Disparities in influenza mortality and transmission related to sociodemographic factors within Chicago in the pandemic of 1918

Kyra H. Grantza,b,1, Madhura S. Ranec,c,1, Henrik Saljéd,e, Gregory E. Glassb,f, Stephen E. Schachterle,9, and Derek A. T. Cummingsa,b,d,2

aDepartment of Biology, University of Florida, Gainesville, FL 32611; bEmerging Pathogens Institute, University of Florida, Gainesville, FL 32611; cDepartment of Epidemiology, University of Washington, Seattle, WA 98195; dDepartment of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205; eMathematical Modelling Unit, Institut Pasteur, Paris 75015, France; fDepartment of Geography, University of Florida, Gainesville, FL 32611; and Epidemiology, Worldwide Safety & Regulatory, Pfizer Inc., New York, NY 10017

Edited by Burton H. Singer, University of Florida, Gainesville, FL, and approved September 29, 2016 (received for review August 10, 2016)

Social factors have been shown to create differential burden of influenza across different geographic areas. We explored the relationship between potential aggregate-level social determinants and mortality during the 1918 influenza pandemic in Chicago using a historical dataset of 7,971 influenza and pneumonia deaths. Census tract-level social factors, including rates of illiteracy, homeownership, unemployment, and age, were assessed as predictors of pandemic mortality in Chicago. Poisson models fit with generalized estimating equations (GEEs) were used to estimate the association between social factors and the risk of influenza and pneumonia mortality. The Poisson model showed that influenza and pneumonia mortality increased, on average, by 32.2% for every 10% increase in illiteracy rate adjusted for population density, homeownership, unemployment, and age. We also found a significant association between transmissibility and population density, illiteracy, and unemployment but not homeownership. Lastly, analysis of the point locations of reported influenza and pneumonia deaths revealed fine-scale spatiotemporal clustering. This study shows that living in census tracts with higher illiteracy rates increased the risk of influenza and pneumonia mortality during the 1918 influenza pandemic in Chicago. Our observation that disparities in structural determinants of neighborhood-level health lead to disparities in influenza incidence in this pandemic suggests that disparities and their determinants should remain targets of research and control in future pandemics.

Significance

The pervasiveness of influenza among humans and its rapid spread during pandemics create a false sense that all humans are affected equally. In this work, we show that neighborhood-level social determinants were associated with greater burdens of pandemic influenza in 1918 and several other diseases in a major US city. We show that literacy, homeownership, and unemployment were associated with cumulative influenza mortality as well as measures of the speed of transmission using a unique dataset describing the home location and week of death of individuals who died during the influenza pandemic in 1918. Our results suggest that, similar to other infectious diseases, social disparities should be a focus of research and public health response in future pandemics.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the Dryad Digital Repository, https://datadryad.org (10.5061/dryad.48rv3).

See Commentary on page 13557.

K.H.G. and M.S.R. contributed equally to this work.

1To whom correspondence should be addressed. Email: datal@ufl.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1612838113/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1612838113

PNAS | November 29, 2016 | vol. 113 | no. 48 | 13839–13844
Fig. 2 shows the weekly point pattern mortality overlaid on a map of illiteracy rates by census tracts. The first deaths were observed in census tracts with higher illiteracy rates. As the epidemic progressed, the disease spread to the neighboring tracts but remained concentrated in the high-illiteracy areas.

A Poisson model fit with GEE was used to estimate average mortality per census tract with robust variance. The clustering variable in the model was the census tract, and the correlation in data in each census tract was assumed to be unstructured. Assuming correlation within census tracts, we analyzed association between mortality and illiteracy rate, homeownership, unemployment, and population density. The model also accounted for any unknown correlations between the outcomes and was offset by the total underlying population of each census tract. The multivariate models were also adjusted for age, because pandemic mortality was known to vary by age (1).

The univariate and multivariate analyses’ estimates for risk ratios of mortality rate per census tract are presented in Table 1. Univariate analysis found that illiteracy rate, homeownership, and unemployment were independently statistically associated with mortality. The association between population density and incidence of mortality was nonsignificant. In the multivariate model, for every 10% increase in illiteracy rate within a given census tract, mortality increased by 32.2% [95% confidence interval (95% CI) = 22.2%, 43.0%]. Mortality incidence decreased by 33.7% (95% CI = 24.5%, 41.8%) per 10% increase in homeownership, 19.6% (95% CI = 13.8%, 25.0%) per 10% increase in unemployment rate, and 4.3% (95% CI = 1.1%, 5.5%) per 10% increase in population density. All sociodemographic predictors of mortality were statistically significant in the multivariate model. Mortality was statistically significantly associated with the number of individuals in each census tract over the age of 45 y old but not associated with the number of individuals in other age classes (Table S2). Regression results were insensitive to inclusion of week of epidemic in the analysis (Table S3).

