composition of influenza surface proteins, irrespective of previous influenza vaccination or infection by influenza, leave an individual susceptible to the new influenza strain. Two types of variation exist: antigenic drift and antigenic shift. Antigenic drift refers to the evolution of a new influenza virus strain from the accumulation of point mutations in the HA and NA genes. This new variant strain escapes immune recognition generated by previous strains and results in outbreaks in inter-pandemic years among both influenza A and B viruses [2]. Antigenic shift occurs when antigenic change leads to the development of influenza A virus with a novel HA or NA glycoprotein. Profound antigenic change, like that occurring with antigenic shift, in a susceptible human population often leads to a p andemic influenza strain. Genetic reassortment between human and animal influenza viruses or direct animal to human transmission is the source of antigenic shift [1]. Acquisition of HA and NA genes through genetic reassortment coincided with pandemics in 1957, 1968, 1977, and more recently, the H1N1 pandemic in 2009 [5].

1.3 EPIDEMIOLOGY

1.3.1 Transmission and Clinical Illness

The virus replicates in respiratory epithelial cells where it gains access to respiratory secretions enabling direct spread to occur through the formation of droplets created by coughing, sneezing, or speaking [1]. The opportunity for infection through indirect contact by touching an infected surface (fomite) has also been observed. The median incubation period for influenza A virus is 1.4 days, though the period ranges from 1 to 3 days [6]. Upon infection, influenza illness is marked by an abrupt onset in adults and children and symptoms include fever, chills, cough, sore throat, myalgia, and a headache [7, 8]. Gastrointestinal symptoms have been observed in children and symptoms persist for several days [1, 7, 8]. Viral shedding is likely to be an important determinant in transmission and infectivity and the median duration is 5 days though children are thought to shed longer than adults [9, 10].

1.3.2 Morbidity and Mortality

Seasonal epidemics in the Northern Hemisphere occur in the winter months between October and April. Within the United States, peak activity most often occurs in February [11]. Estimating the burden of influenza, specifically incident cases and mortality due to influenza, remains a challenge. Passive surveillance systems may not capture all of the incident cases and influenza-specific deaths are often not confirmed virologically or listed on hospital death certificates [12]. Thus, estimates in influenza morbidity and mortality vary by reporting method. In the United States the attack rate ranges from 10%-20% resulting in an estimated 25-50 million yearly

infections [13, 14]. Age-specific differences in the attack rate have been noted with the greatest attack rates among persons < 20 years of age [13, 15]. Hospitalizations are highest among children and are similar to adults aged 50-64 years [16, 17]. Though, the elderly (> 65 years of age) had the highest rates of influenza-associated hospitalizations and this trend has been increasing over the past two decades as a result of the aging population [16]. Pregnant women are at an increased risk of developing cardiopulmonary events as the duration of the pregnancy increases [18].

Analyses of influenza-attributable mortality in the United States have produced a range of estimates. Using an approach based on excess deaths above an epidemic threshold, Simonsen et al. estimated 21,000 influenza deaths per year (1972-1992) [19]. Thompson et al. proposed an annual estimate of 36,000 deaths associated with influenza and further provided age-specific death rates [12]. More recently, Dushoff et al. predicted an annual average of 41,400 deaths attributed to influenza which is consistent with previous estimates and provides additional confirmation that influenza is an important contributor to seasonal excess deaths [20]. Globally, approximately 250,000 to 500,000 deaths are attributed to influenza each year [21].

Cumulative and age-specific mortality rates differ significantly for seasonal and pandemic influenza. The "Spanish flu" of 1918-1919 is widely referenced as one of the most devastating pandemics in history. The United States experienced > 500,000 deaths and globally at least 20 million deaths occurred [22]. Aside from the notable mortality rates, the proportion of excess deaths in persons < 65 years was greater than 99%. The age-related trends continued in the subsequent pandemics of 1957-58 and 1968-69 as younger persons were at a higher risk of death from the pandemic strains [2, 22]. These pandemics also noted significantly lower mortality rates particularly the 1968-69 pandemic which had a mortality rate more consistent

with a typical seasonal epidemic in the United States [23]. Higher attack rates among children < 20 years were observed in both the United States and England for the recent 2009 H1N1 pandemic [24, 25]. Pre-existing immunity might have served as a protective factor among the older age groups during the pandemics [22, 24, 25].

1.3.3 Seasonal Epidemics

Aside from observations of seasonality and the morbidity and mortality burden, seasonal epidemics are often characterized by age-specific patterns, risk factors in transmission, timing and magnitude of epidemics, and differences in the subtypes. The reemergence of A/H1N1 in 1977 has led to the co-circulation of two influenza A viruses (A/H1N1 and A/H3N2) each season for the first time in inter-pandemic history [5]. Previously, the pandemic strain replaced the circulating subtype. As a result, seasonal epidemics are often classified by the predominance of influenza A subtype and influenza B. Differences in spatiotemporal trends, timing, magnitude, and age-related patterns among subtypes and influenza viruses are evident.

Age-specific attack rates exist for both seasonal epidemics and pandemics, and they are driven by factors such as host immunity, virulence, and contact network. Several studies have quantified the attack rates among children and evidence suggests that school-age children are important disseminators of influenza particularly within households [15, 26-28]. Because of the role school children have in propagating influenza epidemics, school closure has received significant attention as a public health intervention [29, 30]. Estimates of transmission are difficult to generate, though an analysis suggests transmission varies by setting with schools and workplaces accounting for 37%, households for 30% and the general community accounting for 33% [31]. Within a household, approximately 40-48% of the secondary cases are attributable to

transmission from a child [26]. Risk factors for the transmission of influenza within households from other than school-aged children include exposure to preschool index patients, exposure to those who attend a child care center, asthma and household density [26, 32, 33]. Households with children less than 5 years of age experience a higher rate of influenza B and rhinovirus infection and influenza-associated hospitalizations than school-aged children [27, 34].

Several differences are noted in the epidemiology of influenza B epidemics compared to influenza A. Influenza B has fewer recognizable cell-surface receptors and genetic reassortment with a novel avian or swine heamagglutinin or neuroaminidase is not likely to occur [1]. As a result, evolutionary changes through antigenic drift are less frequent and opportunities for antigenic shift are non-existent. Since the 1980's, two distinct lineages of influenza B have been co-circulating globally: the Victoria and Yamagata lineages [35, 36]. Seasons dominated by influenza B and A/H1N1 experienced less severe illness and mortality compared to seasons dominated by A/H3N2 [19, 37, 38]. However, since 1980 there have been significantly fewer seasons dominated by influenza B transmission, and it has not occurred since 1992/93 [11](Stark unpublished data). The co-circulation of A/H3N2 and A/H1N1 has led to alternating seasons of subtype dominance which may be the result of epidemic interference or cross-immunity between the subtypes [39]. Within the United States, epidemics dominated by A/H3N2 experience stronger regional synchronization owing to the stronger person to person transmission, opportunities for multiple seeding, or less signal detection in seasons dominated by A/H1N1 and B [38]. Similar inter-hemispheric synchronization results are also observed [39]. In addition, recent findings observe peak timing of B epidemics occurring later in the season than influenza A (H3 and H1) [34, 39].

Seasonality of influenza epidemics during the winter months of temperate climates is an important epidemiological phenomenon that has significant public health implications. Stimuli that lead to a wintertime increase in influenza activity are paramount to understanding seasonal host-pathogen interactions. Distinguishing between climatic factors, host immunity and the school calendar on seasonality is challenging given the similar periodicities [40-42]. However, recent experimental and epidemiologic studies conclude that absolute humidity modulates influenza transmissibility [43, 44]. Furthermore, anomalously low levels of absolute humidity are associated with the onset of influenza activity [43]. This climatic stimulus does not trigger the reintroduction of a latent viral population among host populations resulting in the next season's epidemic. Phylogenetic analysis reveals inter-hemispheric global migration of influenza A viruses contributes to the introduction of seasonal epidemics [45]. These results suggest that geographic importation of genetically different viral strains is responsible for new epidemics rather than natural selection of local strains that persist during non-epidemic periods [45, 46]. These results highlight the importance of year round surveillance to monitor the antigenic characteristics of influenza A specifically in East-Southeast Asia where continuous circulation of A/H3N2 serves as a reservoir seeding epidemics in temperate climates [47].

1.3.4 Spatial and Temporal Dynamics

Despite the regularity of seasonal epidemics, only a few studies have empirically investigated the spatio-temporal patterns of influenza epidemics. Over a 7 year period from 1992-1999, Sakai et al. observed an epidemic spread of influenza in concentric circles from western-central Japan to eastern Japan [48]. The timing of peak influenza-like illness (ILI) activity varied by influenza type; new antigenic variants of A/H3N2 demonstrated a rapid and homogenous spreading pattern

over a shorter time period. Similarly, spatial and temporal clusters of influenza incidence existed in the large urban areas of the Fukukoa prefecture in Japan and gradually diffused to the rural areas coinciding with the major transportation networks in that region [49]. Paget et al. observed a west to east spread of influenza in 4 consecutive seasons across 20 countries in Europe from 2001-2005 and a North-South movement in 3 non-consecutive seasons [50].

An alternative approach to evaluating the spatiotemporal patterns involves estimating spatial synchrony. Synchrony, referred to as how the timing and amplitude in incidence between two series covaries in geographical space, has been used previously to describe the host-pathogen relationship in measles and other ecological systems [51, 52]. Synchrony typically declines with distance, thus nearby populations often experience synchrony. While a ubiquitous presence in populations dynamics, changes in disease transmission from mass vaccination desynchronized measles epidemics [53, 54]. Greene et al. assessed space-time patterns of influenza mortality among the elderly in the United States and observed clustering of influenza in states that shared borders [38]. In addition, the investigators demonstrated that average synchrony declined over shorter distances in seasons when A/H1N1 or B dominated compared to A/H3N2. Evaluation of three Italian regions also displayed a high level of synchrony [55]. Similar excess death estimates in each region irrespective of subtype were highly correlated.

Ecological mechanisms including host dispersal, community interaction processes (predator-prey), and climatic factors can induce regional synchrony [53]. An investigation of influenza spread and human movement revealed a significant association between human movements with epidemic synchrony. The investigators observed that the rates of adults commuting to and from work demonstrated the strongest correlation with mortality from influenza [56]. Furthermore within the United States, a spatial trend in peak timing from Western

States to New England was observed [57]. In this analysis, the Western states including California, Nevada, and Utah peaked approximately two to three weeks earlier than the New England states. Alternatively, environmental forces such as temperature and humidity could have a direct influence on virus and host susceptibility; thus modulating annual oscillations and disease spread [42-44, 58]. In accounting for a seasonal travelling wave in Brazil originating at the Equator in the North and extending towards the Southern regions, Alonso noted that the direction of the seasonal spread may be independent of population factors and could be attributed to climatic forces relevant in the geographically diverse country [59]. The authors suggested that a co-dependence on population movement and the environment could be the driving force behind the travelling wave.

Simulation modeling has provided valuable insights into the spatial spread of influenza. Many simulations have been extended to pandemics and have incorporated international air travel to simulate the global spread of disease [31, 60-63]. These sophisticated models can estimate the effectiveness of prevention and control strategies while modifying the parameter assumptions to evaluate a range of conditions. Additional modeling strategies featured weekly disease counts to derive how an epidemic behaves in space and time. Bonabeau et al. used weekly influenza case counts from a network of general practitioners in France to model the homogenous diffusion of disease spread throughout the country [64]. Mugglin et al. developed a Bayesian hierarchical model to understand the evolution of an epidemic in space and time from an influenza epidemic in Scotland in 1989 [65].

1.4 SURVEILLANCE

1.4.1 Systems

The collection of structured data through routine reporting provides the foundation of influenza surveillance systems. In a period heightened by the threat of emerging pathogens particularly pandemic influenza, surveillance is critical. Current influenza surveillance systems exist across a wide scale of administrative and political boundaries. Global influenza surveillance centers are the focus of the World Health Organization's (WHO) Influenza Surveillance Network. This network routinely surveys samples throughout the year from over 175,000 patients located in 105 countries [66]. In doing so, the network serves as an alert system in the identification of new strains and provides information for annual influenza vaccine strain selection [47, 67]. Additional global networks including the European Influenza Surveillance Network and the United States military's global laboratory-based network also play critical roles in monitoring the epidemiology of influenza and emerging diseases [50, 68, 69].

The United States, specifically the Centers for Disease Control and Prevention (CDC), participates in the WHO's network as a collaborating center designed to provide surveillance data and serve as a testing center for the global samples. The CDC organizes a comprehensive surveillance program featuring multiple data sources designed to address several goals including: timing of influenza activity, defining circulating subtypes, track influenza-like illness, hospitalizations and mortality [70]. To achieve these objectives the CDC collects data from an outpatient illness network, the National Respiratory and Enteric Virus Surveillance System (NREVSS) laboratories, metropolitan vital statistics offices, hospital networks, and State and Territorial Epidemiologists Reports from all 50 states and multiple territories [11].

Pennsylvania participates in many of these surveillance activities for the CDC. The state health department coordinates influenza surveillance by focusing on passive acceptance of case reports, school absenteeism data, influenza-like illness reporting, and syndromic surveillance of hospital emergency rooms. In 2003, Pennsylvania instituted an electronic surveillance system (PANEDSS) as a mechanism to improve efficiency and reporting of influenza cases from hospitals, clinics, and laboratories [71]. This system was the first of its kind and is used in conjunction with syndromic surveillance of respiratory illnesses at hospital emergency rooms to make informed policy decisions on the geographic and temporal attributes of seasonal epidemics.

Syndromic surveillance, including the system utilized by the Pennsylvania Department of Health, gained functionality and acceptance following the events of September 11, 2001. In response to concerns over bioterrorism, surveillance systems were designed to track chief complaints in hospital emergency room visits, pharmacy sales (including prescription and over the counter medications and medical supplies such as thermometers), and employee absenteeism [72, 73]. These systems are beneficial for detecting increases in viral illnesses and to compliment traditional surveillance systems used by state, county or city health departments.

1.4.2 Limitations

Development and implementation of surveillance systems specific to influenza present unique challenges to local, national, and global agencies. Challenges arise in providing timely, accurate, and uniform reporting. Typically, local and state influenza surveillance systems monitor illness during the epidemic season (October-April). However, the emergence of H1N1 pandemic flu in 2009 provided the impetus for local and state health departments to extend surveillance year round as a means to observe unexpected outbreaks and improve our knowledge on the annual

burden of disease. The development of syndromic surveillance systems supports year round surveillance. Moreover, passive surveillance systems are dependent on those with illness to seek care as the mechanism for entry into the system; thus individuals with severe illness or access to health care are the persons who often seek medical attention. Additionally, surveillance systems that do not require virological confirmation of disease or rely on tests with poor sensitivity will not capture the true measure of influenza in the population and may describe other respiratory infections as well. During periods of low influenza prevalence these limitations will underestimate the true incidence of influenza. Nonetheless, these systems provide valuable data on disease incidence and further our knowledge of the epidemiologic profile of influenza.

1.5 CONTROL AND PREVENTION

1.5.1 Vaccines

The significant health burden posed by influenza has led many public health agencies worldwide to recommend vaccination as the leading preventive strategy. Each year, the Food and Drug Administration's Vaccine and Related Biological Products Advisory Committee in consultation with the World Health Organization and the US Centers for Disease Control and Prevention use global surveillance data to predict which influenza strains will be circulating the following year in the Northern Hemisphere. A recommendation for inclusion into the vaccine contains two representative viruses from influenza A (H3N2 and H1N1) and one influenza B variant. Accurately forecasting future viral circulation is challenging and vaccine effectiveness declines when the vaccine poorly matches the circulating strain. Antigenic mismatch with the influenza B

variant has occurred in 5 of the past 10 influenza seasons and notably with influenza A in 2003/04 [36, 74]. Additional concerns regarding vaccines focus on production, contra-indications and achieving significant coverage. Despite these challenges, efforts continue to produce safer vaccines and improved immunization strategies in order to reduce the burden of disease.

Two methods of influenza vaccination are currently approved in the United States: an inactivated subunit vaccine given intramuscularly and a live-attenuated vaccine with intranasal administration. Each vaccine induces a different immune response, affords varying levels of protection, recommended for different age groups, and presents unique challenges. The trivalent inactivated vaccine (TIV) elicits a serum immune response through the production of anti-heamagglutinin and anti-neuradminidase antibodies [75]. Though this vaccine is produced in eggs, thus contraindicated to individuals with an egg allergy, it is well-tolerated and the efficacy is 60-90% in children and adults and the 2010-2011 season saw the introduction of higher dose vaccine ("Fluzone Higher Dose") specific for persons 65 and older [1, 75, 76]. Vaccine effectiveness, defined as the reduction in rate of clinical illness, varies by outcome measure and ranges from a 30-70% reduction [3, 75, 76]. The inactivated design is licensed for adults and children > 6 months and there are multiple manufacturers. Several important distinctions between the inactivated vaccine and live attenuated influenza virus vaccine (LAIV) concern the age of indication and the immunogenicity. LAIV induces mucosal antibodies and a cell-mediated response and thus may elicit a broader immune response than the inactivated type. And while it is only indicated for healthy individuals 2-49 years, the efficacy among children is significantly higher than TIV [75, 76]. A reduction of clinical illness among adults was also evident [77]. The

LAIV vaccine represents a new development in influenza vaccine development and additional technological advances such as cell-culture vaccines, quadrivalent (two influenza B variants), and other adjuvant-based delivery systems are in development [3, 36].

1.5.2 Anti-virals

Two classes of anti-viral drugs exist and have been approved for use in the United States: M2 ion channel inhibitors, amantadine and rimantadine and the neuraminidase inhibitors, zanamivir and oseltamivir. The M2 ion channel inhibitors are specific for influenza A viruses, and thus, do not provide protection from influenza B viruses. Both classes are effective in reducing the duration of symptoms and can be used for prophylaxis; although widespread resistance to the M2 ion channel inhibitors among A/H3N2 has decreased their usefulness as a prophylactic drug [78, 79]. Both neuraminidase inhibitors have proven efficacious in the prophylaxis of seasonal influenza in exposed adults and children [78]. In examining clinical endpoints such as hospitalization, symptoms, otitis media and pneumonia, oseltamivir has also proven effective in reducing rates [3, 78].

1.5.3 Non-pharmaceutical Interventions

The practice of non-pharmaceutical interventions as a method to interrupt influenza transmission has gained visibility in recent years. Concerned with vaccine and anti-viral shortages during a pandemic, public health officials have developed, evaluated, and implemented non-pharmaceutical interventions as an additional approach to blunt seasonal epidemics or a pandemic. These interventions have focused on infection control methods such as isolation and

quarantine, hand hygiene and use of face masks, and social distancing through school and work place closures [80-82]. The effect of these approaches have been evaluated through both empirical and simulation analysis. Observational studies point to mixed success. A review of hand washing studies showed significant reduction in respiratory illnesses among children and adults [83]. In addition, acceptance of these practices among school children has been well received [84, 85]. A clinical trial of hand washing and face mask use demonstrated only moderate success in reducing the secondary attack rate [86]. A multi-layered intervention demonstrated the ability to reduce influenza A infection, though not influenza B among school children [87]. Simulations of multi-layered interventions have proven to be more effective than individual strategies [31, 60, 88]. Nevertheless, the ease of application, minimal cost, success in educating others in the practice and demonstrated benefit in reducing transmission of respiratory disease makes non-pharmaceutical interventions a practical yet potent public health prevention method.

In summary, influenza is one of the most widely researched infectious diseases. Within the past several years, advances in phylodynamics, application of synchrony techniques, and evaluation of climatic factors has provided insights into global migration patterns, mechanisms of spread, and seasonality. Despite these advances, several questions remain such as: do the spatial temporal dynamics at narrow spatial scales correspond to established patterns at larger scales and what is the mechanism of spread between smaller spatial units such as counties. Answering these questions will provide a more complete picture of seasonal influenza patterns across multiple geographic scales.

2.0 MANUSCRIPT 1: LOCAL SPATIAL AND TEMPORAL PROCESSES OF INFLUENZA IN PENNSYLVANIA, USA: 2003-2009

Manuscript in preparation

James H. Stark[1], Ravi Sharma[2], Stephen Ostroff[3], Derek A.T. Cummings[4], Bard Ermentrout[5], Samuel Stebbins[1], Donald S. Burke[1], Stephen R. Wisniewski[1]

[1] Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh PA, [2] Department of Behavioral and Community Health Sciences, Graduate School of Public Health, University of Pittsburgh, Pittsburgh PA, [3] Bureau of Epidemiology, Pennsylvania Department of Health, Harrisburg, PA, [4] Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, [5] Department of Mathematics, School of Arts and Sciences, University of Pittsburgh, Pittsburgh, PA

2.1 ABSTRACT

Influenza is a contagious respiratory disease responsible for annual seasonal epidemics in temperate climates. An understanding of how influenza spreads geographically and temporally within regions could result in improved public health prevention programs. We evaluated the spatial and temporal patterns of laboratory-confirmed influenza cases in Pennsylvania, United States from six influenza seasons (2003-2009). Using a test of spatial autocorrelation, local clusters of elevated risk were identified in the South Central region of the state. Multivariable logistic regression indicated that lower monthly precipitation levels during the influenza season (OR = 0.52, P = 0.0319), fewer residents over age 64 (OR = 0.27, p = 0.01) and fewer residents with more than a high school education (OR = 0.76, P = 0.0148) were significantly associated with membership in this cluster. In addition, time series analysis revealed a temporal lag in the peak timing of the influenza B epidemic compared to the influenza A epidemic. Further examination of the regional transmission dynamics within these clusters may be useful in planning public health influenza prevention programs.

2.2 INTRODUCTION

Each year significant resources are expended by public health officials and health care providers to prevent and mitigate influenza epidemics. Decisions on how to allocate resources for

prevention programs and vaccination campaigns often rely on macro-level information and recommendations without regard to spatially and temporally explicit illness patterns. Knowledge of local geographic spread and variation would likely improve the ability of public health agencies to allocate human and material resources and allow improved targeting and timing of prevention and control measures.

Despite the need for community-based influenza analyses, few studies have explored the spatial and temporal dynamics of incidence on a narrow geographic scale (state or county) appropriate to inform local public health officials [34, 49, 89]. An analysis of influenza hospitalizations in Colorado, United States, noted differences in regional peak timing, influenza B temporality, and age group-specific rates for influenza B hospitalizations [34]. Crighton et al. noted spatial heterogeneity in pneumonia and influenza hospitalization rates within urban and rural counties across age groups in Ontario, Canada [89]. These analyses help to explain the regional spatiotemporal patterns of influenza within a state or province; however, hospitalization data used for these analyses often represents estimates of severe morbidity and may not accurately reflect timing of either peak influenza activity or the true incidence patterns.

Further evaluations of seasonal transmission dynamics have concentrated on broad geographic scales such as a country or continent, often using data aggregated at larger spatial scales [38, 39, 48, 50, 56, 59, 65, 90]. Analyses conducted at smaller spatial scales may capture unique local trends in disease structure potentially concealed in analyses of data aggregated at large scales. The details of local spatial dynamics may reveal the effect of population structure or environmental factors on influenza incidence.