To explore whether the Poisson model accounted for spatial correlation in the data, Moran’s I (MI) statistic of autocorrelation was calculated for model standardized residuals. The MI estimate for model residuals was statistically significant (MI = 0.09, P < 0.001), although significantly reduced compared with the MI estimate for the raw mortality rates (MI = 0.14, P < 0.001), suggesting that the model accounted for some of the existing spatial autocorrelation.

We investigated the relationship between illiteracy and other diseases in earlier years in Chicago (SI Text) and found that incidence in the central and southcentral regions of the city, whereas the northern, western, and southernmost areas of the city experienced a lower disease burden (Fig. 1). The average case fatality rate city-wide was 17.5% during the pandemic period, although this rate is likely to be an overestimate because of underreporting of nonfatal cases.

Relationship with Sociodemographic Factors. Fig. 2 shows the weekly point pattern mortality overlaid on a map of illiteracy rates by census tracts. The first deaths were observed in census tracts with higher illiteracy rates. As the epidemic progressed, the disease spread to the neighboring tracts but remained concentrated in the high-illiteracy areas.

A Poisson model fit with GEE was used to estimate average mortality per census tract with robust variance. The clustering variable in the model was the census tract, and the correlation in data in each census tract was assumed to be unstructured. Assuming correlation within census tracts, we analyzed association between mortality and illiteracy rate, homeownership, unemployment, and population density. The model also accounted for any unknown correlations between the outcomes and was offset by the total underlying population of each census tract. The multivariate models were also adjusted for age, because pandemic mortality was known to vary by age (1).

The univariate and multivariate analyses’ estimates for risk ratios of mortality rate per census tract are presented in Table 1. Univariate analysis found that illiteracy rate, homeownership, and unemployment were independently statistically associated with mortality. The association between population density and incidence of mortality was nonsignificant. In the multivariate model, for every 10% increase in illiteracy rate within a given census tract, mortality increased by 32.2% [95% confidence interval (95% CI) = 22.2%, 43.0%]. Mortality incidence decreased by 33.7% (95% CI = 24.5%, 41.8%) per 10% increase in homeownership, 19.6% (95% CI = 13.8%, 25.0%) per 10% increase in unemployment rate, and 4.3% (95% CI = 1.1%, 5.5%) per 10% increase in population density. All sociodemographic predictors of mortality were statistically significant in the multivariate model. Mortality was statistically significantly associated with the number of individuals in each census tract over the age of 45 y old but not associated with the number of individuals in other age classes (Table S2). Regression results were insensitive to inclusion of week of epidemic in the analysis (Table S3).

To explore whether the Poisson model accounted for spatial correlation in the data, Moran’s I (MI) statistic of autocorrelation was calculated for model standardized residuals. The MI estimate for model residuals was statistically significant (MI = 0.09, P < 0.001), although significantly reduced compared with the MI estimate for the raw mortality rates (MI = 0.14, P < 0.001), suggesting that the model accounted for some of the existing spatial autocorrelation.

We investigated the relationship between illiteracy and other diseases in earlier years in Chicago (SI Text) and found that
illiteracy was significantly positively associated with all-cause mortality, diphtheria, whooping cough, tuberculosis, and pneumonia (Table S4). Illiteracy was negatively associated with incidence of scarlet fever and measles.

We next used a GEE model to assess the relationship between pandemic mortality rates and all-cause mortality and disease burden at the ward level. Within Chicago, pandemic mortality rates were negatively associated with rates of scarlet fever, measles, and whooping cough infection and positively associated with rates of tuberculosis and pneumonia (Table S5). Pandemic mortality was significantly linearly associated with all-cause mortality in Chicago and three other cities (Fig. S1).

Simulations of Spring Outbreak. There were reports of small influenza outbreaks in Chicago in March of 1918 (24). Such herald waves were common around the world but varied substantially in magnitude. In Geneva, the first pandemic influenza wave occurred in July of 1918 and had an estimated $R_t$ of less than one-half of that for the fall wave (25). However, two other studies found that transmissibility was significantly higher in the spring wave of three Scandinavian cities, although lower case fatality rates in the spring meant the vast majority of deaths still occurred in the later fall wave (26, 27).

Although the spring wave in Chicago seems to have been isolated to a handful of office buildings and industrial establishments (24), it is possible that this herald wave targeted certain social classes more than others, leaving disproportionately high levels of immune individuals in some census tracts. This underlying population immunity would then impact the spread of the pandemic wave in the fall and threaten our inferences about the impact of sociodemographic disparities in the fall wave. To test this hypothesis, we simulated, with a stochastic, fixed time step susceptible–infectious–recovered (SIR) model, a spring wave in which all individuals in a given census tract were initially susceptible and then, a fall wave in which only individuals who escaped the spring wave were susceptible. We tested multiple initial conditions and relationships between transmissibility and illiteracy and averaged the results of 1,000 simulations in each census tract. A univariate GEE model predicted the relationship between the final outbreak size of the fall wave and illiteracy.

In most cases where transmissibility was defined to be positively associated with illiteracy in both waves, there was still a strong and significant positive association between illiteracy rate and influenza incidence in the fall wave (Fig. S2 and Table S6). When transmissibility was negatively related to illiteracy in both waves (that is, the disease targeted higher social status individuals) and attack rates were low, illiteracy was consistently negatively associated with pandemic incidence in the fall. At higher attack rates, however, or when there was no social dependence imposed in the fall wave, the GEE model could incorrectly predict a relationship between illiteracy and pandemic incidence that does not match the defined relationship. However, the coefficients of these falsely predicted associations are much lower than those observed in the data (Table 1).

Table 1. Results of univariate and multivariate GEE regression models of pandemic influenza mortality.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Univariate regression unadjusted RR (95% CI)</th>
<th>Multivariate regression adjusted RR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage illiterate</td>
<td>1.056 (1.048, 1.063)</td>
<td>1.028 (1.020, 1.036)</td>
</tr>
<tr>
<td>Population density (per acre)</td>
<td>1.001 (0.999, 1.002)</td>
<td>0.996 (0.994, 0.997)</td>
</tr>
<tr>
<td>Percentage homeowners</td>
<td>0.971 (0.961, 0.982)</td>
<td>0.960 (0.947, 0.972)</td>
</tr>
<tr>
<td>Percentage unemployed</td>
<td>0.958 (0.952, 0.964)</td>
<td>0.978 (0.972, 0.985)</td>
</tr>
</tbody>
</table>

Age was included but largely nonsignificant (Table S2). RR, risk ratio.

Reproduction Number Estimates and Association with Sociodemographic Factors. Estimates of reproduction number, $R_t$ were calculated for each census tract for the entire 7-wk period (Methods). Because of the spring wave or cases early in the fall possibly inducing some level of immunity in the population, these estimates are likely lower than the true basic reproduction number, $R_0$. The mean $R$ was 1.22 (95% CI = 1.21, 1.23). Table 2 shows the relationship between reproduction number estimates and sociodemographic factors. The strongest correlation was found with population density. $R$ also significantly correlated with illiteracy and counterintuitively, was negatively associated with unemployment. There was no significant relationship between transmissibility and homeownership.

We calculated the reproduction number for each census tract in each week of the epidemic to further explore the effects of illiteracy and population density on transmission at multiple points in the epidemic. Fig. 3A shows mean weekly reproduction number estimates for census tracts grouped by illiteracy rate. As expected, the census tracts with higher rates of illiteracy have greater $R_t$ estimates throughout the peak of the epidemic. There is a statistically significant correlation between illiteracy and reproduction number in weeks 3–6 (Table S7). As the epidemic slowed, the reproduction numbers for all tracts reduced toward one, and the association between illiteracy and transmissibility became weaker and eventually, nonsignificant. Interestingly, when census tracts are grouped by population density, there is not a clear relationship with transmissibility (Fig. 3B). Only in week 2 is the correlation between population density and $R_t$ significant ($\rho = 0.20, P = 0.001$) (Table S7).