In 2003, a new Pennsylvania law led to mandatory influenza case reporting from all laboratories, providers and hospitals resulting in a detailed spatio-temporal data source not

previously available. As a result, a new opportunity exists to assess the local trends in disease. We conducted an exploratory ecological study evaluating the spatial and temporal patterns of laboratory-confirmed influenza cases in Pennsylvania from six consecutive influenza seasons (2003-2009) using Pennsylvania's National Electronic Disease Surveillance System (PA-NEDSS). Specifically, we assessed spatial incidence clusters, predictors, and temporal variation. Pennsylvania's diverse geography and population structure make it a unique locale to evaluate these dynamics.

2.3 METHODS

Laboratory-confirmed cases of influenza from 2003-2009 were obtained from Pennsylvania's National Electronic Disease Surveillance System (PA-NEDSS) managed by the Pennsylvania Department of Health [71]. The Pennsylvania National Electronic Disease Surveillance System is used to conduct surveillance of reportable diseases including influenza. The passive surveillance system began in 2003 and the system accepts PCR, culture and antigen tests from laboratories, hospitals, clinics, and individual providers in the form of online, electronic, paper or phone reports. Case reports are sent to NEDSS on average 5 days post-specimen collection date. The primary variables extracted from the database for this report included temporal attributes (sample specimen collection date, sample NEDSS report date), spatial attributes (subject home address latitude, and longitude, and zip code), influenza type, gender, reporting method, and date of birth.

For each season, the influenza season defined by the surveillance system ranged from October 1 through April 30 of the subsequent year. Cases were aggregated by week beginning

with October 1 and each subsequent 7 days formed the next week. Specimen collection date was considered the date of diagnosis and used for all temporal and spatial analyses. If this date was not available (13% missing dates), a multiple imputation method used a Poisson regression to model difference between the specimen collection date and the NEDSS report data (100% complete data). Variables considered to be associated with incomplete reporting were included as covariates for the model (county, report method, season). To determine whether the cases with missing dates displayed spatial and temporal biases, a sensitivity analysis using a reduced data set of only cases with complete temporal properties was performed for all analyses.

The cumulative incidence for all six seasons was compared across counties. The total population of each county derived from annual population estimates of the US census served as the denominator [91]. For the presentation and spatial autocorrelation of the cumulative incidence by county, an Empirical Bayesian smoother was implemented to adjust for the inherent variance instability of the small incidence estimates given the small populations at risk [92, 93].

To assess differences in the duration of epidemics, a Gaussian distribution was fit to each epidemic using a non-linear least squares regression. Estimates of sigma (the width of the peak of the epidemic) for each epidemic were compared with 95% confidence intervals from each season.

Global spatial autocorrelation of the 6 year cumulative incidence was estimated by Moran's *I* statistic. This measure detects departures from spatial randomness; thus, a significant positive value would suggest that neighboring counties have statistically significantly more similar incidence than would be found among randomly selected pairs of counties. A significant negative statistic would indicate that neighboring counties have different incidence. Because the Moran's *I* statistic is a global test of spatial autocorrelation, the local indicator of spatial

association (LISA) was used to detect local spatial clusters. Similar to the Moran's *I* statistic, the Local Moran statistic derives an estimate of significance based on a Monte Carlo permutation of the observations. The result is a thematic map which identifies the type of local clustering. Regions designated high-high or low-low indicated clustering of similar values; whereas, regions of high-low or low-high indicated a county was an outlier in the cumulative incidence relative to the neighboring counties [93, 94].

To identify predictors of an elevated incidence cluster from the LISA cluster analysis, a logistic regression modeled a binary outcome which was 1 if counties were in the high incidence cluster (N=8) or 0 if not (N=59). Each covariate was included separately in the model. A stepwise selection approach was used to identify significant predictors in the multivariable model. Goodness of fit for the multivariable model was assessed using Akaike's Information Criteria. All p-values were two-sided based on a 95% significance level.

Covariates selected for the model reflected three broad categories: socio-demographics, health indicators, and the environment. Each variable has either previously displayed an association with influenza incidence and seasonality or could be a confounder in the relationship between spatial heterogeneity and the observed incidence [26, 33, 56, 59]. Social and demographic variable data obtained from the US Census included: age (proportion greater than 64), education (proportion greater than high school), race (proportion white), household income, population density (per square mile), and housing density (per square mile) [91]. Additional demographic variables summarizing the transportation networks in the region include highway miles (linear miles/total county area square miles), and total road miles (linear miles / total county area square miles) [95]. The health indicator variables obtained from the Area Resource File included county level data of active physicians (3 year mean 2005-2007/1000 persons),

hospitals and rural health clinics (4 year mean 2003-2006 [Hospitals] + 5 year mean 2003-2007 [Rural Health Clinics] /1000 persons), proportion pneumonia and influenza mortality (2003-2005 mean/population), and proportion chronic lower respiratory disease mortality (2003-2005 mean/population), also referred to as chronic obstructive pulmonary disease [96]. Distribution of influenza-like illness sentinel physicians (ILINet) and mean number of specimen submissions by provider were summarized for each county and included as a covariate (Owen Simwale, Pennsylvania Department of Health, Personal Communication). Climatic variables including precipitation (per 10 inches), temperature (in Celsius degrees) and dew point data were obtained for the study period (October – April) of each year and averaged over the time period [PRISM Climate Group, Oregon State University, http://www.prismclimate.org, created 4 Feb 2004]. Absolute humidity was calculated by converting the dew point temperature to vapor pressure and then divided by temperature multiplied by the gas constant for water vapor. Mean elevation (feet) was summarized for each county [97-99].

Statistical analyses were performed using the R statistical package (R Foundation for Statistical Computing, Vienna, Austria). Smoothing, and spatial autocorrelation were performed in STIS, (TerraSeer Inc., Crystal Lake, IL), and GeoDa (University of Illinois Urbana-Champaign, Urbana, IL). Institutional review board approval was obtained from the Pennsylvania Department of Health.

2.4 RESULTS

All 67 counties in Pennsylvania reported at least one case of laboratory-confirmed influenza over the six year period and a total of 57598 cases were reported to the Pennsylvania Department of

Health during the study period (Table 1). The greatest number of reported cases occurred during the 2007/08 influenza season; while the 2006/07 season reported the fewest. Co-circulation of influenza A and B occurred during all 6 seasons; however in 2003/04, the percentage of reported typed viruses that were B was approximately 1%. This is in contrast to the 2008/09 season in which 42% of all typed viruses were B; the most in any of the 6 seasons.

In the time series of reported influenza cases, only the 2003/04 season peaked prior to January 1 (Figure 1). Each of the consecutive seasons peaked post-January 1 and the 2007/08 season had the greatest weekly magnitude. The 2006/07 season exhibited the latest weekly peaks. Season 2003/04 experienced the shortest peak epidemic length (2.33 weeks) which was significantly shorter than the other 4 seasons (Table 2). Seasons 2004/05, 2007/08, and 2008/09 had confidence intervals and point estimates that overlapped indicating that durations were not different. A negative correlation existed between reported epidemic size and duration of an epidemic (ρ=-0.49); however, this effect was not significant (P = 0.3231).

Evaluation of the time series stratified by influenza type yielded two important observations reflecting the subtype epidemics (Figure 2). First, peak incidence of influenza B epidemics lagged influenza A epidemics by approximately 3 weeks (mean = 2.75). Second, the decline in weekly cases coincided for both influenza A and B time series in each of the seasons reporting significant influenza B cases even as surveillance systems were maintained.

The Empirical Bayes smoothed cumulative incidence for the seasonal spatial distributions revealed clusters of elevated incidence in the Central and Northwestern portions of the state (Figure 4). The Southeastern and Northeastern regions of the state experienced consistently lower incidence for each season. The Moran's *I* statistic testing for global spatial autocorrelation of the cumulative incidence was 0.4959 (P = 0.07) indicating that neighboring counties have

similar incidence, although not statistically significantly. In the local autocorrelation analysis, the central portion of the state was designated as “high-high” indicating clusters of similar elevated incidence (Figure 5). These counties included: Bedford, Centre, Fulton, Huntingdon, Juniata, Mifflin, Snyder, and Union. The areas of Philadelphia and Delaware counties and the Northeastern region were designated as “low-low” indicating these counties had local correlation of a lower incidence. Analysis of individual seasons demonstrated similar patterns (data not shown but available from lead author).

Descriptive statistics and results of the generalized linear model evaluating the relationship between membership in the elevated cluster and the predictor variables were presented in Table 3. The bivariate logistic regression found education > high school, age > 64, total miles within the county, number of physicians, clinics, and hospitals, the rate of chronic lower respiratory disease, and precipitation associated with membership in the cluster ($P < 0.05$). When including all predictors in a multivariable model, only mean monthly precipitation, age > 64 and education > high school remained significant ($P < 0.05$) (Table 4). For a one percent increase in the proportion of individuals aged over 64, the odds of membership in the cluster decreased adjusting for the other variables in the model (OR=0.27, CI= 0.10, 0.73). Similarly the odds of membership in the cluster decreased for a percent increase in the proportion of individuals with more than a high school decgree (OR= 0.76, CI=0.61, 0.95). An inch increase in monthly precipitation results in a 48% decrease in membership of the cluster (OR=0.52, CI=0.28, 0.94).

A sensitivity analysis using the reduced data set, consisting of only cases with a collection date (N=50421) was performed to assess whether the cases with missing dates displayed spatial and temporal biases. The sensitivity analyses reported limited differences in the spatial and temporal entities and did not influence membership in the cluster.

2.5 DISCUSSION

This was the first study to evaluate the spatial and temporal patterns of laboratory-confirmed influenza cases at the county level within a single state. There was evidence of spatial heterogeneity in the distribution of influenza in Pennsylvania. Using a test of spatial autocorrelation, local clusters of elevated incidence existed from Centre County in the central portion of the state extending to the Southern border counties of Fulton and Bedford. The extent of these elevated risks in this region persisted in each season. A combination of both demographic (age and education) and climatic variables (monthly precipitation) were significantly associated with membership in the elevated incidence cluster. Additionally, this study confirmed a previous finding that influenza B epidemics occur later in the season than influenza A [34, 39].

Time series analysis of weekly influenza surveillance identified by the World Health Organization and National Respiratory and Enteric Virus Surveillance System (WHO/NREVSS) collaborating laboratories for the entire United States and the Mid-Atlantic region (New York, New Jersey, Pennsylvania) showed similar timing of influenza A peaks compared to the PA-NEDSS data for most seasons under study [11]. Coinciding epidemic fade outs of influenza A and B were observed on a national level and within the Mid-Atlantic region from recent seasons:

2005/06 through 2008/09 (data not shown). Other regions of the country observed similar patterns of simultaneous declines. The concurrent weekly decline in reported cases for Pennsylvania may be the result of several factors including environmental drivers, host factors, diminished surveillance, and a small sample size. Changes in temperature and humidity as the winter shifts to spring may alter virus stability and influence patterns of crowding and host mixing leading to a simultaneous decline in incidence [42, 59]. Alternatively, diminished state surveillance as providers stop collecting and submitting specimens for influenza testing can lead to unreliable case estimates at the end of an epidemic producing an artifactual constraint on the epidemic time series. Seasonal time series encompassing longer surveillance periods are needed to control for the confounding effects of time in order to validate these findings.

This study is consistent with previous findings that the influenza B epidemic typically occurs later in the season than the influenza A epidemic. Finkelman et al. aggregated weekly incidence values over a nine year study period and demonstrated that influenza B temporally lags both the A/H3 and A/H1 subtypes in the Northern Hemisphere [39]. The degree of temporal similarity in peak epidemic timing of influenza A and B across the geographic scales (counties and continents) suggests that the factors driving the timing of the subtype epidemics could be similar within the Northern Hemisphere.

Estimates of the epidemic widths showed similarities to the peak durations observed among larger seasonal epidemics in Japan [48]. However, the range of epidemic lengths from the Pennsylvania results (2.33-5.9) did not correlate to epidemic magnitude as noted in the seasonal epidemics from Japan. The non-significant correlation could be attributed to less variation in epidemic size or having a smaller data set. Differences in circulating influenza subtypes, particularly the introduction of new A/H3N2 antigenic variants in the Japan epidemics resulted

in shorter peak activity periods [48]. This result was in contrast to seasons without new variants leading to epidemics that were smaller and displayed longer periods to attain peak activity. Without data specific to Pennsylvania, reviewing the strain-specific information for the United States shows that the 2003/04, 2004/05, 2005/06, and 2007/08 seasons were dominated by A/H3N2. In 2003/04 the A/Fujian/411/2002 A/H3N2 virus predominated and accounted for 88.8% of A/H3N2 isolates characterized leading to a less than optimal vaccine match [100]. Both the 2004/05 and 2007/08 seasons reported new A/H3N2 variants (A/California/7/2004-like and A/Brisbane/10/2007-like) and these epidemics reported peak durations of at least 1 week shorter than the 2005/06 season [101, 102]. In 2008/09, approximately 42% of all Pennsylvania cases were antigenically characterized by influenza B viruses. Nationally, influenza A cases were predominated by A/H1N1 (pre-novel H1N1) [103]. The 2008/09 season was not dominated by a new A/H3N2 variant, yet the epidemic length observed in this study from 2008/09 is not significantly different than the results from 2004/05 and 2007/08 when a new A/H3N2 antigenic variant appeared. In Pennsylvania, the first identified illness due to 2009 pandemic influenza A/H1N1 virus did not occur until the end of April and its appearance does not impact the data in this analysis. During the 2008/09 season the circulating A/H1N1 viruses were related to the vaccine component while less than 49% of the circulating influenza B viruses were related to the vaccine strain. Similar to the 2003/04 season, the vaccine mismatch among the influenza B virus may have contributed to the overall short epidemic duration. Historically, within the United States, seasons dominated by A/H3N2 displayed a higher average synchrony and greater mortality than seasons dominated by A/H1N1 or B; though duration of B epidemics is not known [19, 38]. Nevertheless, the time period under study may not be representative of other influenza seasons and a longer time series is needed to confirm these results.

Discovery of the elevated incidence cluster in the central portion of the state warranted further investigation. The logistic model was designed to assess differences in characteristics for counties within and outside of the cluster with the specific intent of answering the question: what factors can explain the cluster of elevated incidence. Only age, education, and precipitation remained significant in the multivariable model.

The association of both age and education with membership in the cluster may be a reflection of differences in vaccination coverage between the counties. Poor vaccination coverage would create upward pressure on seasonal incidence rates and mortality [75, 104, 105]. Regional vaccination differences have been reported in urban and rural areas, age groups, and with increasing levels of education [106, 107]. According to the Behavioral Risk Factor Surveillance System (BRFSS), vaccination rates among the elderly (Age > 65) in Pennsylvania only recently approached the 70% Healthy People 2010 threshold [108]. The proportion of residents greater than 64 years and with more than a high school education was significantly lower among the counties in the cluster; which may suggest a lower vaccination rate in the cluster. Without available county-explicit data estimating seasonal influenza vaccination coverage, interpretation of the regional trends should proceed with caution.

Environmental factors including temperature and humidity have been long-associated as the driving force in the severity, spread and seasonality of influenza [38, 58, 59, 109, 110]. More recently, experimental and epidemiologic simulation studies have concluded that absolute humidity modulates influenza transmissibility leading to the observed seasonality in temperate climates [111, 112]. This report presented the results of multiple environmental factors including temperature, precipitation, dew point, and absolute humidity. In this study we found a significant relationship with precipitation but not with absolute humidity, nor with any other environmental

variables. The relationship of influenza incidence and precipitation has been inconsistent across studies as the associations tend to differ by country and influenza type [113-116]. Associations of precipitation with the onset of influenza B have been observed, though these associations have not persisted with influenza A. For the climatic variables used in this analysis, the monthly results were averaged over the study period which is in contrast to previous studies that evaluated monthly differences in an effort to estimate the timing of influenza incidence or the onset of the influenza season which may have contributed to the contrasting results. There is no notable spatial correlation structure in the evaluation of influenza A and B in this dataset; thus, these comparisons cannot be made.

A sensitivity analysis using the reduced data set, consisting of only cases with a collection date (N=50421), reported limited differences in the spatial and temporal analyses. Spatial comparisons revealed > 2 fold differences in the cumulative incidence for Juniata, and Mifflin counties. These relative differences were to be expected as these three counties reported > 50% of cases without a collection date. Influenza did occur in these regions; however, without implementing the missing data imputation, the analysis would have erroneously ignored actual cases. The local indicators of spatial autocorrelation analysis (LISA) observed a similar pattern of spatial auto correlation in the center of the state except for Juniata, Snyder, and Union counties (data not shown). Similarities in the local spatial autocorrelations among the imputed and reduced data sets point to the strength of these results that the central region of the state is subject to an elevated risk of influenza.

The passive surveillance system of PA-NEDSS creates reporting limitations. Even though Pennsylvania law mandates physicians, providers, hospitals, and laboratories to report specific disease data to PA-NEDSS, significant non-compliance has resulted in several types of

ascertainment biases [117]. First, the expected annual number of incident cases in the United States is estimated between 10%-20% which is substantially higher than the reported number of cases to PA-NEDSS [13, 118]. Many cases of influenza go undetected because the patient fails to seek treatment or is not tested for the disease. Moreover, seasonal variation also exists in the reporting due to severity of disease. A severe influenza season is likely to result in more symptomatic cases and more cases seeking treatment. The consequence of an increased case load also creates a burden on reporting for the provider or laboratory potentially resulting in fewer cases reported to PA-NEDSS. In addition, there is concern over the methods of data collection and submission from Philadelphia County which are not consistent with the remainder of the state. A sensitivity analysis conducted without Philadelphia County did not note any significant changes to these results, thus the extent of the underreporting for these analyses was limited. Finally, spatial differences observed could also have been affected by testing practices of health care providers; those with access to free testing and a greater interest in influenza could result in a surge of testing. Inclusion of variables reflecting spatial location and submission history of influenza-like illness sentinel providers, who have access to free testing, was not associated with the cluster of elevated incidence; thus super testers are not likely to affect the spatial results observed.

The future of longitudinal data analysis within this data system is likely to be affected by the emergence of the H1N1 pandemic influenza subtype. Shifts in age distributions of pneumonia and influenza mortality have been noted in post-pandemic periods, which may have implications for the spatial distributions particularly in regions with younger populations [22]. Furthermore, there may be differences in the transmission parameters of the newly emerged influenza A subtype and the previous A/H1 subtypes in circulation resulting in further

longitudinal distortions of the data. Despite these potential shortcomings, analysis of the transitional pandemic period remains an essential area for further exploration of these specific issues.

In conclusion, the epidemiology of influenza in Pennsylvania can be defined by a distinguishing spatial pattern. County level analysis revealed unique spatial patterns; a strength of this study. State and county public health officials should consider these findings in the utilization of human and economic public health resources to improve control strategies aimed at minimizing transmission through targeted vaccinations, directed hygienic advertisements, and informed surveillance. Additional research should focus on extending the analysis to the states of Maryland, Virginia, and West Virginia to determine if the spatial regime extends beyond the administrative borders.

2.6 TABLES

Table 1. Characteristics of reported influenza cases in Pennsylvania, USA, 2003-2009 influenza seasons

	Influenza season, no. (%)							
Variable	**Cumulative**		**2003-2004**		**2004-2005**		**2005-2006**	
Number of Cases	57598		8836	15.34%	11293	19.61%	8717	15.13%
Flu Type								
A	35307	71.33%	5670	64.17%	8557	75.77%	6547	75.11%
B	8169	16.50%	59	0.67%	1369	12.12%	1692	19.41%
Unknown	6023	12.17%	3107	35.16%	1367	12.10%	477	5.47%
Gender								
Male	23057	46.58%	4151	46.98%	5154	45.64%	4098	47.02%
Female	26395	53.32%	4683	53.00%	6135	54.33%	4616	52.96%
Unknown	47	0.09%	2	0.02%	4	0.04%	2	0.02%
Age**								
Mean	34		31		45		33	
Median	27		19		46		24	
Under 5 years	9396	16.32%	2871	32.50%	1253	11.10%	1338	15.35%
5 to 19 years	14461	25.12%	1626	18.41%	1817	16.09%	2632	30.19%
20 to 44 years	14929	25.93%	1499	16.97%	2483	21.99%	1942	22.28%
45 to 64 years	8479	14.73%	866	9.80%	2216	19.62%	1224	14.04%
65 years and over	10314	17.91%	1972	22.32%	3524	31.21%	1581	18.14%

* Nineteen subjects have missing date of birth

**Number of Cases

Table 1 continued

Variable	Influenza season, no. (%)					
	2006-2007		**2007-2008**		**2008-2009**	
Number of Cases	3997	6.94%	16657	28.92%	8098	14.06%
Flu Type						
A	3264	81.66%	11269	67.65%	4550	56.19%
B	563	14.09%	4486	26.93%	3404	42.04%
Unknown	170	4.25%	902	5.42%	144	1.78%
Gender						
Male	1937	48.46%	7717	46.33%	3881	47.93%
Female	2055	51.41%	8906	53.47%	4196	51.82%
Unknown	5	0.13%	34	0.20%	21	0.26%
Age*						
Mean	27		35		22	
Median	19		31		17	
Under 5 years	684	17.11%	2077	12.48%	1173	14.49%
5 to 19 years	1348	33.73%	3562	21.41%	3476	42.92%
20 to 44 years	1063	26.59%	5520	33.17%	2422	29.91%
45 to 64 years	499	12.48%	2918	17.54%	756	9.34%
65 years and over	403	10.08%	2563	15.40%	271	3.35%

* Nineteen subjects have missing date of birth

Table 2. Epidemic width estimates and confidence intervals

Season	σ^2	95% Confidence Interval
Season 2003/04	2.33	2.26, 2.39
Season 2004/05	3.6	3.2, 4.0
Season 2005/06	4.89	4.58, 5.2
Season 2006/07	5.9	5.19, 6.60
Season 2007/08	3.72	3.56, 3.87
Season 2008/09	3.82	3.64, 4.01

* Sigma measures the epidemic length

Table 3. Descriptive statistics and results of logistic regression model. Dependent variable are counties designated HIGH-HIGH in Moran's LISA cluster analysis, counties = 8.

Variable	Cluster = Yes Mean	Cluster = No Mean	Odds Ratio (OR)	P-value
DEMOGRAPHICS (N=11)				
Household size	2.507	2.473	78.5707898	0.2940
Proportion of families w/ 1 child < 18 years	0.4341	0.4394	0.00043812	0.5880
Proportion of families w/ 1 child < 6 years	0.1745	0.1747	0.67139341	0.9850
Race (proportion white)	0.9647	0.9459	2.13165782	0.4890
Education > high school	0.3038	0.3778	2.6397E-08	**0.0319***
Age > 64	0.1448	0.1642	0.65856	**0.0308***
Household income	35035	37467	0.99994100	0.3910
Population density per square mile	84.09	503.2	0.99133773	0.1010
Housing denisty per square mile	34.77	215	0.97170793	0.0795
Total road miles per area square miles	1.4892	3.1867	0.23015552	**0.0102***
Highway miles per area square miles	0.08902	0.14375	0.00027550	0.1700
HEALTH INDICATORS (N=8)				
Active physicians	71.75	619.7	0.9938191	**0.0144***
Active physicians per 1000 persons	1.13	2.151	0.5231432	0.2010
Rural clinics and hospitals	1.688	4.766	0.56254	**0.0387***
Rural clinics and hospitals per 1000 persons	0.03656	0.04483	0.01993	0.6730
ILI Sentinel Physicians	0.625	0.8305	0.8376960	0.6470
ILI Submissions	0.9023	0.6601	193.05980	0.1990
P&I mortality	2.40E-04	2.67E-04	0.967113	0.5200
Chronic lower respiratory disease	0.0004663	0.0005414	0.904023	**0.0423***
ENVIRONMENT (N=6)				
Elevation	1035.4	1227	0.999300	0.3370
Precipitation	3.174	3.501	0.591656	**0.0097***
Minimum temperature	-1.711	-2.1358	1.209128	0.4438
Maximum temperature	8.963	8.502	1.252322	0.3945
Dew point	-2.172	-2.5419	1.488992	0.322
Absolute Humidity	868.9	848.5	1.00630	0.351

* Significance: P-value < 0.05

Table 4. Results of multivariable logistic regression model

MULTIVARIABLE MODEL		
Variable	Odds Ratio (OR)	**P-value**
Age > 64[†]	0.27	**0.0100***
Education > high school[†]	0.76	**0.0148***
Average precipitation (2003-2009)[‡]	0.52	**0.0319***

***** Significance: P-value < 0.05

[†] Interpreted as a 1% units

[‡] Interpreted as a 1 inch units

2.7 FIGURES

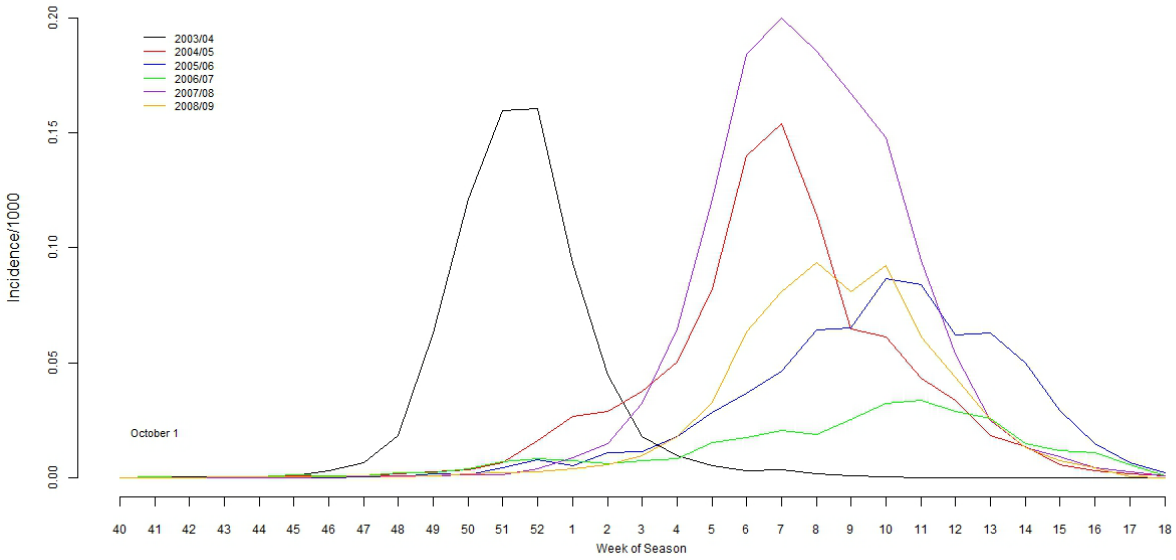

Figure 1. Seasonal influenza incidence in Pennsylvania (sum of all counties), 2003-2009

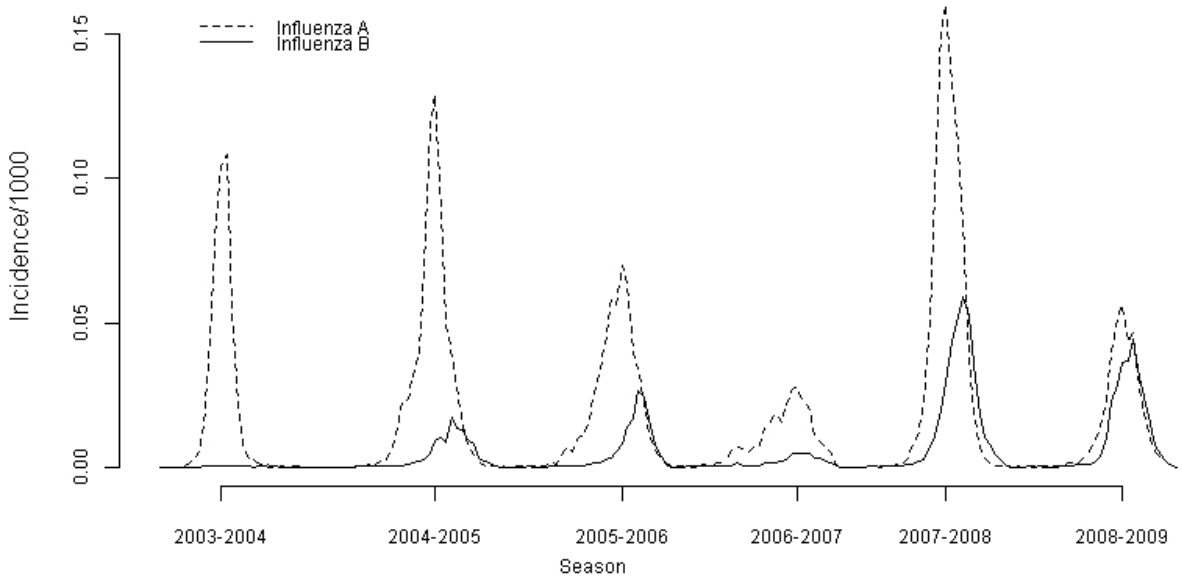

Figure 2. Seasonal influenza incidence by influenza type in Pennsylvania, 2003-2009

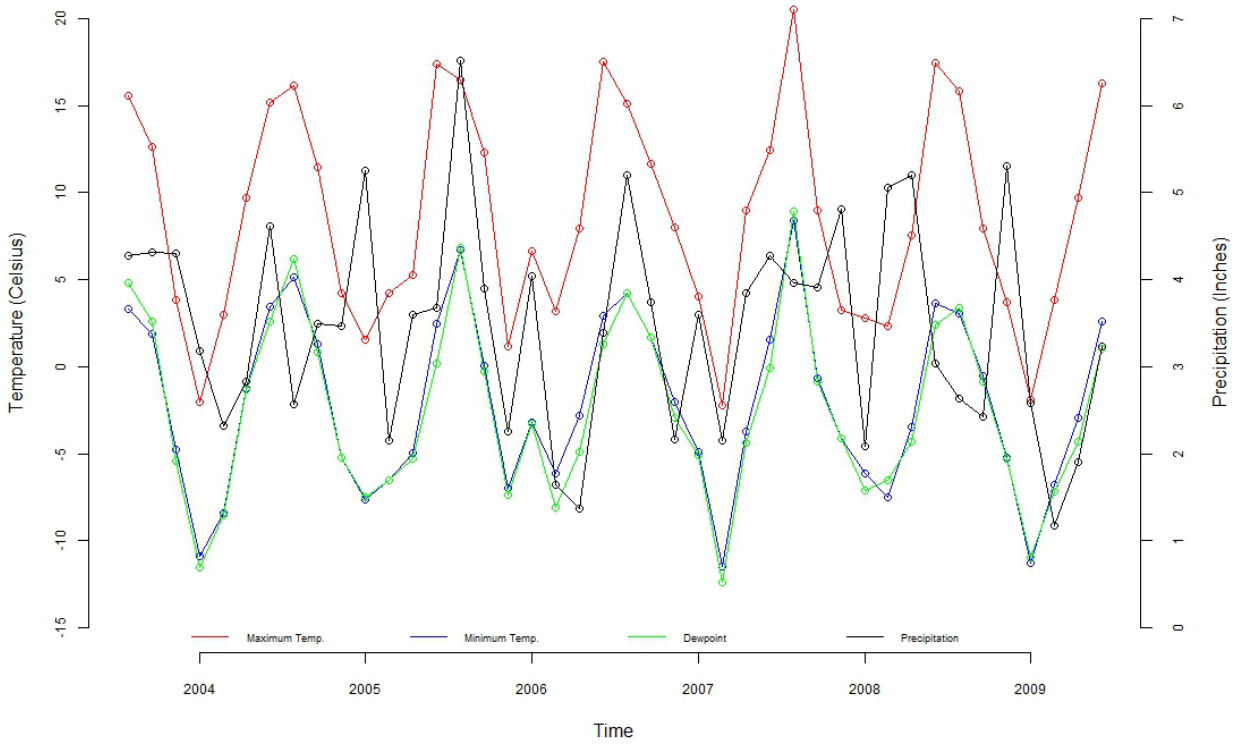

Figure 3. Monthly changes in environmental variables (October-April) averaged across all 67 counties

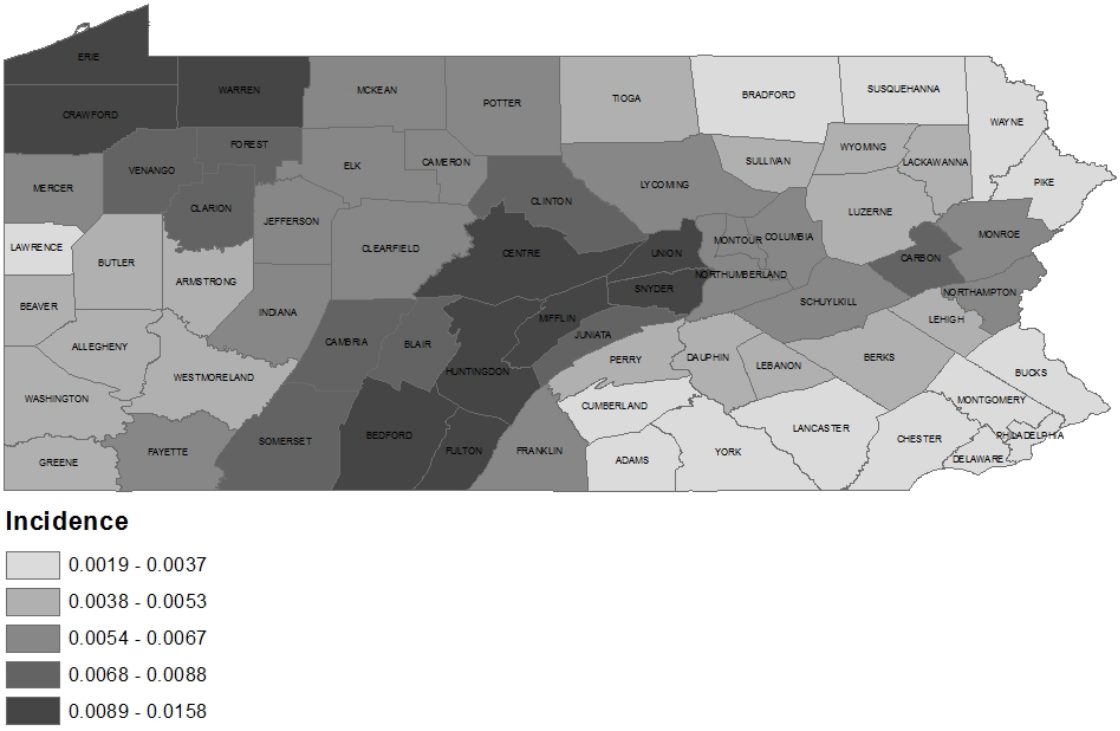

Figure 4. Cumulative incidence of six influenza seasons (2003-2009)

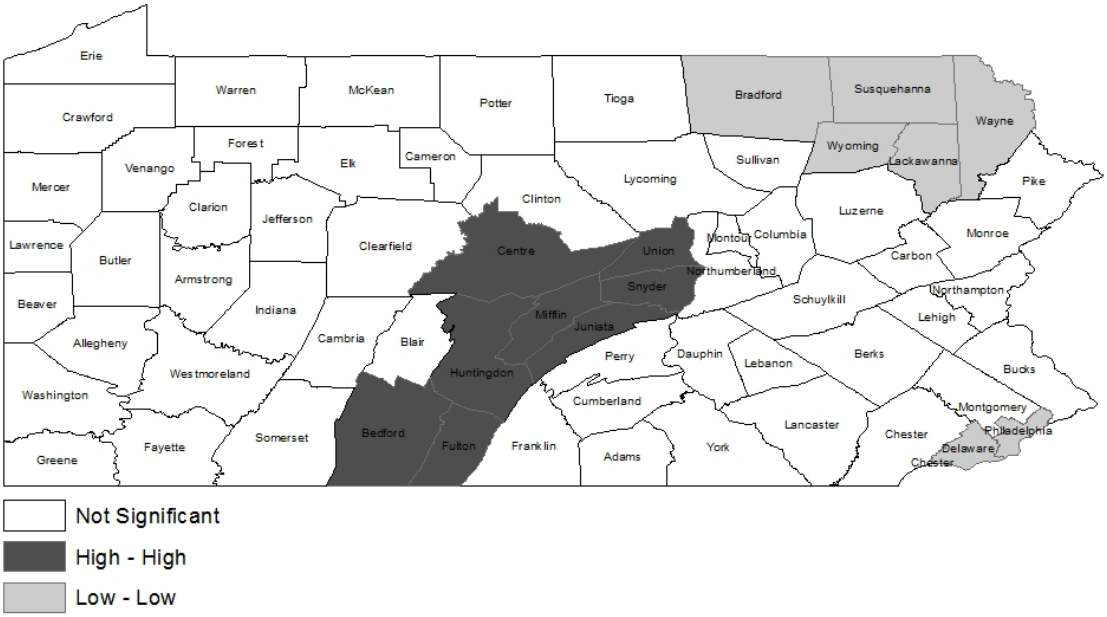

Figure 5. Spatial autocorrelation of 6-year cumulative incidence for 67 counties in Pennsylvania

3.0 MANUSCRIPT 2: LOCAL VARIATIONS IN SPATIAL SYNCHRONY OF INFLUENZA EPIDEMICS

Manuscript in preparation

(Authorship to be determined)

James H. Stark[1]

Donald S. Burke[1], Derek A.T. Cummings[2], Bard Ermentrout[3], Jiawei Huang[4], Stephen Ostroff[5], Ravi Sharma[6], Samuel Stebbins[1], Stephen R. Wisniewski[1]

[1] Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh PA, [2] Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, [3] Department of Mathematics, School of Arts and Sciences, University of Pittsburgh, Pittsburgh, PA, [4] Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh PA, [5] Bureau of Epidemiology, Pennsylvania Department of Health, Harrisburg, PA, [6] Department of Behavioral and Community Health Sciences, Graduate School of Public Health, University of Pittsburgh, Pittsburgh PA

3.1 ABSTRACT

Understanding the mechanism of influenza spread across multiple geographic scales is not complete. While the mechanism of dissemination across regions and states of the United States has been described, understanding the determinants of dissemination between counties has not been performed. The paucity of high resolution spatial-temporal influenza incidence data to evaluate disease structure is often not available. Here, we report on the underlying relationship between the spread of influenza and human movement between counties of one state. Significant synchrony in the timing of epidemics exists across the entire state and decay with distance (regional correlation = 62%). Synchrony as a function of population size display evidence of hierarchical spread with more synchronized epidemics occurring among the most populated counties. A gravity model describing movement between two populations is a stronger predictor of influenza spread than adult movement to and from workplaces suggesting that non-routine and leisure travel drive local epidemics. These findings highlight the complex nature of influenza spread across multiple geographic scales.

3.2 INTRODUCTION

Despite the regularity of influenza epidemics, understanding the nature of influenza spread remains unclear. Inferences reflecting the spatiotemporal patterns of disease spread has been

advanced in recent years through availability of detailed spatial-temporal data and the application of synchrony and time-frequency decomposition methods [51, 119]. Evidence of spatial synchrony and traveling waves has been reported in infectious diseases such as measles and dengue resulting in novel insights into urban and rural infection hierarchies and the impact of spatial heterogeneities of the host population of incidence waves [51, 120, 121]. These approaches have been extended to influenza which has observed population density, human movement, and antigenic dominance as key determinants of influenza spread at the country scale [38, 48, 56, 122, 123].

The current understanding of the intrinsic properties of influenza epidemics is limited by the geographic scales used to evaluate the data. Often the spatial scale of analysis is the continent or country [38, 56, 90]. Analyses conducted at larger spatial scales may potentially conceal local trends in disease structure. High resolution spatial-temporal infection data is often not available. As a result, there are few opportunities to validate findings at large spatial scales with finer spatial scale observations. The mechanism of influenza spread is one such example. Brownstein et al. showed the importance of air travel in the dissemination of influenza cases across census regions in the United States [123]. Viboud et al. used state-specific mortality data to demonstrate the relative importance of workflows compared to distance and other movement metrics in capturing the spatial synchrony of influenza mortality in the United States [56]. While these finding are relevant to understanding the spread of influenza within the United States, confirmation of these results using more spatially refined incidence data would test the consistency of these relationships across a broad geographic spectrum.

Gravity models have been used to explain spatial dynamics of epidemics [56, 124-126]. They were developed in transportation theory to model the flow of travelers across a landscape

[127]. The gravity model describes the magnitude of travel between two locations as a function of the population sizes in the two locations and the distance between those locations. Simulations of a gravity model fitted to workflows captured influenza spread among the states with good agreement in timing and duration of epidemics compared to historical data [56]. Because a gravity model estimates a general pattern of movement without preconditions on type or geographic features of the location, evaluation of a gravity-model may provide insight into local interactions not captured by well-defined mechanisms of travel.

In this report, laboratory confirmed influenza cases from Pennsylvania, United States are used to compare county-specific incidence patterns. As the sixth most populated state in the United States, Pennsylvania is divided among 67 counties, of which two counties, Allegheny and Philadelphia, account for greater than 22% of the state's population. The state is trifurcated by two major interstate highways with limited transportation networks in the northern counties and has international airports on oppos ite ends of the state. With extreme segmentation in the population structure and a divisive transportation network, Pennsylvania is a unique locale to assess the predictors of influenza spread at a local level.

This is the first report to evaluate the underlying relationship of disease spread and human movement using county-specific influenza cases. Estimates of spatial synchrony are evaluated using correlation coefficients and the Mantel statistic to determine whether synchrony is associated with large numbers of adult workflows or gravity-like estimates of interaction. Understanding the mechanism of spread at a fine spatial scale would provide an improved level of understanding not previously available for local county and city public health officials to implement surveillance and response activities.

3.3 METHODS

3.3.1 Data

Weekly estimates of reported influenza cases from 2003-2009 were provided by the Pennsylvania Department of Health. A description of the passive surveillance system employed by the Department of Health has been previously described (Stark Dissertation). Briefly, the Pennsylvania National Electronic Disease Surveillance System (PA-NEDSS) is a computer application used to conduct surveillance of reportable diseases including influenza. Case reports are routinely collected by providers and laboratories and are transmitted electronically to the PA-NEDSS system. The surveillance system defines each influenza season to begin in the 40th week of the calendar year through the last week of April of the following year. Influenza data occurring during this entire time period were used for this analysis. Cases specified to date of specimen collection and home address or zip-code was aggregated by week and to one of 67 Pennsylvania counties, respectively. The total number of reported influenza cases during the study period (2003-2009) was 57598. The US census provided annual population estimates to calculate seasonal incidence for each county [128]. Rates of human work flux data between counties for the year 2000 was obtained from the US Census [129]. The workflow data describes in which county people work and in which county they reside; thus approximations of flow between counties can be calculated.

3.3.2 Sampling

The analysis for this study uses 186 weeks of surveillance data accumulated over 6 seasons (31 weeks/year). Small case counts from counties with a small population may result in increased sampling variability and poor correlation with other counties. Even with uniform reporting efficiencies in each county, we expect small counties to report more weeks with zero cases which may lead to greater variability. Thus, it is difficult to differentiate the effect of population size and reporting error on disease spread. To address these concerns, a sampling method adapted from Grassly et al. was employed to estimate a new correlation matrix with additional sampling error for the largest populated counties described by the binomial distribution [130]. For the 30 counties with the largest population, the reported incidence rates at each time point of these counties were resampled 1,000 times from a binomial distribution with a sample size equal to the remaining 37 counties (randomly sampled with replacement). These counties are designated the large populated counties. The selection of 30 counties for the resampling was based on a natural break in the distribution of populations and reflected a large enough sample to get meaningful results. This provided a new distribution of a time series for the larger counties as if they had sampling error equivalent to the smaller counties (N=37). Using the binomially-generated predicted time series, the mean of 1000 new pair-wise correlation matrices were generated. The mean correlations were ranked and the 25^{th} and 975^{th} values were extracted and compared with the mean correlation of the observed correlation matrices for the 37 smaller counties. A statistically significant difference in the correlation between large and small counties will result if the distribution of binomial sampled correlations excludes the mean correlation of the smaller counties. Thus, a statistically significant result is not likely to reflect differences in sampling error and provide further confidence in the synchrony and correlation analysis.

3.3.3 Synchrony and Mantel Correlation Analysis

Spatial synchrony provides an estimate of the correlation of an epidemic time series across a geographic region [51, 53]. For this analysis, spatial synchrony was measured as the Spearman rank correlation of the pairwise comparisons of weekly cases for each county over the entire study period. Algorithms for the spatial correlation function estimating the relationship between synchrony and Euclidean distance were obtained from the NCF library for R, specifically the non-parametric covariance function [131].

Mantel tests were used to compare the matrix of pair-wise Spearman correlations of influenza time series to matrices describing pair wise county to county human movement, geographic distance, and population size [132, 133]. Thus the Mantel statistics estimates the correlation of the comparative elements between two matrices. A workflow matrix was created composed of the number of individuals who reported commuting from county i to county j in the US Census dataset by summing the movement to and from each county resulting in a symmetric 67x67 matrix. Distances between counties were represented by a Euclidean distance matrix. The population matrix consisting of the product of counties i and j was also tested. Partial Mantel's tests were used to measure the association of two matrices in the presence of a third matrix.

Two Pennsylvania counties may have limited movement between one another but may engage in substantial workflow contact through a third non-Pennsylvania county; thus having an indirect effect on t he epidemic synchrony. In order to explore whether this workflow might explain the pattern of correlations of influenza observed in Pennsylvania, an additional workflow matrix capturing these second-order movements (inter-state) for counties in Border States was created and included in the Mantel tests. This matrix incorporated workflows to and from 302

counties from the six states bordering Pennsylvania (Delaware, Maryland, Ohio, New Jersey, New York, and West Virginia).

We estimated the correlation of gravity flows estimated using a published model with our pair-wise influenza correlation matrix [6]. This gravity model was estimated using national influenza mortality data. We also estimated the parameters of a gravity model that maximized the Mantel correlation coefficient of the gravity flow matrix with the Pennsylvania influenza time series and workflows using Nelder-Mead optimization.

Model 1

$$C_{ij} = \theta \frac{P_i^{\tau_1} P_j^{\tau_2}}{d_{ij}^{\rho}}$$

A gravity model for workflows and disease spread (Cij) is parameterized by the population of counties i and j (Pi, Pj) and the distance between the two counties (Dij). The exponents $\tau 1$, $\tau 2$, and ρ, estimated by the model, quantify the attraction of the receiving and generating counties by population size and the distance between two counties. Theta, θ , is the proportionality constant.

The Mantel test compared the pair-wise Spearman correlations of influenza time series and the gravity matrix. An additional Mantel statistic measured the association of influenza time series and the gravity matrix fitted to national influenza mortality data [56].

Institutional review board approval was obtained from the Pennsylvania Department of Health.

3.4 RESULTS

Of the 186 weeks of influenza data analyzed for each county, the mean number of weeks with at least one case was 76 weeks and the range was 11 weeks (Cameron) to 141 weeks (Allegheny). When comparing the proportion of total weeks reporting a case, counties with large populations reported 70% more weeks with a case than counties with small populations. The partition of counties by population size was determined by a natural break in the data. Additional statistics describing the differences between the large and small population counties were presented in Table 5.

Results of the binomial sampling demonstrated that sampling error has limited effect on the correlation of epidemic time series between counties. The mean correlation of the 1000 pair-wise binomial sampled correlations was 0.692 (95% CI: 0.658, 0.726). The mean Spearman correlation from the correlations of the 30 larger populated counties was 0.76 and 0.54 for the 37 smaller populated counties. Because the confidence interval of binomial sampled correlations excludes the mean correlation of the smaller counties, we concluded that differences in the correlation were more likely to reflect differences in the population structure than in the sampling error of the smaller counties. As a result, we are confident in using the incidence data for all counties to further evaluate estimates of synchrony and the predictors of disease spread.

Estimation of spatial synchrony from all 67 counties used Spearman rank correlations of the epidemic time series and a distance matrix composed of county centroids. Considerable correlation existed across the entire state as the regional correlation was 62% (Figure 6a). Adjacent counties had a high mean correlation of 80%; although, synchrony declined with distance and approached the regional mean correlation at 127km. The lower bound of the 95% confidence interval crosses the regional correlation at 36km. Prior to this distance, the local

synchrony is statistically significantly different than the state correlation. Fewer than 2% of county pairs have county centroids separated by 36km or less, thus the correlation in epidemic time series between neighboring counties was not extensive. The rising, yet not significant, increase in synchrony over distance (U-shaped curve) reflects strong correlation among the larger population regions separated by several hundred kilometers. The U-shaped trend is also noted in Figure 7 and is a unique feature of Pennsylvania's geography.

Measurement of spatial synchrony as a function of population sizes and county workflows also revealed interesting patterns. Synchrony increased as the product of the county population size increased ranging from a median correlation of 0.48 in the smallest quartile to a median correlation of 0.80 in the largest quartile (Figure 6b). A positive, but not significant trend existed for synchrony and county to county workflows (Figure 7a). These county-specific synchrony results were consistent with the observations of distance, population size, and workflow observed by Viboud et al. using state-specific mortality time series [56]. The inter-state workflows consisting of neighboring counties of Pennsylvania also exhibited a positive trend, though less variation between quartiles compared to the intra-state workflows (Figure 7b). Figure 3 describes the three dimensional relationship between workflows, distance, and population size.

When evaluating the yearly synchrony with distance, population, and workflows, similar trends as the cumulative study period (2003-2009) existed for each year (Appendix A). However, the 2006/07 season experienced substantially lower adjacent and statewide correlations and lower correlations for each population and workflow quartile particularly the first quartile (smallest population/workflow counties) which observed a correlation closer to 20%.

The Mantel statistic describing the relationship between the epidemic time series and distance, population, and human movement are presented in Table 6. Only Euclidean distance

was not significantly associated with influenza when evaluating all 67 Pennsylvania counties. The inter-state workflow matrix had a smaller correlation than the intra-state workflows likely implying that work-related movement of individuals from the neighboring counties was not strongly associated with disease spread within Pennsylvania. Many counties in Pennsylvania did not experience work-related movements to all of the border state counties; thereby, necessarily reducing the correlations.

The gravity matrix fitted to Pennsylvania county disease data was the strongest predictor of influenza spread within the state. After adjusting for population size, distance, and workflows, the gravity model remained the strongest predictor of influenza spread. Similar to the 3-dimensional figure of workflows, distance, and population (Figure 8), distance as a function of the gravity model also displays a U-shaped pattern (Figure 9a). Of the county pairs with high gravity values at a distance > 400 km, the majority involved Allegheny County. Allegheny County paired with Berks, Bucks, Chester, Delaware, Lackawanna, Lehigh, Luzerne, Montgomery, Northampton, and Philadelphia County. The remaining set of pairs involved Philadelphia County which paired with Beaver, Butler, Erie, Washington, and Westmoreland County. Only Montgomery and Westmoreland County presented a pair that did not involve Allegheny or Philadelphia at a distance greater than 400 km apart. As expected, these counties are among the largest populated counties. The gravity model fitted to Pennsylvania-specific workflows was not a strong predictor of disease spread (Table 6) and the trend over distance noted in the gravity model fitted to disease data did not materialize (Figure 9b). Furthermore, the small correlation between intra-state workflows and the Pennsylvania gravity model ($\square = 0.39$, $P < 0.001$) indicated that movement within the state was not completely dependent on workflow. A comparison of the parameter estimates fitted by the gravity model is presented in Table 7.

The gravity matrix fitted to the parameters obtained from United States gravity model did not correlate well with disease spread using Pennsylvania's county-specific influenza data. Differences in strength of correlation between gravity matrices may be the result of local variations within Pennsylvania captured more efficiently such as the range of county size and distance. It is important to note that Viboud et al. did not present a correlation of the gravity matrix to the epidemic time series; thus, a direct comparison of the impact of the gravity model on disease spread scaled to the country was not possible [56].

3.5 DISCUSSION

Few studies have explored synchrony of influenza epidemics and the predictors that drive influenza spread. This study further evaluated these quantities though at a finer spatial scale than previously reported. These results demonstrated evidence of spatial-temporal correlation in the incidence of influenza across counties of Pennsylvania. Significant synchrony among neighboring counties existed. A gravity model describing movement between two populations was the best predictor of influenza spread.

Comparison of these results to inter-regional influenza spread and to the United States mortality data analysis could reflect differences in the mechanisms of spread at different geographic scales. Analysis of influenza incidence among the US Census regions demonstrated the importance of air travel in long-range dissemination. While adult workflows effectively captured the spread of influenza across the United States, a gravity model did better at the smaller county to county scale. Interstate commerce and other opportunities for interstate workflows may be responsible for the majority of interactions at these larger distances. Within

one state, other interactions including those for errands, leisure, and school may be relatively more important. A gravity model may capture these interactions more effectively than workflows. The small correlation between intra-state workflows and the Pennsylvania gravity model ($\rho = 0.39$) indicated that movement within the state was not completely dependent on workflows. This notion was further confirmed by the differences in distance as a function of gravity models fitted to disease and workflows where the movement trends did not coincide at longer distances. Thus mechanistically, work-related commuting did not account for the majority of movement at longer distances and disease synchrony within Pennsylvania, and the epidemics between counties in Pennsylvania were synchronized by non-routine travel.

Estimating the movement kernel has important implications for accurately simulating disease spread. Multiple large-scale epidemic simulations have used a gravity-like model to simulate movement patterns [31, 60]. A simulation of pandemic influenza in the United States used a power law model for commuting data at the census tract resolution and fit a distribution of travel to work distances up to 200km reasonably well [31]. The gravity model fitted to workflows in the United States mortality analysis displayed evidence of a distance threshold whereby limited work movements occurred beyond distances of 119km [56]. A similar distance threshold existed for the Pennsylvania gravity model fitted to workflows where work movements declined rapidly until 200km; this further validates the movement kernel used for the simulation modeling.

A comparison of the exponents between the Pennsylvania gravity models (disease and workflows) highlighted differences in the movement kernel. As expected, for travel to work, the gravity model fitted to workflows produced larger distance and population exponents than the gravity model fitted to disease spread. The larger distance exponent reflected a rapid decline in

movement which was more common with routine work commuting. The estimated distance exponent of 0.098 from the Pennsylvania gravity model fitted to disease spread was only slightly larger than 0 which indicated that movement is independent of distance which was evidenced by the U-shaped curve of distance as a function of gravity. The smaller population exponent for the gravity model fitted to disease revealed the importance of smaller populations in the movement of non-routine travel and ultimately in the spread of disease.

The 2006/07 season displayed significantly lower synchrony as a function of distance, population size, and workflows. This season reported the fewest number of cases during the study period; and as a result, the smaller populated counties are more likely to have fewer cases (data not shown). This was evident as no cases occurred in Forest, Sullivan and Susquehanna counties. Fewer weeks with zero cases will produce lower Spearman correlations between all counties, but most notably in the counties with smaller populations and less movement. Additionally, influenza A/H1N1 and influenza B dominated the 2006/07 season. These viruses are known to result in lower morbidity and mortality and synchrony across the United States [19, 38, 56]. Similar patterns could be occurring at the smaller spatial scale; however, these patterns were not noted during another A/H1N1 and B season 2008/09. Additional time series are needed to confirm these results.

The correlation matrix based on the gravity parameters derived from the complete United State county data was not a strong predictor of influenza spread within Pennsylvania. Local variations within Pennsylvania may be masked among a larger county analysis of the United States as a result of differences in the range of area, population size, and distance between counties. A gravity model fitted to the United States smoothed over these differences and conceals the variation in smaller states.

It was difficult to validate the findings from the United States mortality analysis because of the quality of the reported data. Accounting for sampling error among the smaller communities with the binomial sampling method was one approach to adjust for the inherent problems associated with passive influenza surveillance system data. However the extent of this sampling (reporting) error was not known and may not be fully accounted for in the analysis. Sampling error could have presented in the form of noncompliance in reporting, subject failure to seek testing, and severity of illness. These biases led to fewer reported cases and potentially affected the timing of the cases resulting in smaller correlations. These types of ascertainment could have had its greatest impact on synchrony over longer distances and thus underestimated the true association with distance. Without data on complete reporting for any one county in Pennsylvania, it was difficult to assess the extent of the bias in the correlations. Spatial analysis of influenza sentinel reporters did not observe spatial variation with the incidence of reporting suggestive of a potential limited role in a spatial reporting bias (Stark Dissertation). Additionally, variation in vaccination rates across the counties could have impacted these results. Lower vaccination rates among rural, smaller populated counties, would have lead to an increase in the number of cases; thereby, overestimating the Mantel correlation for population with smaller counties [106]. County-specific vaccination rates for Pennsylvania are not known; however, vaccinations rates among the elderly (Age > 65) have met the 70% Healthy People 2010 goals suggesting vaccination rates for this at-risk population were quite high [107, 134].

Determining edge effects remains a challenging task in spatial analysis. For this analysis, special concern was devoted to adult movement across state borders which necessitated the development of an intrastate workflow matrix. Incorporating a total of 302 counties from the bordering states, including Pennsylvania, resulted in minimal flow between several counties

outside of Pennsylvania and those within Pennsylvania, thus, not significantly impacting the correlations with disease spread. Though, the correlation between the workflow matrices was 70%, indicating nearly a third of work-related travel occurs across the state borders. While the correlation with this matrix was not a strong predictor of overall disease synchrony, the opportunity for border transmission still exists in the form of non-routine travel. We did not account for interstate long distance or air travel as these forms of travel are negligible for each county of the state.

Age-specific attack rates vary by influenza strain and subtype [27, 34]. Influenza B and A/H1N1 typically infect younger populations which may be more mobile within communities but are less likely to be accounted for in the workflow matrix or gravity model. The elderly are also less likely to appear in the workflow matrix and this population is strongly associated with influenza hospitalization and mortality. Mortality analyses among the elderly have shown greater synchrony among A/H3N2 seasons than seasons dominated by A/H1N1 and B [38, 56]. Nonetheless, strain and subtype-specific analyses would further illuminate the determinants of disease spread between counties. However, small influenza B samples from each county and limited data on influenza A subtypes prevented further analysis.

This study documents the gravity-like spread of disease within the state of Pennsylvania; thus placing less emphasis on the value of administrative borders for public health prevention methods. Public health officials should target interventions to multiple counties to effectively capture the flow of residents and the spread of disease. Interventions targeted to patches of the state that display significant gravity-like spread of disease might be more efficient than statewide campaigns and provide greater public health value.

The precision gained from using county-specific disease and exposure data improved our knowledge of spatial-temporal predictors of disease spread enabling this study to delineate differences in mechanisms dependent on geographic scale. While this study incorporated workflows from neighboring states, it did not include disease data. Future studies should incorporate disease data from the neighboring states to confirm the gravity-like spread of disease across a larger administrative boundary. Through analysis of county-specific data, these results can be used to inform mathematical models of influenza spread at a narrow spatial scale.

3.6 TABLES

Table 5. County characteristics

	All Counties (N=67)	Large Counties (N=30)	Small Counties (N=37)
Mean population size	183,300	349,600	48,440
Population range	4,946 – 1,518,000	120,000 - 1,518,000	4,946 – 94,640
Total number of weeks*	12462	5580	6882
Proportion of total weeks with a case	41%	53%	31%
Mean number of weeks with a case for each county	76	98	58

* 186 weeks over 6 epidemic seasons (31*6)

Table 6. Observed Mantel statistics. Pearson correlation of the dissimilarity matrices and Spearman rank correlations of the epidemic time series for all counties (N=67). P-values and the corresponding confidence intervals (CI) are presented. Gravity Pennsylvania refers to the gravity model fitted to Pennsylvania-specific data.

	All counties (N=67)			
Matrix	**Correlation**	**P-value***	**Lower CI**	**Upper CI**
Euclidean distance	-0.03	0.5528	-0.079	0.006
Workflow (Intra-state)	0.14	0.0001	0.129	0.157
Workflow (Inter-state)	0.08	0.0260	0.058	0.099
Population	0.33	0.0004	0.310	0.389
Gravity (United States)†	0.11	0.0013	0.094	0.140
Gravity (Pennsylvania – Workflows)	0.19	0.0001	0.169	0.245
Gravity (Pennsylvania - Disease)	0.63	0.0001	0.593	0.656
Gravity(Pennsylvania - Disease) adjusting for:				
Euclidean distance	0.63	0.001		
Workflow (Intra-state)	0.62	0.001		
Population	0.60	0.001		

† Gravity matrix generated using parameters derived from Viboud et al.

* Significance is determined at $P < 0.05$

Table 7. Parameter estimates for gravity models (disease spreads and workflows)

Coefficient (exponents)		**Gravity Model - Disease**	**Gravity Model - Workflows**
$\tau 1$ $\tau 2$	Population	0.265	0.47
ρ	Distance	0.098	1.76

3.7 FIGURES

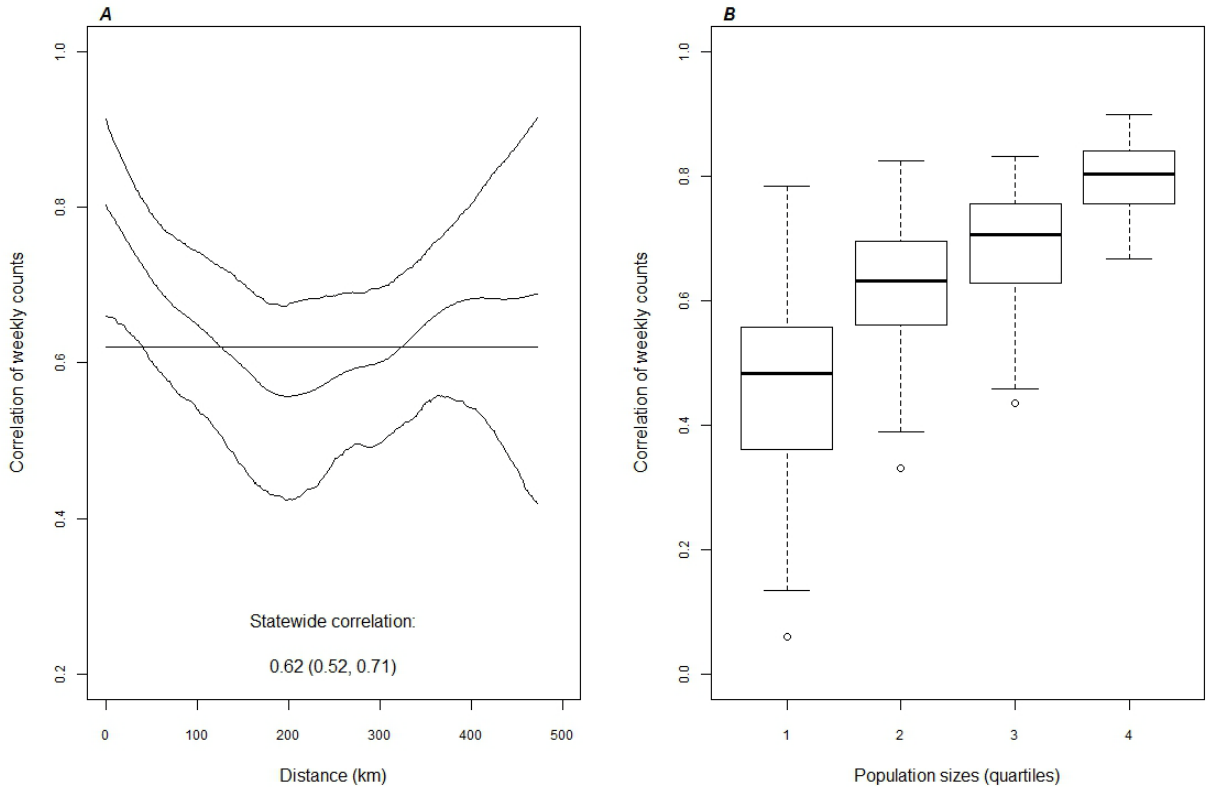

Figure 6. Correlation of weekly time series with distance and population size. A) Synchrony as a function of distance. The spline function (middle curve) is presented with a 9 5% confidence interval (outer curves). B) Synchrony as a function of population size (product of population i, j). The distribution of population was categorized by quartile. The boxplot within each quartile represent the distribution of the correlation of population between pairs of counties.

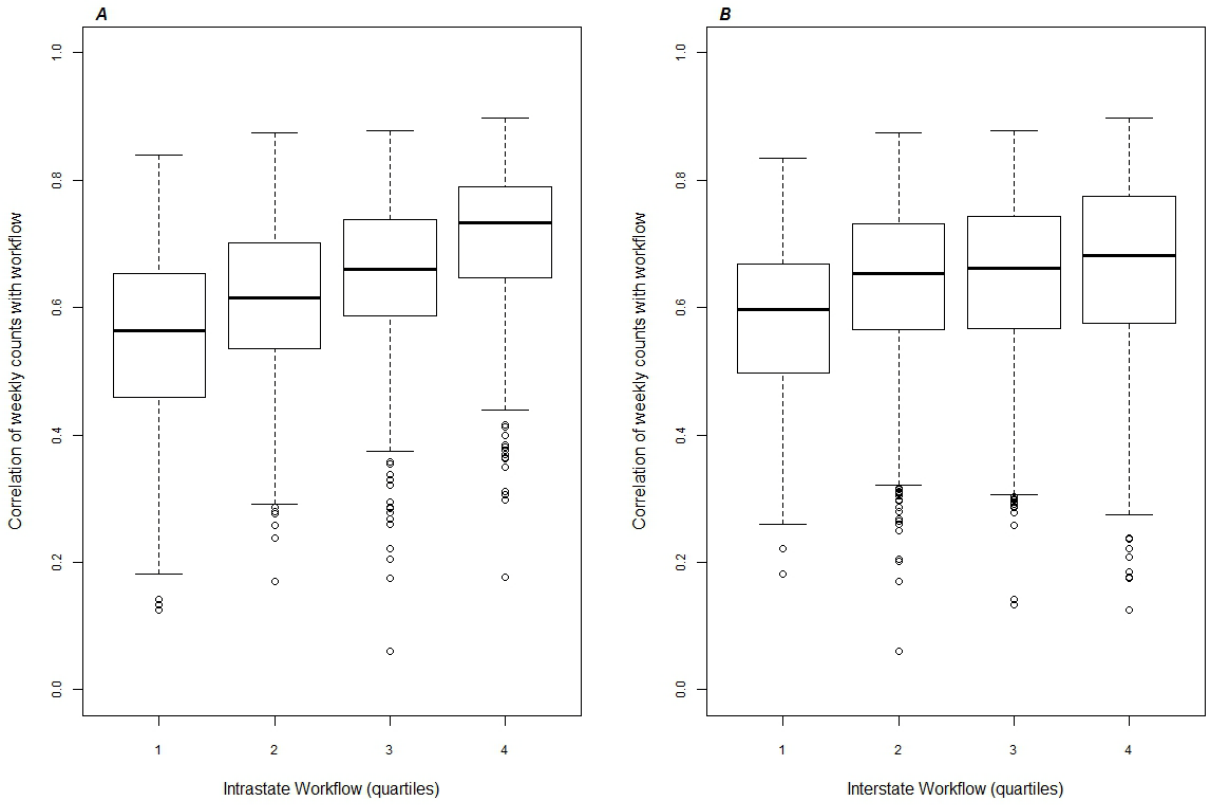

Figure 7. Correlation of weekly time series with human movement. A) Synchrony as a function of workflows. B) Synchrony as a function of Pennsylvania and neighboring county workflows. The distribution of workflow was categorized by quartile. The boxplot within each quartile represent the distribution of the correlation of workflow between pairs of counties.

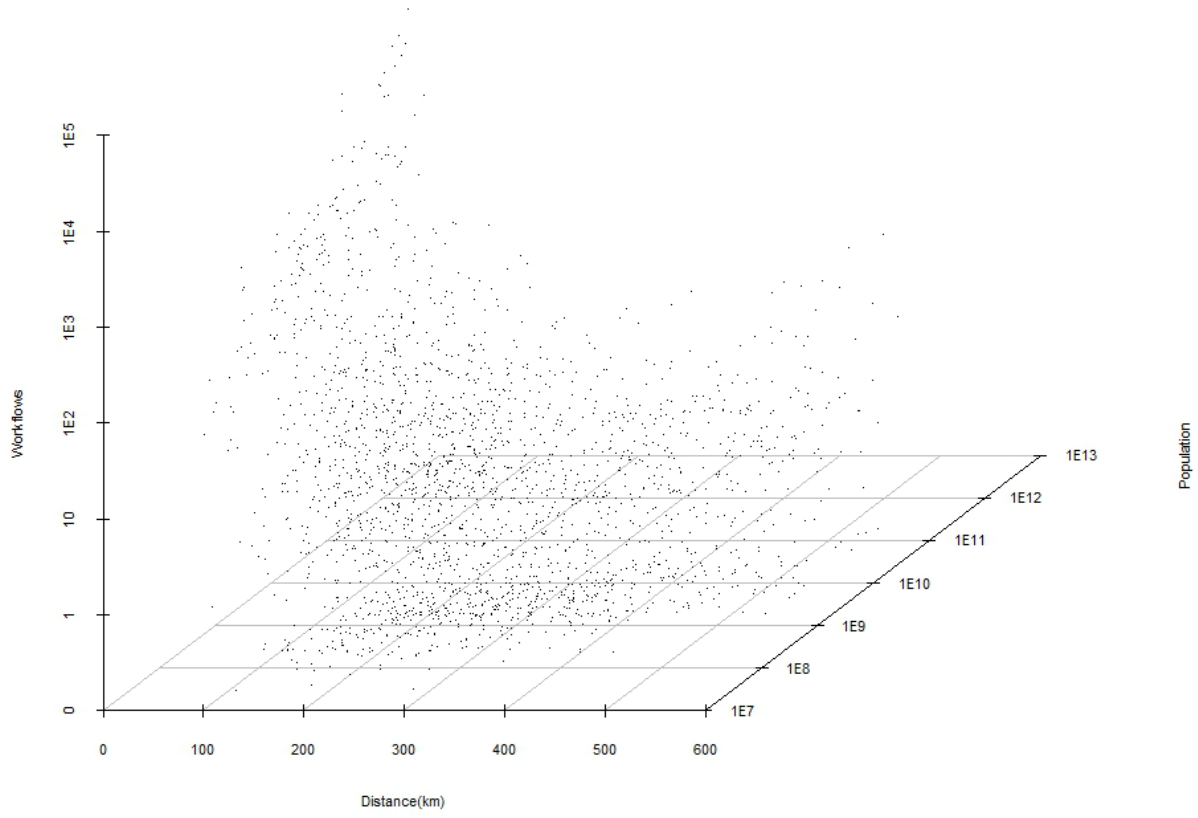

Figure 8. Association of workflows, population and distance. (y-axis, z-axis log10 scale)

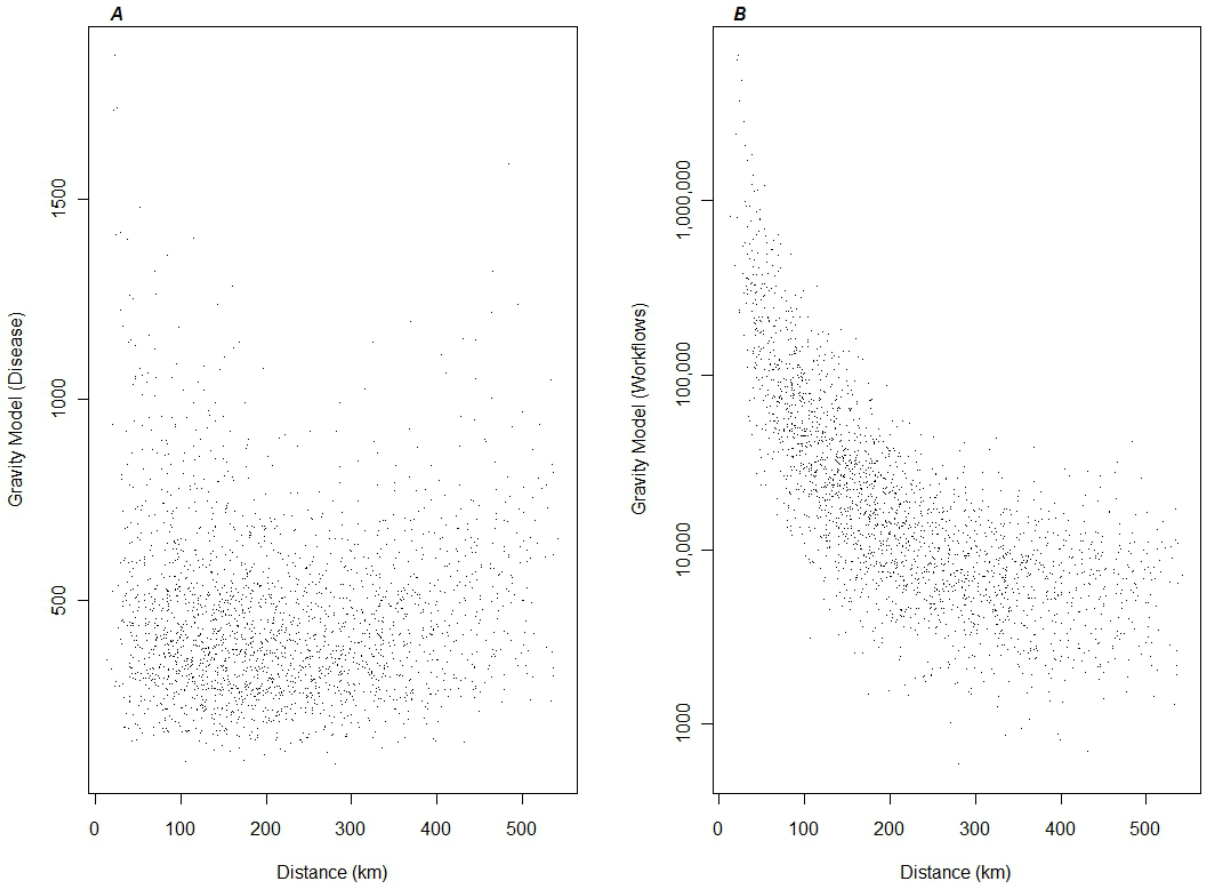

Figure 9. Correlation of gravity model with distance. A) Distance between two counties as a function of the gravity model fitted to disease. B) Distance between two counties as a function of the gravity model fitted to workflows (y-axis log10 scale)

4.0 MANUSCRIPT 3: ORGANIZING FUNDAMENTAL UNITS OF TRANSMISSION DYNAMICS: ONTOLOGIC UNITS OF EPIDEMIOLOGY

Manuscript in preparation

(Authorship to be determined)

James H. Stark[1]

Donald S. Burke[1], Derek A.T. Cummings[2], Bard Ermentrout[3], Stephen Ostroff[4], Ravi Sharma[5], Samuel Stebbins[1], Stephen R. Wisniewski[1]

[1] Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh PA, [2] Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, [3] Department of Mathematics, School of Arts and Sciences, University of Pittsburgh, Pittsburgh, PA, [4] Bureau of Epidemiology, Pennsylvania Department of Health, Harrisburg, PA, [5] Department of Behavioral and Community Health Sciences, Graduate School of Public Health, University of Pittsburgh, Pittsburgh PA

4.1 ABSTRACT

An important aspect of a public health system is to ensure that the key inputs and processes deliberately account for the health event in question. Challenges to this system arise when the geographic boundaries of a disease do not align with administrative, political, and legal units. One such example occurs among infectious diseases where the intrinsically spatial nature may produce boundaries incongruent with known spatial entities; thus, minimizing the effectiveness of public health surveillance, diagnosis, and the design of interventions. Here, we introduce the term ontologic unit of epidemicity to describe the spatiotemporal unit of transmission for an infectious disease. We review the methodological and philosophic framework in support of this concept and provide an example using influenza mortality data.

4.2 INTRODUCTION

Infectious disease dynamics are intrinsically spatial and directly transmitted diseases reflect human navigation of the physical landscape and the socioeconomic, demographic and behavioral heterogeneities in communities. Transmission patterns may not align with administrative boundaries, a characteristic expressed by the epidemiologic axiom that “disease knows no boundaries”. However, pathogens may reveal boundaries that are not readily

apparent. Understanding these boundaries as a subsystem of a partially decomposed system may be important to the success of public health surveillance and response activities.

Here, we present a review of conceptualizations of spatial units of transmission that have been formulated directly for transmission in related areas. In support of this approach, we introduce a new term: ontologic unit of epidemicity. We apply a set of algorithms to identify the spatial ontologic unit of epidemicity for influenza transmission using mortality data from the United States.

4.3 HISTORICAL PERSPECTIVE

Historically, defining the epidemiology of a disease has relied on existing territorial units for data acquisition, and comparison of social, demographic and economic measures. Administrative boundaries and the associated legal jurisdictions not only dictated the extent of epidemiological analyses but continued to influence disease surveillance. Presentation and access to seasonal influenza data located on the FluView website hosted by the CDC is organized by the nine US census divisions or the ten US Health and Human Service (HHS) regions [11]. The US census divisions initially took the form of colonial groupings and topographical regions. Though the three great regions and subdivisions were later adjusted in 1900 to correspond with political units to enable comparisons over time and improve data processing [135]. Researchers and public health officials have utilized the standardized data by census division and tracts to make valuable insights into public health issues. However, interpreting data and deriving innovative spatially focused public health measures is ultimately limited by the territorial boundaries that produced the data and the census divisions presented by the CDC for influenza challenge this approach.

Recently, data on human mobility patterns illuminated effective human connectivity borders that do not fit comfortably with existing administrative and political borders [136]. Moreover, membership in social networks such as Facebook reveals connectivity partitions based on the strength of geo-proximity to "friends" [137]. Human connectivity as demonstrated through analyses and simulations of influenza epidemics is an important contributing factor in the spread of disease [31, 56, 61]. Identifying how these connectivity borders and other drivers such as geography, demographics and climatic conditions are associated with the spatial structure of disease patterns is an important aspect of regional and global disease control; and thus, raises the issue of how to organize the appropriate spatial transmission unit of a host-pathogen system.

In this report, we seek to provide a methodological and philosophic approach to organizing the spatiotemporal unit of transmission for an infectious disease. In doing so, we introduce the term: ontologic units of epidemicity (OUE) formalizing this concept.

The term ontology is referenced in multiple disciplines; however, the concept originates from the branch of philosophy known as metaphysics. Philosophically, ontology simply refers to the study of existence or the nature of being (from the Greek word 'ont' – being) [138]. Over time ontology evolved into a distinct branch of metaphysics which incorporated concepts into establishing not only what is real, but the appropriate questions and classification systems to define what exists [139]. The ontologist, Roman Ingarden in his most influential work, The Literary Work of art, explains the requisite features and the interrelationships for an object to be considered a work of literature [139]. This ontological piece of work itself serves as a basis for defining ontologies in other disciplines including biology and artificial intelligence as a hierarchy

or classification of experiences, entities, and meanings. We extend the concept of ontology to public health as we explore the existence of epidemics, the components of a spatial transmission unit, and what relationship a spatial community structure has to physical entities.

4.4 PARTIALLY DECOMPOSABLE SYSTEMS

The idea that subsets of systems could be identified and analyzed in isolation from other subsets of a system has been suggested in many fields. Herbert Simon introduced the term partially decomposable systems to describe systems that could be divided into sub-systems that have much greater interactions within their sub-system than with other parts of the system [140]. Sub-systems of these partially decomposable systems (or nearly decomposable systems) could be studied or controlled effectively by considering sub-systems in isolation from the wider system. Decomposability could be a result of static weak coupling between sub-systems or a mix of time scales of interactions where within system interactions take place at fast time scales compared to longer time scales between subsystem interactions. As an example, island populations of wildlife might be considered decomposable systems when movement between islands occurs infrequently. The results of analysis or simulation that assumes decomposability provides an approximation of the full system which ignores coupling between all components of the full system. The accuracy of this approximation depends on the degree of coupling between sub-systems. Natural systems exist on a continuum between decomposable and non-decomposable systems. Useful empirical methods to identify subsystems should be able to characterize the strength of evidence supporting the existence of sub-systems in order to distinguish decomposable systems from non-decomposable systems [140].

Simon described the process of identifying these sub-systems as critical to decision making [140]. When considering all the possible connections and ramifications of decisions for every possible interaction, decision makers could effectively be paralyzed. Scientists engage in a similar process when deciding which components to include in observation or simulation and which to exclude. The claim in this manuscript is that the transmission cycle of human-to-human transmissible infectious diseases can be considered a partially decomposable system. These systems might be decomposed into distinct populations of the host and pathogen determined by geography, climatic conditions, demographic indices or some other unknown drivers. Identifying the sub-systems of this system can reveal what these drivers are and allow for progress to be made in determining the relationships between the multiple variables that might affect the level of transmission in any one population.

4.5 INFECTIOUS DISEASE EXAMPLES

Spatial transmission units have been observed for several infectious diseases. Geographic variation in HIV clade structure is a distinguishing feature of the HIV pandemic with more than 9 subtypes and multiple circulating recombinant forms dominating regional epidemics [141]. Africa retains the distinction of having the highest prevalence among continents and the greatest subtype diversity resulting in many distinctive geographic patches of subtype dominance. The reasons for global diversity remain elusive; however, founders effects, host genetics, and socio-behavioral features have been postulated [141].

Well-documented measles epidemics throughout the world have provided an extensive opportunity to evaluate and report on spatial units of transmission. Within Cameroon, patterns of

measles incidence differed by region of the country with the Northern provinces having experienced annual epidemics while the Southern provinces experiencing a three year periodicity [142]. Barriers to migration and social contact may have contributed to the regional differences, although, differences in birth rates, population density, and vaccination coverage may have been contributing factors. The consequence of differing measles dynamics within the country is reflected by the supplemental immunization campaign undertaken in the Northern provinces. Additionally, measles infection hierarchies observed in England and Wales noted phase differences among urban centers, small towns and rural areas [121]. Urban centers serve as a reservoir of infection for smaller spatial entities; as a result, concentrating vaccination in these urban areas could have dual purposes in reducing infection and the limiting the role of the reservoir.

Spatial heterogeneity of serotypes is an important characteristic associated with *Streptococcus pneumoniae*. The most prevalent serotypes differ between developed and developing countries and not all of the countries are endemic to the same serogroups [143]. Both serotypes 1 and 5 display significant global geographic variation. These serotypes are more likely to be found in South America and other developing countries as opposed to the United States, Canada and in similar developed countries. [143, 144]. In this example and similarly for the HIV example, the distribution of subtypes demonstrates the interaction networks that produced these spatial variations. In a well-mixed system, we would not expect to observe the homogenous subtypes.

Another subsystem of *Streptococcus pneumoniae* to consider is the relationship of the pathogen to age of the host. Specific subtypes are more associated with younger age groups which may be related to the maturity of the immune system [144]. For example, the odds of

infection with serotypes 6, 14, a nd 19 significantly decline after the first decade in life [144, 145]. Thus, defining the appropriate serotype-specific OUE for both the geographic and age-related subsystem is imperative for public health strategy.

4.6 HISTORICAL AND SOCIOLOGICAL PHENOMENA

Recognition of ontologic units of is not limited to infectious disease dynamics. The field of dialect geography is concerned with the geographic distribution of dialects and the fundamental organizing principles. The Atlas of North American English (ANAE) presents a comprehensive analysis of the current state and diversification of North American urban dialects and sound changes [146]. Through a telephone survey of all major urbanized areas, the researchers conducted an extensive acoustic analysis of phonological and phonetic features from native speakers of that city. The resulting analyses identified multiple distinct dialect regions (isoglosses) within the United States and areas of clear internal homogeneity within the Midlands and Eastern dialect areas. The North East is defined by multiple isoglosses while the majority of states in the traditional American west are concentrated into one isogloss [146]. Collectively, these isoglosses represent ontologic units of dialects.

Consistent with themes from patterns of disease spread, interaction networks described by diffusion and population movement are driving factors behind the spatial distribution of dialects. Both a gravity model and cascade model of diffusion support patterns of linguistic feature spread [147]. The latter model observes linguistic change spreading from the largest city to the smallest city in a hierarchical fashion proportional to the population size of the city and inversely proportional to the distance separating them. Studies in Norway, Iran, and Northern

Illinois support the models of diffusion and the Northern Cities Shift composing the Inland North isogloss of North America may have been the result of an urban diffusive spread [147-149]. Alternatively, some evidence suggests the phonological features of the Western states isogloss in North America have been attributed to the mixed settlement routes originating in the Northern, Southern, and Midland regions [146].

The concept of geographic variation in phonological features extends to social media on the internet. The microblogging tool, Twitter, enabled Eisenstein et al. to evaluate the geographic variation in the written lexicon noting strong regional differences in slang terms suggesting the importance of geography in slang usage [150]. The authors observe regional clustering with an emphasis of clusters in the major metropolitan areas. The geographic distribution of words is not new to linguistics. For example, the various terms used for a carbonated beverage are divided into 5 regions which do not necessarily mimic phonological or other lexical isoglosses [146, 151].

Another popular social media tool, Facebook, reveals interesting spatial patterns among networks of friends. Clear regional patterns exist emphasizing strong connections locally but fewer connections outside the cluster [137]. Variation also exists within each cluster as small cities within the West cluster had more ties to distant cities within the Western region, in contrast to the Pacific region in which the cities are tightly connected. While the Facebook spatial unit doesn't necessarily reflect movement or face to face contact, it does potentially represent a previous direct connection between two individuals. Though there is no guarantee of future contact, the Facebook ontologic unit represents the bounded likelihood of contact.

Human traveling behavior expressed through dispersal of bank notes has also led to the observation of unique spatial mobility boundaries [136, 152]. Through the use of an online

database of bank note trajectories for the United States, Brockmann et al. estimated a set of scaling laws to describe the dispersal of bank notes; and in essence, a proxy for human travelling behavior [152]. Further analyses revealed effective geographic boundaries for the United States [136, 153]. These human mobility borders correlated significantly with the Appalachian Mountain range and state boundaries but displayed unique partitions particularly with Pennsylvania and Missouri where the mobility border divided the state in two halves. The geographic information derived from the human mobility units has important implications in our knowledge of the geographical spread of disease and other human-related activities.

4.7 PUBLIC HEALTH SIGNIFICANCE

The utility of recognizing and defining a spatial ontologic unit of epidemicity has important public health implications. Strategies for public health prevention are inherently an exercise in defining; steps are taken to define surveillance activities, response activities, and coordination of these approaches. Recognition of the appropriate geographic unit to define these activities can strengthen the public health response and minimize inefficient prevention mechanisms.

Identification of a spatial ontologic unit impacts public health surveillance in multiple ways. Surveillance defined as the systematic collection, analysis and interpretation of public health data can be used to monitor disease occurrence, observe the natural history and clinical course of the disease, and provide epidemiologic insights [154]. Monitoring sudden changes and secular trends in an OUE may reveal shifts in underlying patterns of host movement and behavior that may require urgent or novel interventions within the OUE or potentially in the definition of a new OUE. Furthermore, changes in disease frequency within one region of an

OUE may prompt increased surveillance and prevention activity among the entire OUE. Because surveillance can support planning, implementation and evaluation of prevention methods, surveillance within a climate-driven OUE can inform the timing of an appropriate intervention.

One such example improving surveillance by defining an OUE is the Mekong Basin Disease Surveillance (MBDS) which is a group of countries in the Mekong Basin that formed a surveillance network as a result of cross-border disease transmission [155]. The Mekong Region in Asia consists of China, Thailand, Myanmar, Laos, Cambodia, and Vietnam, and is widely considered, as part of greater South East Asia, to be the epicenter for the emergence of pandemic influenza viruses and opportunities for zoonotic infections [4, 46]. Recognizing the destructive effects of morbidity and mortality on local populations from cross-border disease outbreaks, the countries of the Mekong Region established and strengthened partnerships to coordinate outbreak investigation, improve surveillance, share epidemiologic data, and develop response mechanisms [155]. In this example, the OUE of a particular host-pathogen relationship was not necessarily defined, rather, the territorial extent of multiple host-pathogen subsystems was acknowledged and the public health response was mounted.

Effective vaccine policy also benefits from defining the OUE. In the example noted earlier, Cummings et al. observed different oscillatory patterns of measles incidence between the northern and southern provinces in Cameroon which highlight the necessity of different vaccine policies for the two regions [142]. Regarding the HIV vaccine, the current strategy focuses on subtype and region-specific vaccines reflective of the ontologic units of epidemics; though, this approach has been questioned in light of data regarding conserved epitopes [156-158].

Another example demonstrating the public health implications of OUE recognition involves the pneumococcal vaccine. As stated, country and age-specific serotype coverage

influences the choice of subtypes used to compose the pneumococcal conjugate vaccine. As a result, the range of serotype coverage varies in each formulation in order to maximize regional and age effectiveness. This is especially true in the developing countries where serotype diversity is at its greatest [143]. Alternative strategies have focused on increasing the number of subtypes in the vaccine from 7-valent (pneumococcal conjugate vaccine) to 23-valent (pneumococcal polysaccharide vaccine); however, the 23-valent vaccine is not recommended for all groups as children <2 do not have the ability to mount an effective immune response to some of the serotypes [159, 160].

Finally, refinement of model assumptions based on knowledge of the ontologic unit will improve epidemiologic model development. Understanding the fundamental unit of transmission, the appropriate aggregation level, and the drivers of the host-population subsystem will enable models to more accurately predict disease spread and evaluate disease containment policies.

4.8 DEFINING CRITERIA

A spatial ontologic unit of epidemicity is a group of locations for which disease dynamics are more highly correlated with each other than would be expected based on the overall correlation measured in all areas. It is similar to community structure in network analysis in which a network is partitioned by the strength of connections within groups. As a theoretical construct, the spatial OUE reflects a threshold of similarity in the temporal patterns of incidence among spatial entities. The non-metric multidimensional scaling (MDS) algorithm can aid in discrimination of clusters and patterns as the spatial ontologies are self-defining. As part of this

exploratory process, multidimensional scaling can discover directional relationships and illuminate the relationship between epidemic space and geo-space. Non-congruency between epidemic space and geo-space is of potential interest as it identifies areas where spatial relationships do not entirely predict similarity in dynamics.

4.9 INCOMPATIBLE ONTOLOGIC UNITS

Failure to correctly define an OUE may lead to the wrong conclusion or mask the true association. Certain communicable diseases display seasonal periodicities leading to an annual reset of the disease. Thus, each epidemic season is one unit of observation. Partitioning a single season into multiple temporal observations neglects the inherent correlation between observations during a single epidemic. Inappropriately characterizing epidemic seasons may result in incompatible associations. Checkley et al. surmised that the El Nino phenomenon in the form of increasing ambient temperature resulted in an elevated risk for hospital admissions of diarrheal diseases among Peruvian school children [161]. The analysis focused on da ily observations over a twenty month El Nino event which consisted of two diarrheal epidemics. While the association posed and demonstrated by the authors may exist for this particular El Nino-driven event, establishing an OUE of diarrhea in Peru driven by weather-induced changes from El Nino necessitates analysis of multiple exposure periods from multiple surveillance sites. Another such example focuses on a Chikungunya virus outbreak in coastal Kenya beginning in 2004 [162]. The authors postulated unusually dry conditions precipitated a change in water storage and transmission efficiency. Similar to the previous example, the study focused on multiple pre-epidemic control seasons and only a single epidemic season. Persistent outbreaks

over multiple seasons would provide greater confidence in establishing an associated with climatic conditions. Finally, winter freezes have been implicated as a mechanistic force of St. Louis encephalitis virus epidemics in Florida [163]. The analysis of monthly sentinel data from 17 surveillance sites is limited to two epidemic seasons over a twenty year period. While winter freezes were positively associated with two epidemics, the absence of an epidemic following multiple winter freezes was also noted; and thus, confirmation and identification of a climate-driven OUE is incomplete.

4.10 EXAMPLE: INFLUENZA

Influenza is a contagious respiratory pathogen responsible for annual seasonal epidemics in temperate climates resulting in significant morbidity and mortality. Epidemics in the northern hemisphere typically occur between November and April with peak cases reported in the winter months. Each year significant resource demands are placed on local, state, and national public health officials to provide accurate and timely surveillance in order to respond with the effective prevention methods. Studies evaluating the spatiotemporal spread within the United States have been well-documented leading to numerous insights into the timing, magnitude, burden of disease, and transmission dynamics [12, 16, 19, 27, 164, 165]. Moreover, recent studies point to the synchrony of seasonal influenza epidemics within the United States and the world [38, 39, 56, 90]. These reports highlight the mechanistic forces that synchronize epidemics across multiple geographic scales; however, neither of these studies evaluates the epidemic unit. Knowledge of the ontologic unit of epidemics for influenza could support public health officials in multiple ways. Identification of sub-networks could identify and prioritize effective

surveillance zones, create improved vaccine distribution regions, and minimize duplicated efforts.

Some evidence of an influenza spatial ontology exists. Wenger et al. demonstrate a spatial trend in peak timing from Western to Eastern states with the New England region peaking the latest [57]. Peak timing patterns are most evident among the New England and Western states. Phylogenetic analyses of the H1N1 pandemic virus discovered distinct viral lineages that created clade-specific spatial patterns suggesting a role for founders effects [166]. Ultimately, this spatial heterogeneity disappeared during the second wave of the pandemic ceding to a single viral lineage from New York City [167]. These phylogenetic analyses are limited by available data from specific states and cities and are not representative of all regions.

To evaluate the presence of an influenza spatial ontologic unit in the United States, we analyzed pneumonia and influenza mortality for the 48 contiguous states from 1972-2002. Methods to collect and prepare the mortality data have been described elsewhere [56]. Pneumonia and influenza deaths were selected from the National Center for Health Statistics and organized by week and administrative region. For each season's epidemic, the period from December through April, a Spearman correlation of the weekly mortality rate between two states was calculated resulting in a 48x48 matrix (30 matrices total). Spearman correlations serve as a mechanism to quantify the similarities in epidemic timing between two states. Because of the weaker and potential inconsistent connections with the continental United States, Alaska and Hawaii were excluded from the analysis.

4.10.1 Statistical Analysis

Identifying regions with a similar pattern of epidemic time series is performed by non-metric multidimensional scaling (MDS); an ordination method used to identify spatial structure in data. This method orders objects based on the rank order of their similarities such that objects dissimilar from one another appear farther apart than objects that are similar in N-dimensional space [132, 168]. As opposed to ordering objects based on the distance separating them, an epidemiologic parameter, such as mortality rates, serves as the distance metric. Thus, the location of objects on a map corresponds to the degree of similarity in their mortality rates [126]. For this analysis, a Spearman correlation matrix of the epidemic time series are converted to distances using the square root of the reciprocal of the coefficient and applied to the sammon function in R.

Infectious disease dynamics have modeled population movements using a gravity model [56, 124, 126]. Originally developed for transportation theory, a gravity model captures the behavior of human movement [169]. For this analysis, we consider the generalized gravity model such that the interaction of sub-populations is inversely proportional to the distance separating them [124]. Using parameters previously identified for a gravity model of the United States, we estimate the stress of an MDS with an initial configuration of this gravity model. We compare this stress to an MDS with an initial configuration of a United States geographic map (state centroids). A lower stress reflects the overall fit of the algorithm and how well epidemic space coincides with a gravity model or geo-space.

All statistical analyses were performed using the R statistical package version 2.11 (R Foundation for Statistical Computing, Vienna, Austria).

4.11 RESULTS

Exploring the geometric relationships of influenza mortality rates over a 30 year period in the United States revealed several notable spatial patterns. Based on a threshold of similarity, several groups of states are clustered together similar to their spatial proximity as states in the United States. New England, parts of the South (Florida, Georgia, South Carolina, Alabama, Mississippi, and Arkansas), the upper Midwest (Michigan, Wisconsin, Illinois, Indiana, and Ohio) and many of the Western states have more synchronized influenza epidemics with each other than neighboring states or regions (Figure 11). However, there are several states clustered together more on the influenza map than expected based on the spatial proximity in the United States (California, Texas, Florida, Pennsylvania, Ohio, North Carolina, and New York). These larger populated states are pulled inward, and in some instances displacing the natural location of other smaller populated states suggesting a population factor may account for this map contraction. To test the hypothesis that the influenza maps created by dimension reduction are more consistent with a gravity model of movement than just spatial proximity, we compared the MDS stress iterations of the flu map, United States map and a gravity-formulated map. The gravity map resulted in the lowest first iteration stress (Gravity Stress = 32, US Stress = 47) indicating the flu correlations are more closely aligned to gravity space than geo-space; a conclusion supported elsewhere [56, 170].

Visualization of correlation thresholds provides additional evidence supporting the spatial synchronization of epidemics within regions (Figure 12). Each edge (line) between two states (vertices) represents the correlation of the epidemic during the study period. As the correlation threshold increases, fewer states display similar patterns in the timing of flu epidemics. Synchronized epidemics remain concentrated in the heavily populated Eastern states.

Connections between California and Texas and four states on the interstate 95 Northeast corridor (Massachusetts, New York, New Jersey, and Pennsylvania) have the most similar timing in epidemics. Furthermore, the epidemic correlations display a strong similarity to population density in the United States. Few epidemics in the Western states are synchronized beyond the $\rho = 0.45$ threshold. This lack of synchronization verifies the limited connectivity of the large population centers in this region compared to the Eastern region of the United States.

4.12 CONCLUSIONS

Identifying the spatial ontologic unit of epidemicity has meaningful public health consequences. In this report, we introduce the concept of an OUE and provide an introductory framework to identify the unit of transmission. Though our current knowledge of this concept is incomplete, future observations will improve our understanding.

The exploratory influenza analysis revealed interesting spatial units and patterns not previously reported. A previous analysis demonstrated a West to East trend in the timing of peak epidemics; however, we report on the directional relationship of synchronized yearly epidemics [57]. These clusters of synchronized epidemics present novel insight into the spread of influenza and influenza surveillance regions. Identifying discrete thresholds of similarity illustrate the relationship of epidemics in neighboring states such as Washington and Oregon and New Jersey and New York. The largest correlations in the Mid-Atlantic region suggest the scope of surveillance encompasses both New York and New Jersey and potentially Pennsylvania. Alternatively, the threshold of correlations between Washington and Oregon is not as strong. This pair of states may indicate the presence of a spatial surveillance unit but the strength is

limited. Nonetheless, these states should be more concerned with surveillance of one another than trends between other states in the region. Another implication of the data is the limited correlation in the Northwest and West with itself and other parts of the country. Weaker connectivity networks and greater distances between population centers may account for this finding.

In defining surveillance regions it is important to note the limitations of the available data. Analysis of mortality data between states lacks the granularity to capture the true spatial ontologic unit. Pennsylvania's relationship with New York, New Jersey and Massachusetts may reflect baised mortality patterns in the Eastern region of the state that is closest to New York, New Jersey and Massachusetts. The Western portion of Pennsylvania because of its unique geography may have stronger connections to the Midwestern states. County or census tract analysis is likely to reveal the true ontologic unit and the most appropriate region for optimal surveillance.

Congruency of the multidimensional scaling influenza map and the United States geo-space map suggest that spatial proximity between states is an important aspect in the correlation of epidemics. However, non-congruency between the maps implies another factor may be driving the spatial relationships. We tested the hypothesis that the spatial correlation in incidence is more consistent with human movement flows than spatial proximity; however, climate and regional and historical differences in reporting may also drive the spatial relationships.

Identification of influenza transmission units has additional implications for constructing transmission models. Typically, simulations model the explicit administrative unit. However, these analyses suggest a more refined, accurate model would be to aggregate the administrative unit to reflect the spatial units of transmission such as collapsing New England into one unit or

developing coinciding models depending on p opulation size. Moreover, by understanding the disparate relationship among Western and Northwestern states, predictive capabilities of models can be further refined.

Many statistical techniques are available for the identification of spatial ontologies. Here, we present the results of a multi-dimensional scaling algorithm that visualizesthe directional relationships between states. Other methods such as hierarchical clustering, principal component analysis, factor analysis, and latent class analysis are all capable of realizing spatial transmission partitions.

In conclusion, the realization of infectious disease spatial ontologies can have profound effects on approaches to surveillance and public health policies. Currently, public health policies and prediction models craft strategies based on the administrative unit used to analyze the data. However, consideration of the transmission unit is more prudent.

4.13 FIGURES

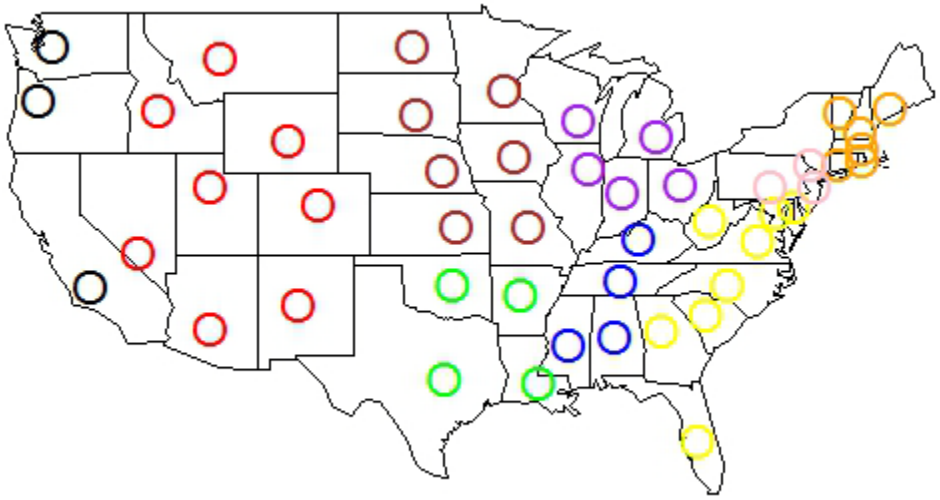

Figure 10. Map of the United States with color-coded circles representing US Census regions

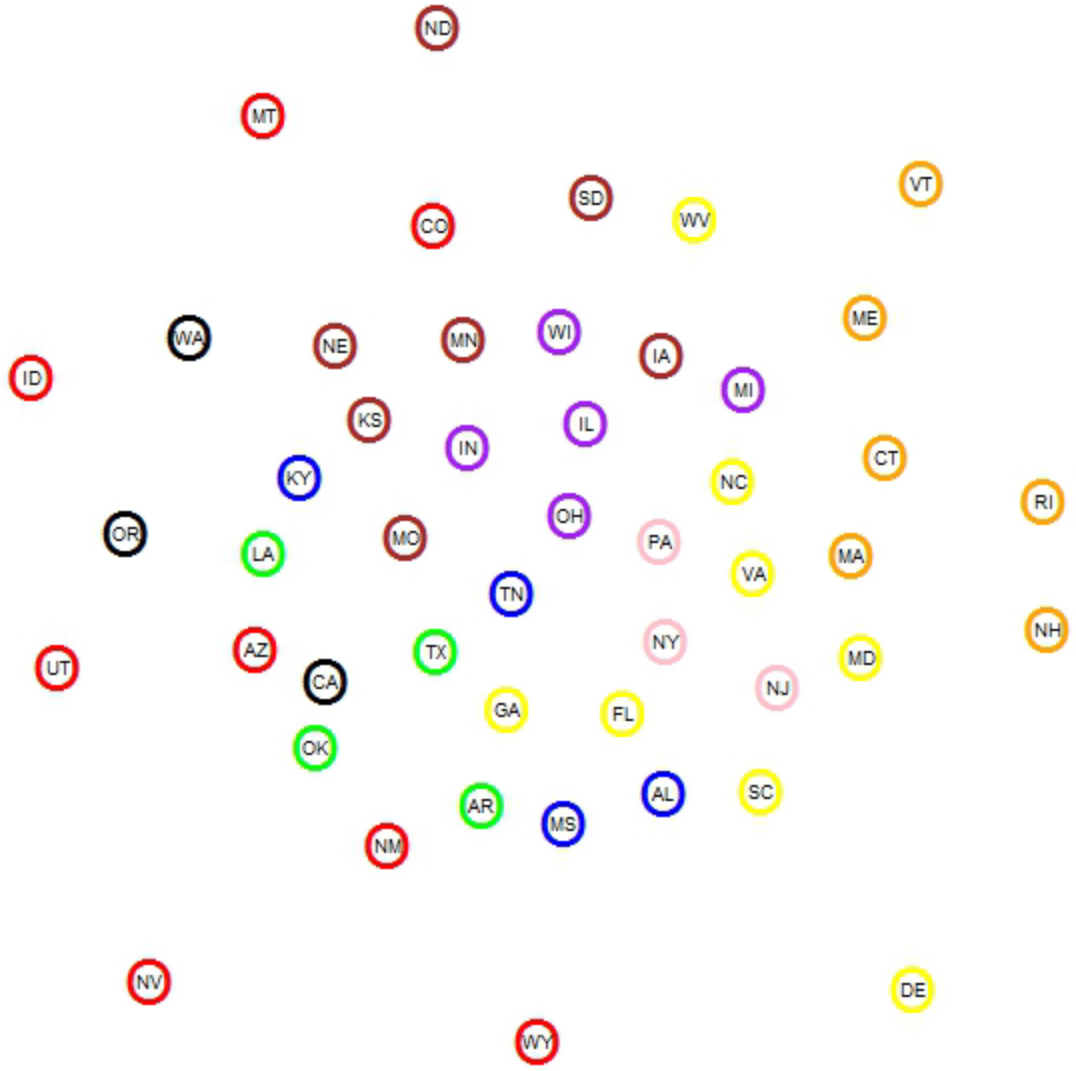

Figure 11. Map of influenza spatial structure from multidimensional scaling using color coded circles to represent US Census regions

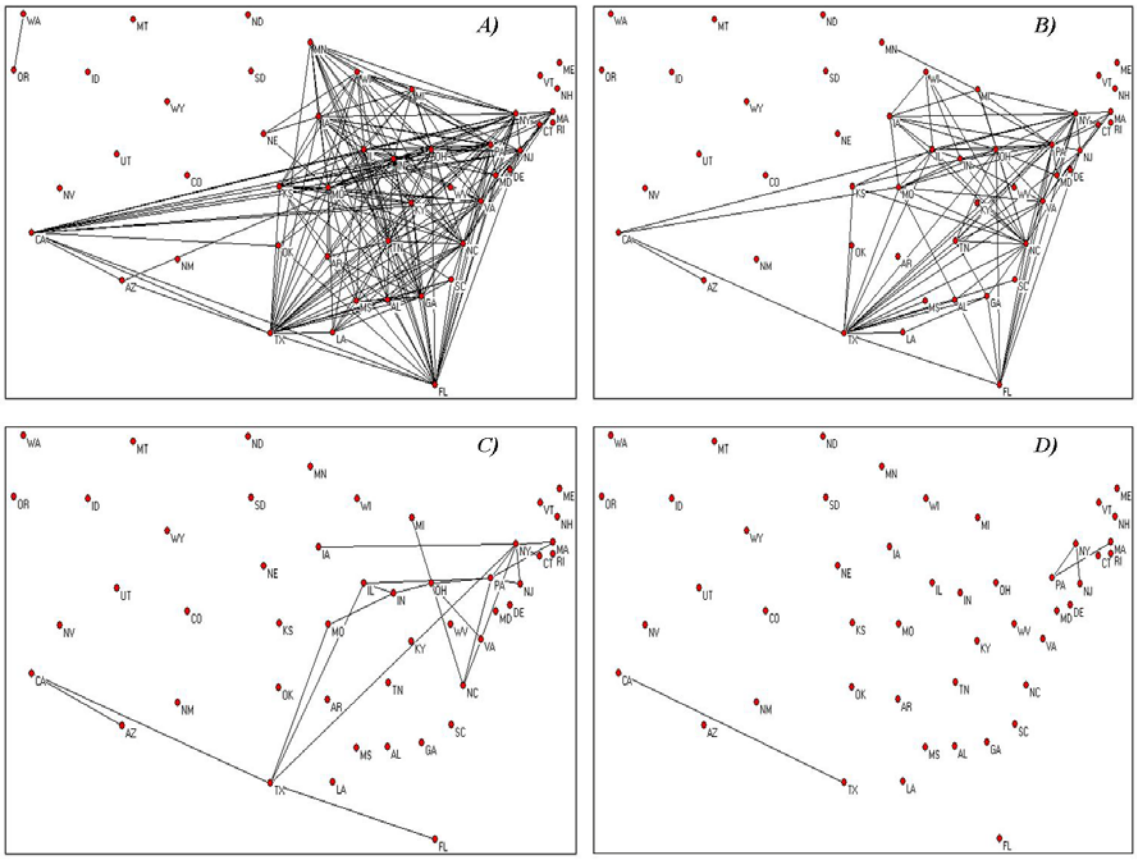

Figure 12. Spatial network representations of correlation thresholds. Each edge (line) represents a correlation between two states. A) Correlations > 0.45. B) Correlations > 0.5. C) Correlations > 0.55. D) Correlations > 0.6

5.0 CONCLUSIONS

By using spatially and temporally explicit influenza incidence data from the Commonwealth of Pennsylvania and influenza mortality data from the United States, this dissertation characterized seasonal influenza epidemics, evaluated factors that drive local influenza epidemics, and provided an initial assessment of how administrative borders influence surveillance for local and regional influenza epidemics. Understanding the geographical differences in the spatio-temporal dynamics of seasonal influenza epidemics has important implications for public health systems as practitioners can gain insight into local and regional differences to improve public health planning and utilization of resources.

Despite the regularity of seasonal influenza epidemics, differences in the intrinsic properties of disease dynamics may exist across spatial scales. Previous research has described trends in timing, magnitude, and synchrony for countries and continents; however, similar analyses have not been completed at a finer spatial scale such as a county. Inferences reflecting the spatiotemporal patterns of influenza spread conducted at larger spatial scales may conceal local trends in disease structure. This dissertation reports the results of the first study to examine the spatio-temporal dynamics of influenza between counties within one state. Several important findings were noted, though not all were consistent with patterns observed across larger spatial scales. The analyses revealed similar temporal trends and synchrony in the timing of epidemics with epidemics across the United States. Synchrony existed across the entire state and decayed

with distance. Furthermore, synchrony as a function of population size displayed evidence of hierarchical spread with more synchronized epidemics occurring among the most populated counties. Despite the similarities in synchrony across spatial scales, the mechanism of disease spread differed. Dissemination across regions of the United States is driven by air travel and adult work flux drives interstate spread. This dissertation demonstrated that a gravity model was the strongest predictor of influenza spread and suggests that non-routine and leisure travel drive local epidemics. This finding has not been previously reported and highlights the complex nature of influenza spread across multiple geographic scales.

Spatial analysis revealed heterogeneity in the distribution of influenza across the state. A test of autocorrelation revealed a cluster of cases occurring in the South Central region of the state consisting of 8 counties (Union, Snyder, Juniata, Mifflin, Centre, Huntingdon, Fulton, and Bedford). This spatial regime existed each year. Multivariable logistic regression indicated that lower monthly precipitation levels during the influenza season, fewer residents over age, and fewer residents with more than a high school education were significantly associated with membership in this cluster. Association with age and education may have reflected differences in vaccination coverage. These results provide the Pennsylvania Department of Health with the first indication of influenza case clustering in the state. These results also provide an opportunity for future research and collaboration. The possibility exists for extension of this spatial regime across the state borders into West Virginia and Maryland. The next phase of research should concentrate on acquisition and analysis of county-specific influenza incidence data for these two states to observe if the spatial cluster extends across state administrative units.

Understanding how administrative borders influence surveillance and prevention methods for local and regional epidemics presents an additional challenge to the public health system as

noted in the spatial cluster from Pennsylvania. Defining the epidemiology of disease often relies on the administrative boundaries for data collection, covariate analysis, and intervention. However, the intrinsically spatial nature of infectious diseases may result in transmission boundaries that do not align with existing territorial units. Pathogens, such as influenza, may expose transmission boundaries that are not readily apparent. As a result, efforts to implement effective surveillance and response measures may be inefficient and inadequate. Recognizing the spatial transmission unit is important for the success of public health activities.

This dissertation introduced the term ontologic unit of epidemicity (OUE) to describe the spatiotemporal unit of transmission for an infectious disease. Spatial ontologies have been observed for many infectious diseases including measles in Cameroon, global HIV clades, and country-specific subtypes of *Streptococcus pneumoniae*. Extension of this concept to sociologic phenomena such as dialect regions and online social networking suggest an underlying pattern of connectivity in the United States. Through analysis of influenza mortality data for the United States, this dissertation provided a framework for OUE definition and analysis. The geometric relationships of influenza correlations revealed strong congruency of epidemic space and geo-space suggesting epidemics are well-synchronized in space and were consistent with a gravity-like spread. Though regional clusters of connectivity existed, most notably in the Northeastern states of Massachusetts, New York, New Jersey, and Pennsylvania. The Western states had a lower threshold of similarity which may reflect weaker connectivity networks and greater distances between population centers. Realization of these spatial transmission units has implications for regional surveillance and simulation modeling.

5.1 PUBLIC HEALTH SIGNIFICANCE

Seasonal influenza is a significant cause of morbidity and mortality with nearly 30,000 deaths and over 200,000 hospitalizations occurring in the United States each year. Analyzing the spatial temporal dynamics across multiple geographic scales provides greater understanding to the complex patterns of influenza spread. This dissertation provided novel insight into influenza dynamics within Pennsylvania and the United States which can be used to improve public health surveillance and response measures.

5.2 FUTURE RESEARCH

Future research should concentrate on further illuminating the role of administrative borders in defining the epidemiology of disease. Spatial analyses are needed in the Southern Border States of Pennsylvania to confirm the presence of the elevated incidence cluster. Additionally, identifying congruency between a social movement or human mobility patterns and spatial incidence patterns could reveal the underlying mechanism of influenza spatial ontologies.

APPENDIX A

SUPPLEMENTARY FIGURES FOR MANUSCRIPT 2

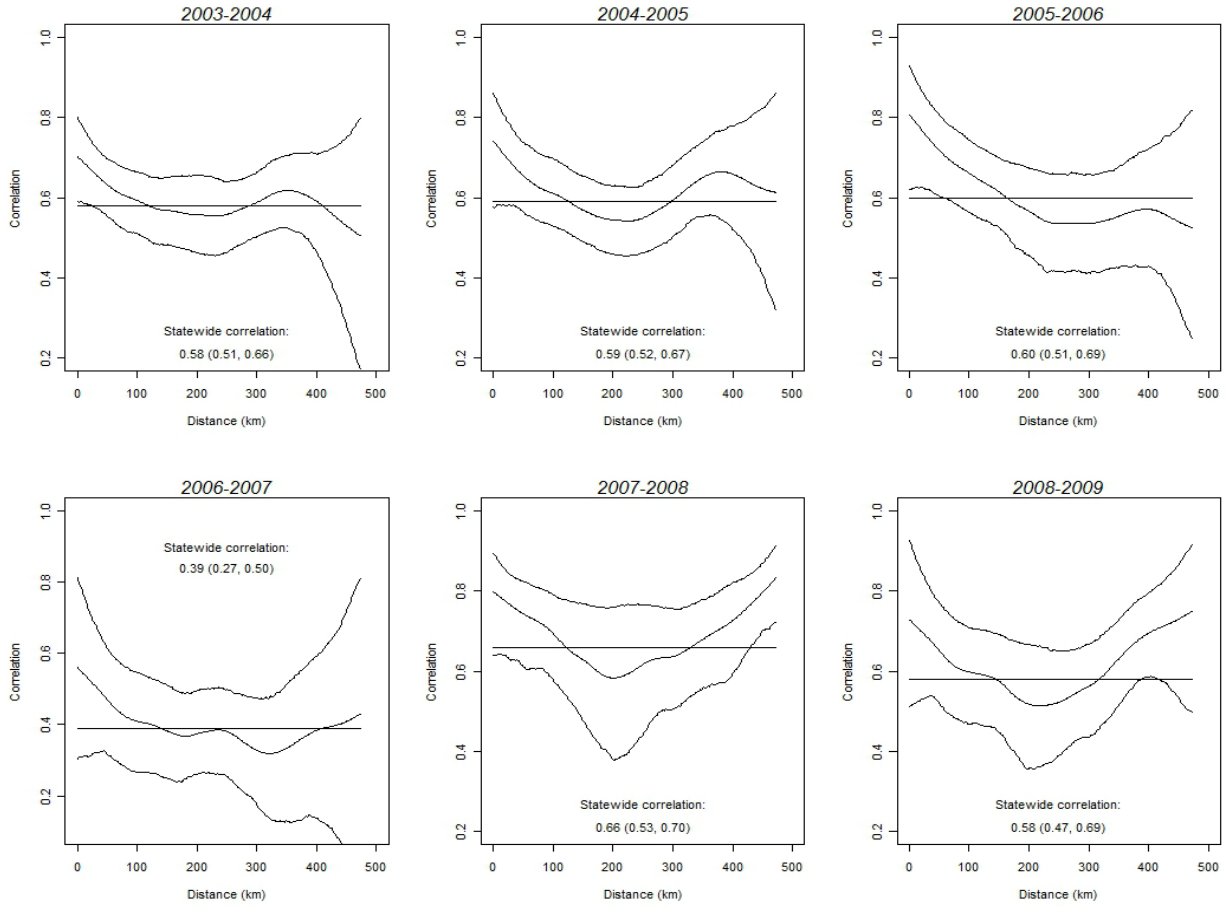

Figure 13. Synchrony as a function of distance for each season

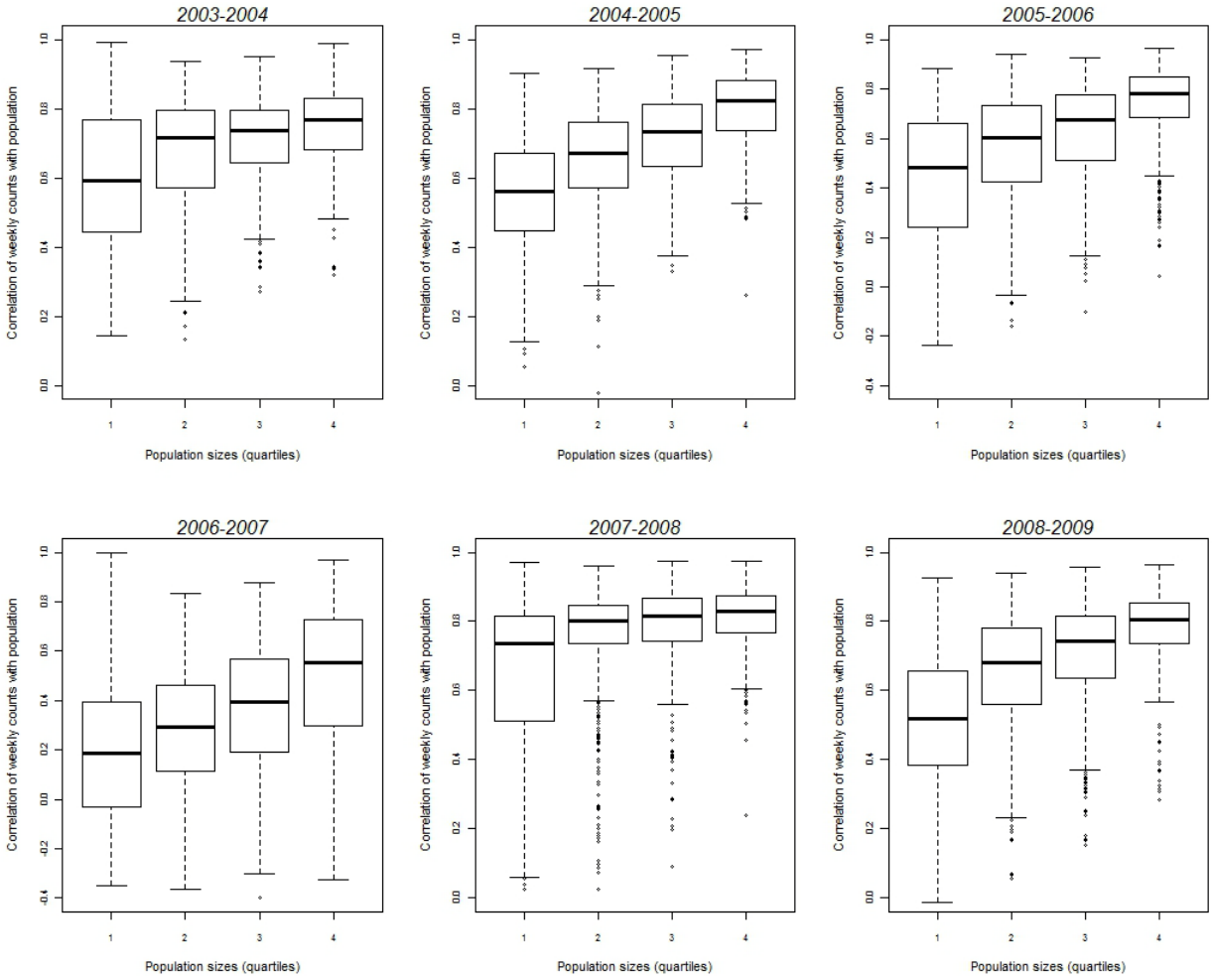

Figure 14. Synchrony as a function of population for each season

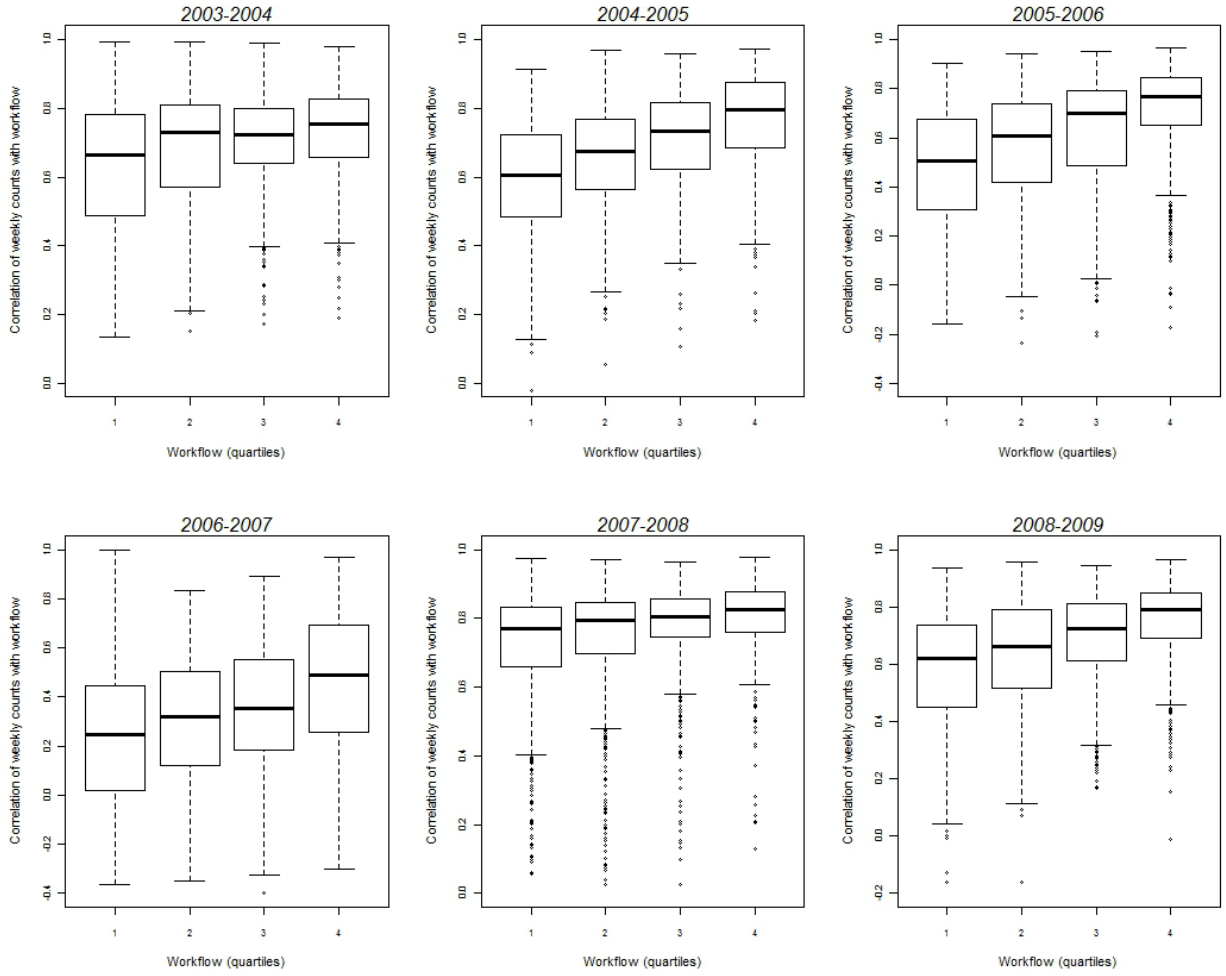

Figure 15. Synchrony as a function of workflows for each season

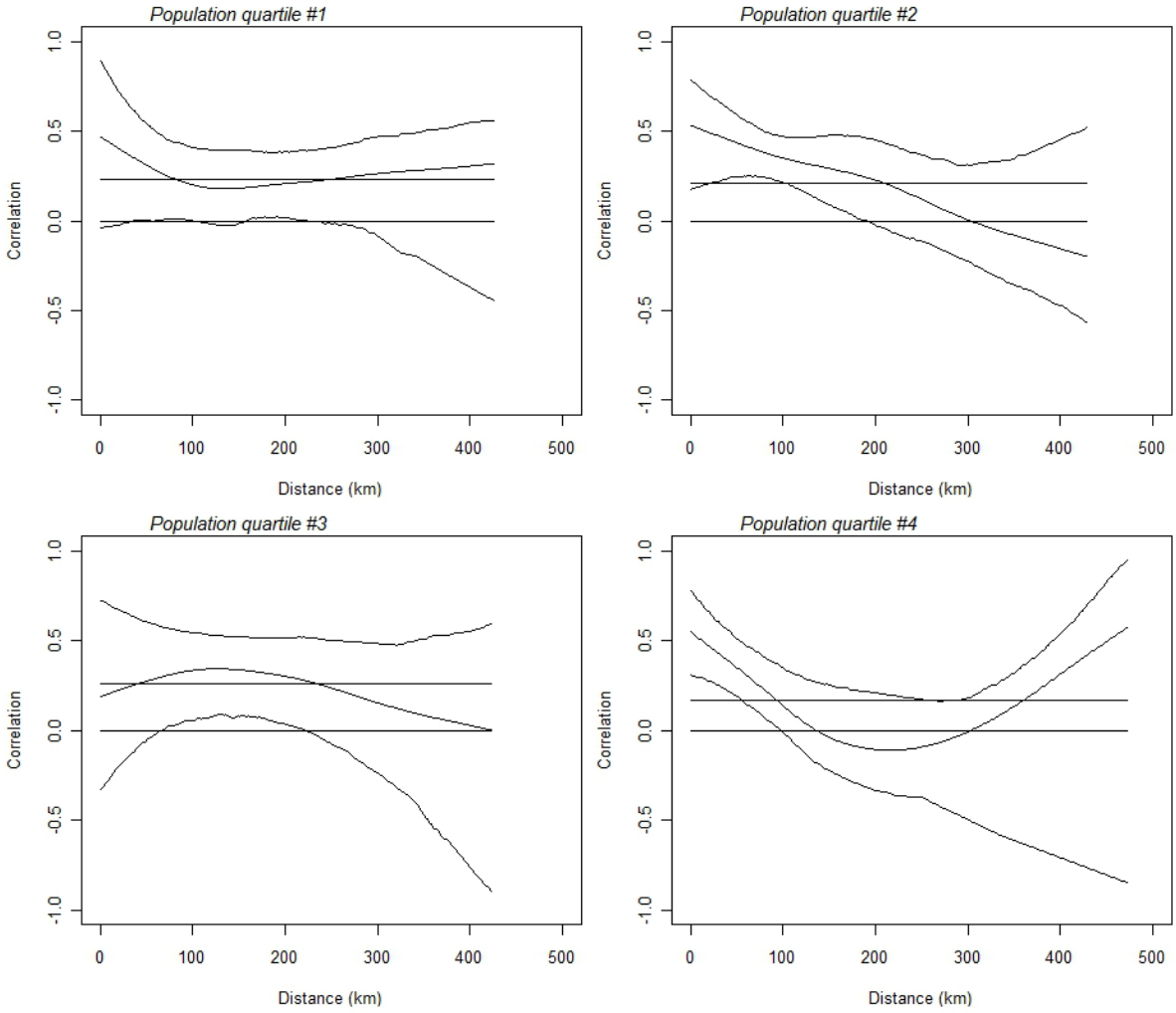

Figure 16. Synchrony as a function of distance for each population quartile

APPENDIX B

SUPPLEMENTARY METHODS

B.1 MORAN'S *I* STATISTIC

To evaluate the extent of spatial similarity of cumulative incidences between Pennsylvania counties, we used the Moran's *I* method. This is a global measure summarizing the spatial autocorrelation across the entire state. The value of the *I* statistic can be defined as the product of the difference between Y_i and Y_j with the overall mean divided by the variance observed among Y_i [93]. The formula is similar to a Pearson's correlation coefficient and can be considered a spatially weighted variant of the Pearson correlation coefficient. Neighboring counties with similar patterns of incidence will have a positive value, while neighboring counties with different incidences will have a negative value. A positive *I* statistic indicates the pattern among counties is clustered. Similar to Pearson's correlation coefficient, the *I* statistic can take on values between 1, -1. Though, it is theoretically possible for regions with extreme values to be heavily weighted and thus have an upper bound > 1. Significance is determined by comparing the observed value to the expected value under the null hypothesis. Permutations of the data through randomization of the counties generate a distribution. The observed value can be compared against the tails of the distribution. If the observed value lies in the tails, then we conclude that

the neighboring counties have statistically significantly more similar incidence than would be expected among randomly selected pairs of counties.

The Moran's *I* statistic represents a global measure to detect spatial autocorrelation over the entire study area; however, this method cannot detect individual clusters within the study area. The Local Indicators of Spatial Autocorrelation (LISA) provides a local measure of similarity between counties. The LISA statistic is formally linked to the global test because the values of the *i*th region sum to the global indicator, thus each LISA value is a component of the global index [93, 94]. Thematic maps are generated of the LISA values such that high-high or low-low values represent clusters of similarity. Regions of low-high or high-low indicate the cumulative incidence of a county is an outlier relative to the neighboring counties.

B.2 MANTEL STATISTIC

We estimated the association between epidemic synchrony and key predictors using a Mantel statistic. The Mantel test is a regression which estimates the comparative elements between dissimilarity matrices summarizing the pair-wise similarities among counties. The predictor variable represents the workflows between counties *i,j* and not just the workflow for county *i* or *j* [133]. Thus, the question to be addressed in this analysis is: do counties that are similar in epidemic timing also tend to be similar in workflows. The Mantel statistic is based on a cross-product term than is normalized to the data and results in a statistic bounded to 1, -1 [133]. Significance is calculated by permutation of the rows and columns of the matrices. Ten thousand iterations of the randomly rearranged matrices generate a distribution of Mantel statistics which is compared to the observed value. A Partial Mantel test is used to measure the association of

two matrices in the presence of a third matrix. The Partial Mantel test can address questions such as: how strong is the association between epidemic synchrony and the gravity model after removing the effects of distance. The Mantel function in R estimates the Mantel statistic.

B.3 MULTIDIMENSTIONAL SCALING

Non-linear mapping in the form of non-metric multidimensional (MDS) scaling presents a mechanism to examine the spatial structure of events. Events in geographic space are represented with an epidemiologic metric instead of a traditional distance metric. As a result, the position of points on a map represents the degree of similarity of an epidemiologic variable [126]. For example, mortality rates of influenza for each state can be mapped such that the location of objects on the map corresponds to the degree of similarity between mortality rates.

The objective of the MDS is to represent the objects with the least number of dimensions while preserving the distance relationship between the objects [126]. Computation of an MDS employs a distance matrix using the square root of the reciprocal of the coefficient. For this study, a Spearman correlation matrix of the weekly mortality rates is converted to a distance metric. An initial configuration of the geographic locations of each state is used to preserve spatial proximity. Distances from the initial configuration are estimated and regressed against the distance matrix to generate predicted ordination distances between each state [171]. A perfect ordination will result if the original ordination from the distance matrix matches the predicted ordination distances. Sammon's projection minimizes the distances between the original and predicted ordinations resulting in the Sammon stress [172]. The stress is a measure of goodness of fit and is iteratively calculated through recalculation of the ordination matrix and the predicted

distances until a specified tolerance is achieved or the minimum stress is found. The lowest stress reflects the ordination that summarizes the rank ordering of distances among the samples.

BIBLIOGRAPHY

1. Cox, N.J. and K. Subbarao, *Influenza.* Lancet, 1999. **354**(9186): p. 1277-82.

2. Cox NJ, S.K., *Global Epidemiology of Influenza.* Annual Review of Medicine, 2000. **51**: p. 407-421.

3. Nicholson, K.G., J.M. Wood, and M. Zambon, *Influenza.* Lancet, 2003. **362**(9397): p. 1733-45.

4. Webster, R.G., et al., *Evolution and ecology of influenza A viruses.* Microbiol Rev, 1992. **56**(1): p. 152-79.

5. Zimmer, S.M. and D.S. Burke, *Historical perspective--Emergence of influenza A (H1N1) viruses.* N Engl J Med, 2009. **361**(3): p. 279-85.

6. Lessler, J., et al., *Incubation periods of acute respiratory viral infections: a systematic review.* Lancet Infect Dis, 2009. **9**(5): p. 291-300.

7. Nicholson KG, W.R., Hay AJ, ed. *Textbook of Influenza*. 1998, Blackwell Science: Oxford.

8. Mandell GL, D.R., Bennett JE, *Principles and practice of infectious disease*. 2005, New York: Elsevier/Churchill Livingstone.

9. Aoki, F.Y. and G. Boivin, *Influenza virus shedding: excretion patterns and effects of antiviral treatment.* J Clin Virol, 2009. **44**(4): p. 255-61.

10. Sato, M., et al., *Viral shedding in children with influenza virus infections treated with neuraminidase inhibitors.* Pediatr Infect Dis J, 2005. **24**(10): p. 931-2.

11. CDC. *Flu Activity and Surveillance*. [cited 2009 August 27]; Available from: http://www.cdc.gov/flu/weekly/fluactivity.htm

12. Thompson, W.W., et al., *Mortality associated with influenza and respiratory syncytial virus in the United States.* JAMA, 2003. **289**(2): p. 179-86.

13. Sullivan, K.M., A.S. Monto, and I.M. Longini, Jr., *Estimates of the US health impact of influenza.* Am J Public Health, 1993. **83**(12): p. 1712-6.

14. Adams PF, H.G., Marano MA, *Current estimates from the national health interview survey, 1996*, N.C.f.H. Statistics, Editor. 1999, Vital Health Statistics: Washington, DC. p. 203.

15. Glezen, W., *Emerging Infections: Pandemic Influenza.* Epidemiologic Reviews, 1996. **18**(1): p. 64-76.

16. Thompson, W.W., et al., *Influenza-associated hospitalizations in the United States.* JAMA, 2004. **292**(11): p. 1333-40.

17. Neuzil, K.M., et al., *The effect of influenza on hospitalizations, outpatient visits, and courses of antibiotics in children.* N Engl J Med, 2000. **342**(4): p. 225-31.

18. Neuzil, K.M., et al., *Impact of influenza on a cute cardiopulmonary hospitalizations in pregnant women.* Am J Epidemiol, 1998. **148**(11): p. 1094-102.

19. Simonsen, L., et al., *The impact of influenza epidemics on mortality: introducing a severity index.* Am J Public Health, 1997. **87**(12): p. 1944-50.

20. Dushoff, J., et al., *Mortality due to influenza in the United States--an annualized regression approach using multiple-cause mortality data.* Am J Epidemiol, 2006. **163**(2): p. 181-7.

21. WHO. *Influenza (fact sheet no. 211)*. 2003 [cited 2011 February 20]; Available from: http://www.who.int/mediacentre/factsheets/fs211/en/.

22. Simonsen, L., et al., *Pandemic versus epidemic influenza mortality: a pattern of changing age distribution.* J Infect Dis, 1998. **178**(1): p. 53-60.

23. Noble, G., *Epidemiological and c linical aspects of influenza*, in *Basic and A pplied Influenza Research*, A. Beare, Editor. 1982, CRC: Boca Raton, FL. p. 11-50.

24. Miller, E., et al., *Incidence of 2009 pandemic influenza A H1N1 infection in England: a cross-sectional serological study.* Lancet, 2010. **375**(9720): p. 1100-8.

25. Zimmer, S.M., et al., *Seroprevalence following the second wave of Pandemic 2009 H1N1 influenza in Pittsburgh, PA, USA.* PLoS One, 2010. **5**(7): p. e11601.

26. Viboud, C., et al., *Risk factors of influenza transmission in households.* Br J Gen Pract, 2004. **54**(506): p. 684-9.

27. Longini, I.M., Jr., et al., *Estimating household and community transmission parameters for influenza.* Am J Epidemiol, 1982. **115**(5): p. 736-51.

28. Cauchemez, S., et al., *Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza.* Proc Natl Acad Sci U S A, 2011. **108**(7): p. 2825-2830.

29. Cauchemez, S., et al., *Estimating the impact of school closure on influenza transmission from Sentinel data.* Nature, 2008. **452**(7188): p. 750-4.

30. Cowling, B.J., et al., *Effects of school closures, 2008 winter influenza season, Hong Kong.* Emerg Infect Dis, 2008. **14**(10): p. 1660-2.

31. Ferguson, N.M., et al., *Strategies for mitigating an influenza pandemic.* Nature, 2006. **442**(7101): p. 448-52.

32. Gordon, A., et al., *Prevalence and seasonality of influenza-like illness in children, Nicaragua, 2005-2007.* Emerg Infect Dis, 2009. **15**(3): p. 408-14.

33. Koch, A., et al., *Risk factors for acute respiratory tract infections in young Greenlandic children.* Am J Epidemiol, 2003. **158**(4): p. 374-84.

34. Proff, R., et al., *Case-based surveillance of influenza hospitalizations during 2004-2008, Colorado, USA.* Emerg Infect Dis, 2009. **15**(6): p. 892-8.

35. Rota, P.A., et al., *Cocirculation of two distinct evolutionary lineages of influenza type B virus since 1983.* Virology, 1990. **175**(1): p. 59-68.

36. Belshe, R.B., *The need for quadrivalent vaccine against seasonal influenza.* Vaccine, 2010. **28 Suppl 4**: p. D45-53.

37. Reichert, T.A., et al., *Influenza and the winter increase in mortality in the United States, 1959-1999.* Am J Epidemiol, 2004. **160**(5): p. 492-502.

38. Greene SK, I.E., Wilson ML, *Patterns of Influenza-associated Mortality among US Elderly by Geographic Region and Virus Subtype, 1968–1998.* American Journal of Epidemiology, 2006. **163**(4): p. 316-326.

39. Finkelman, B.S., 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. **2**(12): p. e1296.

40. Dowell, S.F., *Low attack rate of summertime influenza: could it be the host?* Clin Infect Dis, 2001. **33**(11): p. 1951-2.

41. Dushoff, J., et al., *Dynamical resonance can account for seasonality of influenza epidemics.* Proc Natl Acad Sci U S A, 2004. **101**(48): p. 16915-6.

42. Altizer, S., et al., *Seasonality and the dynamics of infectious diseases.* Ecol Lett, 2006. **9**(4): p. 467-84.

43. Shaman, J., et al., *Absolute humidity and the seasonal onset of influenza in the continental United States.* PLoS Biol, 2010. **8**(2): p. e1000316.

44. Shaman, J., et al., *Absolute humidity and the seasonal onset of influenza in the continental US.* PLoS Curr, 2009. **1**: p. RRN1138.

45. Nelson, M.I., et al., *Molecular epidemiology of A/H3N2 and A/H1N1 influenza virus during a single epidemic season in the United States.* PLoS Pathog, 2008. **4**(8): p. e1000133.

46. Nelson, M.I., et al., *Phylogenetic analysis reveals the global migration of seasonal influenza A viruses.* PLoS Pathog, 2007. **3**(9): p. 1220-8.

47. Russell, C.A., et al., *The global circulation of seasonal influenza A (H3N2) viruses.* Science, 2008. **320**(5874): p. 340-6.

48. Sakai, T., et al., *Geographic and temporal trends in influenzalike illness, Japan, 1992-1999.* Emerg Infect Dis, 2004. **10**(10): p. 1822-6.

49. Onozuka, D. and A. Hagihara, *Spatial and temporal dynamics of influenza outbreaks.* Epidemiology, 2008. **19**(6): p. 824-8.

50. 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 Infect Dis, 2007. **7**: p. 141.

51. Grenfell, B.T., O.N. Bjornstad, and J. Kappey, *Travelling waves and spatial hierarchies in measles epidemics.* Nature, 2001. **414**(6865): p. 716-23.

52. Liebhold A, Koenig WD, and O.N. Bjornstad, *Spatial synchrony in population dynamics.* Annual Review Ecology Evolution Systems, 2004. **35**: p. 467-490.

53. Bjornstad, O.N., R.A. Ims, and X. Lambin, *Spatial population dynamics: analyzing patterns and processes of population synchrony.* Trends Ecol Evol, 1999. **14**(11): p. 427-432.

54. Bolker, B.M. and B.T. Grenfell, *Impact of vaccination on the spatial correlation and persistence of measles dynamics.* Proc Natl Acad Sci U S A, 1996. **93**(22): p. 12648-53.

55. Rizzo, C., et al., *Trends for influenza-related deaths during pandemic and epidemic seasons, Italy, 1969-2001.* Emerg Infect Dis, 2007. **13**(5): p. 694-9.

56. Viboud, C., et al., *Synchrony, waves, and spatial hierarchies in the spread of influenza.* Science, 2006. **312**(5772): p. 447-51.

57. Wenger, J.B. and E.N. Naumova, *Seasonal synchronization of influenza in the United States older adult population.* PLoS One, 2010. **5**(4): p. e10187.

58. Lowen, A.C., et al., *Influenza virus transmission is dependent on relative humidity and temperature.* PLoS Pathog, 2007. **3**(10): p. 1470-6.

59. Alonso, W.J., et al., *Seasonality of influenza in Brazil: a traveling wave from the Amazon to the subtropics.* Am J Epidemiol, 2007. **165**(12): p. 1434-42.

60. Ferguson, N.M., et al., *Strategies for containing an emerging influenza pandemic in Southeast Asia.* Nature, 2005. **437**(7056): p. 209-14.

61. Grais, R.F., J.H. Ellis, and G.E. Glass, *Assessing the impact of airline travel on the geographic spread of pandemic influenza.* Eur J Epidemiol, 2003. **18**(11): p. 1065-72.

62. Grais, R.F., et al., *Modeling the spread of annual influenza epidemics in the U.S.: the potential role of air travel.* Health Care Manag Sci, 2004. **7**(2): p. 127-34.

63. Germann, T.C., et al., *Mitigation strategies for pandemic influenza in the United States.* Proc Natl Acad Sci U S A, 2006. **103**(15): p. 5935-40.

64. Bonabeau, E., L. Toubiana, and A. Flahault, *The geographical spread of influenza.* Proc Biol Sci, 1998. **265**(1413): p. 2421-5.

65. Mugglin, A.S., N. Cressie, and I. Gemmell, *Hierarchical statistical modelling of influenza epidemic dynamics in space and time.* Stat Med, 2002. **21**(18): p. 2703-21.

66. WHO. *WHO Global Influenza Surveillance Network.* [cited 2011 February 9]; Available from: http://www.who.int/csr/disease/influenza/surveillance/en/.

67. Kitler, M.E., P. Gavinio, and D. Lavanchy, *Influenza and the work of the World Health Organization.* Vaccine, 2002. **20 Suppl 2**: p. S5-14.

68. Network, E.I.S. [cited 2011 February 9]; Available from: www.eiss.org.

69. Sueker, J.J., et al., *Global Infectious Disease Surveillance at DoD Overseas Laboratories, 1999-2007.* Am J Trop Med Hyg, 2010. **82**(1): p. 23-7.

70. Thompson , W., L. Comanor, and S. DK, *Epidemiology of Seasonal Influenza: Use of Surveillance Data and Statistical Models to Estimate Burden of Disease.* Journal of Infectious Disease, 2006. **2006**(Suppl 2): p. S92-91.

71. Health, P.D.o. [cited 2009 May 14]; Available from: http://www.dsf.health.state.pa.us/health/cwp/view.asp?Q=230681

72. Heffernan, R., et al., *New York City syndromic surveillance systems.* MMWR Morb Mortal Wkly Rep, 2004. **53 Suppl**: p. 23-7.

73. Wagner, M.M., et al., *Syndrome and outbreak detection using chief-complaint data--experience of the Real-Time Outbreak and Disease Surveillance project.* MMWR Morb Mortal Wkly Rep, 2004. **53 Suppl**: p. 28-31.

74. *Assessment of the effectiveness of the 2003-04 influenza vaccine among children and adults--Colorado, 2003.* MMWR Morb Mortal Wkly Rep, 2004. **53**(31): p. 707-10.

75. Nichol, K.L. and J.J. Treanor, *Vaccines for seasonal and pandemic influenza.* J Infect Dis, 2006. **194 Suppl 2**: p. S111-8.

76. Belshe, R.B., *The burden of influenza and strategies for prevention.* Manag Care, 2007. **16**(8 Suppl 8): p. 2-6; discussion 17-9.

77. Nichol, K.L., et al., *Effectiveness of live, attenuated intranasal influenza virus vaccine in healthy, working adults: a randomized controlled trial.* JAMA, 1999. **282**(2): p. 137-44.

78. Hayden, F.G. and A.T. Pavia, *Antiviral management of seasonal and pandemic influenza.* J Infect Dis, 2006. **194 Suppl 2**: p. S119-26.

79. Bright, R.A., et al., *Adamantane resistance among influenza A viruses isolated early during the 2005-2006 influenza season in the United States.* JAMA, 2006. **295**(8): p. 891-4.

80. *Flu.gov*. [cited 2011 February 9]; Available from: www.flu.gov.

81. Organization, W.H., (WHO), and W. Group, *Nonpharmaceutical Interventions for Pandemic Influenza, International Measures.* Emerg Infect Dis, 2006. **12**: p. 81-87.

82. Organization, W.H., WHO, and W. Group, *Nonpharmaceutical Interventions for Pandemic Influenza, National and Community Measures.* Emerg Infect Dis, 2006. **12**(88-94).

83. Lee, T., et al., *Selected nonvaccine interventions to prevent infectious acute respiratory disease.* Am J Prev Med, 2005. **28**(3): p. 305-16.

84. Stebbins, S., J.S. Downs, and C.J. Vukotich, Jr., *Using nonpharmaceutical interventions to prevent influenza transmission in elementary school children: parent and teacher perspectives.* J Public Health Manag Pract, 2009. **15**(2): p. 112-7.

85. Stebbins, S., J.H. Stark, and C.J. Vukotich, Jr., *Compliance with a multilayered nonpharmaceutical intervention in an urban elementary school setting.* J Public Health Manag Pract, 2010. **16**(4): p. 316-24.

86. Cowling, B.J., et al., *Preliminary findings of a randomized trial of non-pharmaceutical interventions to prevent influenza transmission in households.* PLoS One, 2008. **3**(5): p. e2101.

87. Stebbins, S., Cummings DAT, Dato V, Eng H, Mitruka K, RInaldo C, Roth L, Stark JH, Thompson W, Vukotich C, Wagner M, Wisniewski SR, Burke DS, *Reductions in the incidence of influenza A associated with use of hand sanitizer and cough hygiene in schools: a randomized controlled trial.* In Press, 2011.

88. Lee, V.J., D.C. Lye, and A. Wilder-Smith, *Combination strategies for pandemic influenza response - a systematic review of mathematical modeling studies.* BMC Med, 2009. **7**: p. 76.

89. Crighton, E.J., et al., *An exploratory spatial analysis of pneumonia and influenza hospitalizations in Ontario by age and gender.* Epidemiol Infect, 2007. **135**(2): p. 253-61.

90. Viboud, C., et al., *Influenza epidemics in the United States, France, and Australia, 1972-1997.* Emerg Infect Dis, 2004. **10**(1): p. 32-9.

91. Bureau, U.C. [cited 2009 June 11]; Available from: http://factfinder.census.gov/home/saff/main.html?_lang=en.

92. Assuncao, R.M. and E.A. Reis, *A new proposal to adjust Moran's I for population density.* Stat Med, 1999. **18**(16): p. 2147-62.

93. Waller LA, G.C., *Applied Spatial Statistics for Public Health Data*. Wiley Series in Probability and Statistics. 2004, Hoboken: Wiley. 494.

94. Anselin, L., *Local indicators of spatial association-LISA.* Geographical Analysis, 1995. **27**: p. 93-116.

95. Transportation, P.D.o. *Pennsylvania Highway Statistics*. [cited 2010 August 11]; Available from: http://www.dot.state.pa.us.

96. *Area Resource File (ARF)*, H.R.a.S.A. US Department of Health and Human Services, Bureau of Health Professions, Editor. 2008: Rockville, MD.

97. USGS. Naltional Elevation Dataset [cited 2010 August 11]; Available from: http://geonames.usgs.gov/pls/gnispublic/.

98. Gesch, D.B., *The National Elevation Dataset*, in *Digital Elevation Model Technologies and Applications: The DEM Users Manual*, D. Maune, Editor. 2007, Maryland: Bethesda. p. 99-118.

99. Gesch, D., Oimoen, M., Greenlee S., Nelson, C. Steuck, M., Tyler D. , *The National Elevation Dataset: Photogrammetric Engineering and Remote Sensing.* 2002. **68**(1): p. 5-11.

100. *Update: influenza activity--United States and worldwide, 2003-04 season, and composition of the 2004-05 influenza vaccine.* MMWR Morb Mortal Wkly Rep, 2004. **53**(25): p. 547-52.

101. *Update: Influenza activity--United States and worldwide, 2004-05 season.* MMWR Morb Mortal Wkly Rep, 2005. **54**(25): p. 631-4.

102. *Influenza activity--United States and worldwide, 2007-08 season.* MMWR Morb Mortal Wkly Rep, 2008. **57**(25): p. 692-7.

103. CDC. *FluView*. [cited 2010 August 8]; Available from: http://www.cdc.gov.flu/weekly/weeklyarchives2008-2009/weekly15.htm

104. Simonsen, L., et al., *Impact of influenza vaccination on seasonal mortality in the US elderly population.* Arch Intern Med, 2005. **165**(3): p. 265-72.

105. Govaert, T.M., et al., *The efficacy of influenza vaccination in elderly individuals. A randomized double-blind placebo-controlled trial.* JAMA, 1994. **272**(21): p. 1661-5.

106. Jia, H., et al., *Monitoring county-level vaccination coverage during the 2004-2005 influenza season.* Am J Prev Med, 2006. **31**(4): p. 275-280.

107. Merrill, R.M. and J.D. Beard, *Influenza vaccination in the United States, 2005-2007.* Med Sci Monit, 2009. **15**(7): p. PH92-100.

108. *Behavioral Risk Factor Surveillance System Survey Data*, C.f.D.C.a. Prevention, Editor. 1995-2008: Atlanta, Georgia.

109. Viboud, C., et al., *Association of influenza epidemics with global climate variability.* Eur J Epidemiol, 2004. **19**(11): p. 1055-9.

110. Lowen, A.C., et al., *High temperature (30 degrees C) blocks aerosol but not contact transmission of influenza virus.* J Virol, 2008. **82**(11): p. 5650-2.

111. Shaman, J. and M. Kohn, *Absolute humidity modulates influenza survival, transmission, and seasonality.* Proc Natl Acad Sci U S A, 2009. **106**(9): p. 3243-8.

112. Shaman, J., et al., *Absolute humidity and the seasonal onset of influenza in the continental United States.* PLoS Biol. **8**(2): p. e1000316.

113. Tang, J.W., et al., *Comparison of the incidence of influenza in relation to climate factors during 2000-2007 in five countries.* J Med Virol, 2010. **82**(11): p. 1958-65.

114. Tang, J.W., et al., *Incidence of common respiratory viral infections related to climate factors in hospitalized children in Hong Kong.* Epidemiol Infect, 2010. **138**(2): p. 226-35.

115. Chew, F.T., et al., *Seasonal trends of viral respiratory tract infections in the tropics.* Epidemiol Infect, 1998. **121**(1): p. 121-8.

116. de Arruda, E., et al., *Acute respiratory viral infections in ambulatory children of urban northeast Brazil.* J Infect Dis, 1991. **164**(2): p. 252-8.

117. *Notices: Electronic Disease Surveillance System*. Available from: http://www.pabulletin.com/secure/data/vol33/33-20/941.html.

118. Weycker, D., et al., *Population-wide benefits of routine vaccination of children against influenza.* Vaccine, 2005. **23**(10): p. 1284-93.

119. Vora, A., D.S. Burke, and D.A. Cummings, *The impact of a physical geographic barrier on the dynamics of measles.* Epidemiol Infect, 2008. **136**(5): p. 713-20.

120. Cummings, D.A., et al., *Travelling waves in the occurrence of dengue haemorrhagic fever in Thailand.* Nature, 2004. **427**(6972): p. 344-7.

121. Grenfell, B.T., Bolker, B.M, *Cities and villages: infection hierarchies in a measles metapopulation.* Ecology Letters, 1998. **1**: p. 63-70.

122. Wladimir J. Alonso, C.V., Lone Simonsen, Eduardo W. Hirano, Luciane Z. and a.M.A.M. Daufenbach, *Seasonality of Influenza in Brazil: A Traveling Wave from the Amazon to the Subtropics.* American Journal of Epidemiology, 2007. **165**(12): p. 1434-1442.

123. Brownstein, J.S., C.J. Wolfe, and K.D. Mandl, *Empirical evidence for the effect of airline travel on inter-regional influenza spread in the United States.* PLoS Med, 2006. **3**(10): p. e401.

124. Xia, Y., O.N. Bjornstad, and B.T. Grenfell, *Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics.* Am Nat, 2004. **164**(2): p. 267-81.

125. Bharti, N., et al., *Measles on the edge: coastal heterogeneities and infection dynamics.* PLoS One, 2008. **3**(4): p. e1941.

126. Cliff, A.D., Haggett, P, Smallman-Raynor, M, *Measles: An Historical Geography of a Major Human Viral Disease from Global Expansion to Local Retreat, 1840-1990.* 1993, Oxford: Blackwell.

127. Erlander, S., Stewart NF., *The Gravity Model in Transportation Analysis -- Theory and Extensions: Topics in Transportation* Topics in Transportation 1990, Bristol: J.W. Arrowsmith Ltd.

128. Census. *United States Census*. [cited 2010 September 23]; Available from: http://www.census.gov/main/www/cen2000.html.

129. Census. *County-To-County Worker Flow Files*. [cited 2010 September 23]; Available from: http://www.census.gov/population/www/cen2000/commuting.html.

130. Grassly, N.C., C. Fraser, and G.P. Garnett, *Host immunity and synchronized epidemics of syphilis across the United States.* Nature, 2005. **433**(7024): p. 417-21.

131. Bjørnstad, O.N., Stenseth, N.C. and Saitoh, T. , *Synchrony and scaling in dynamics of voles and mice in northern Japan,.* Ecology, 1999. **80**: p. 622–637.

132. P. Legendre, L.L., *Numerical Ecology*. 1998, Amsterdam: Elsevier.

133. *Spatial Analysis in Ecology*. [cited 2010 April 1]; Available from: http://www.nceas.ucsb.edu/files/scicomp/Dloads/SpatialAnalysisEcologists/SpatialEcologyMantelTest.pdf.

134. CDC. *Behavioral Risk Factor Surveillance System*. [cited 2010 January 4]; Available from: http://www.cdc.gov/brfss/.

135. *U.S. Census Bureau*. [cited 2011 February 9]; Available from: http://www.census.gov/history/.

136. Thiemann, C., et al., *The structure of borders in a small world.* PLoS One, 2010. **5**(11): p. e15422.

137. Warden, P. *How to split up the US*. [cited 2011 February 9]; Available from: http://petewarden.typepad.com/searchbrowser/2010/02/how-to-split-up-the-us.html.

138. *Oxford Dictionary*. [cited 2011 February 9]; Available from: http://www.oxforddictionaries.com/definition/ontology?view=uk.

139. *Logic and Ontology*. [cited 2011 February 9]; Available from: http://plato.stanford.edu/entries/logic-ontology/.

140. Simon, H., *The Science of the Artificial*. Third ed. 1996, Cambridge: The MIT Press. 232.

141. McCutchan, F.E., *Global epidemiology of HIV*. J Med Virol, 2006. **78 Suppl 1**: p. S7-S12.

142. Cummings, D.A., et al., *Improved measles surveillance in Cameroon reveals two major dynamic patterns of incidence.* Int J Infect Dis, 2006. **10**(2): p. 148-55.

143. Sniadack, D.H., et al., *Potential interventions for the prevention of childhood pneumonia: geographic and temporal differences in serotype and serogroup distribution of sterile site pneumococcal isolates from children--implications for vaccine strategies.* Pediatr Infect Dis J, 1995. **14**(6): p. 503-10.

144. Scott, J.A., et al., *Serogroup-specific epidemiology of Streptococcus pneumoniae: associations with age, sex, and geography in 7,000 episodes of invasive disease.* Clin Infect Dis, 1996. **22**(6): p. 973-81.

145. Nielsen, S.V. and J. Henrichsen, *Capsular types of Streptococcus pneumoniae isolated from blood and CSF during 1982-1987.* Clin Infect Dis, 1992. **15**(5): p. 794-8.

146. Labov, W., S. Ash, and C. Boberg, *The Atlas of North American English*. 2006, Berlin: Mouton de Gruyter. 318.

147. Labov, W. *Pursuing the Cascade Model*. 2002 [cited 2011 February 9]; Available from: www.ling.upenn.edu/~wlabov/Papers/PCM.html.

148. Callary, C., *Phonological change and the development of an urban dialect in Illinois.* Language in Society, 1975. **23**: p. 1-24.

149. Trudgill, P., *Linguistic change and diffusion: description and explanation in sociolinguistic dialect geography.* Language in Society, 1974. **3**(215-246).

150. Eisenstein, J., et al. *A latent variable model for geographical lexical variation*. in *Proceedings of EMNLP*. 2010.

151. Campbell, M. and G. Plumb. *The great "pop" vs "soda" controversy*. 2002 [cited 2011 February 11]; Available from: http://www.popvssoda.com.

152. Brockmann, D., L. Hufnagel, and T. Geisel, *The scaling laws of human travel.* Nature, 2006. **439**(7075): p. 462-5.

153. Brockmann, D., *Following the Money*, in *Physics World*. 2010. p. 31-34.

154. WHO. *Public Health Surveillance*. [cited 2011 March 29].

155. *Mekong Basin Disease Surveillance* [cited 2011 February 9]; Available from: http://www.mbdsoffice.com/index_2008.php.

156. Gaschen, B., et al., *Diversity considerations in HIV-1 vaccine selection.* Science, 2002. **296**(5577): p. 2354-60.

157. McKnight, A. and M.M. Aasa-Chapman, *Clade specific neutralising vaccines for HIV: an appropriate target?* Curr HIV Res, 2007. **5**(6): p. 554-60.

158. Slobod, K.S., et al., *Clade, Country and Region-specific HIV-1 Vaccines: Are they necessary?* AIDS Res Ther, 2005. **2**(1): p. 3.

159. Whitney, C.G. and L.K. Pickering, *The potential of pneumococcal conjugate vaccines for children.* Pediatr Infect Dis J, 2002. **21**(10): p. 961-70.

160. CDC, *Epidemiology and Prevention of Vaccine Preventable Diseases*. 11 e d. 2009: Public Health Foundation.

161. Checkley, W., et al., *Effect of El Nino and am bient temperature on hos pital admissions for diarrhoeal diseases in Peruvian children.* Lancet, 2000. **355**(9202): p. 442-50.

162. Chretien, J.P., et al., *Drought-associated chikungunya emergence along coastal East Africa.* Am J Trop Med Hyg, 2007. **76**(3): p. 405-7.

163. Day, J.F. and J. Shaman, *Severe winter freezes enhance St. Louis encephalitis virus amplification and e pidemic transmission in peninsular Florida.* J Med Entomol, 2009. **46**(6): p. 1498-506.

164. Longini, I.M., Jr., A.S. Monto, and J.S. Koopman, *Statistical procedures for estimating the community probability of illness in family studies: rhinovirus and i nfluenza.* Int J Epidemiol, 1984. **13**(1): p. 99-106.

165. Longini, I.M., Jr., et al., *Statistical inference for infectious diseases. Risk-specific household and c ommunity transmission parameters.* Am J Epidemiol, 1988. **128**(4): p. 845-59.

166. Nelson, M.I., et al., *The origin and gl obal emergence of adamantane resistant A/H3N2 influenza viruses.* Virology, 2009. **388**(2): p. 270-8.

167. Nelson, M.I., et al., *Phylogeography of the spring and f all waves of the H1N1/09 pandemic influenza virus in the United States.* J Virol, 2011. **85**(2): p. 828-34.

168. Kruskal, J.B. and M. Wish, *Multidimensional Scaling*. Quantitative Application in the Social Sciences, ed. S. University. 1978, Beverly Hills and London: Sage Publications.

169. Erlander, S. and N.F. Stewart, *The Gravity Model in Transportation Analysis -- Theory and Extensions*. Topic in Transportation. 1990, Utrecht: VSP.

170. Stark, J., *Spatial and T emporal Dynamics of Influenza*, in *Epidemiology*. 2011, University of Pittsburgh: Pittsburgh.

171. Holland, S. *NON-METRIC MULTIDIMENSIONAL SCALING (MDS)*. [cited 2011 March 1]; Available from: http://www.uga.edu/strata/software/pdf/mdsTutorial.pdf.

172. Sammon, J., *A nonlinear mapping for data structure analysis.* IEEE Transactions on Computers, 1969. **18**: p. 401–409.