Spatial Dependence of Influenza Mortality. By visual inspection, the map of cumulative mortality suggests that there was some spatial clustering of influenza and pneumonia mortality at the census tract level (Fig. 1). We used the spatiotemporal clustering statistic, $\phi(d_1, d_2)$, to characterize spatial and temporal dependence of individual deaths (28, 29). This statistic estimates the probability of any two deaths occurring within the same week and within a spatial window of one another relative to the expectation if spatial clustering and temporal clustering were independent (Methods). Any underlying heterogeneities, like population density and reporting rate, are included in the numerator and denominator of this statistic and therefore, do not bias our estimates.

There is moderately strong spatiotemporal dependence of individual deaths at small spatial scales (Fig. 4). Influenza and pneumonia deaths were more than 1.3 times more likely to occur in the same week and within 100 m and roughly 1.2 times more likely to occur in the same week and within 200 m given the underlying spatial and temporal distributions. There is significant, albeit much weaker, clustering at distances up to 1,500 m, beyond which there is no significant spatiotemporal clustering.

Discussion

This study showed that the 1918 influenza pandemic had strong patterns of spatial clustering and sociodemographic dependence. Prior studies have produced contradictory views on the association between influenza mortality and social status during the
1918 pandemic. Some studies have argued that, because the 1918 H1N1 influenza virus was highly virulent, it indiscriminately killed, regardless of social class (30, 31), whereas some studies have proposed that sociodemographic differences affected mortality during this pandemic at the scale of countries, provinces, and counties (2, 7, 32, 33). This study found that, despite the highly virulent nature of the virus, influenza did not behave in a wholly democratizing fashion at the within-city scale. Rather, there was significant spatial variation in influenza mortality associated with sociodemographic factors, and there is evidence of localized transmission. This study also found that several other infectious diseases were associated with social factors, speaking to overall higher disease burden in areas of lower social status.

Sociodemographic indicators were associated with both the rate at which deaths per week increased (suggesting that transmission was greater in areas of reduced socioeconomic status) and increased cumulative mortality (suggesting increased incidence of infection and/or case fatality rates) during the 1918 pandemic. Note that it is impossible for us to disentangle the impact of these two mechanisms on cumulative mortality. The fact that we observe associations between specific sociodemographic factors, transmissibility, and mortality during 7 wk of the pandemic within a small geographic area allows us to refine some of the hypotheses of mechanisms associating sociodemographic factors and pandemic mortality. Pandemic onset time and physical geography, which may be important in explaining country-wide variance in mortality, vary little within the city of Chicago and are, therefore, unlikely to be important factors in explaining the observed differential mortality. Illiteracy was positively associated with mortality and transmissibility, likely contributing to lower access to medical care as well as decreased awareness and adoption of intervention measures proposed by public health officials, which increases risk of infection and poor clinical outcome (18). The latter explanation may be particularly relevant to the city of Chicago, where strict public health measures, including mandatory quarantines, school closings, and bans on public gatherings, are believed to have significantly reduced the size of the pandemic (34). Underlying health conditions, like chronic obstructive pulmonary disease, tuberculosis, and malnutrition, are known to have a strong social gradient and likely associated with census tracts with lower literacy rates (14, 16). Our findings that illiteracy is significantly and positively associated with ward-level all-cause mortality and incidence of several other infectious diseases in the years before the pandemic speak to the probable lower underlying health status of census tracts with higher illiteracy rates. In fact, the only two diseases that were not positively associated with illiteracy were measles and scarlet fever, which infected the overwhelming majority of the general population in the prevaccination era, and thus, heterogeneity in incidence rate across wards can primarily be attributed to differences in reporting rates.

Rate of homeownership was significantly negatively related to mortality in this study, likely a reflection of the higher income and socioeconomic status of homeowners. More robust measures of socioeconomic status, including poverty indices and per-capita income, were not available in the 1920 census data. Interestingly, it was observed that unemployment, assumed to be a measure of low income and poverty, had a protective effect on influenza mortality and was associated with decreased transmission. This result might be an artifact of the method of quantifying unemployment, which includes individuals who willingly elected not to work and were likely of higher social status (Methods). Unemployment may also result in fewer social interactions and less migration to overcrowded regions, which would reduce risk of infection and death. Finally, the highest mortality during the pandemic was observed in 21–44 y olds, who also make up majority of the working population. The decrease in mortality might reflect the fact that the unemployed population contained disproportionately few members of the age class at the highest risk of mortality compared with the general population.

The range of reproduction numbers across census tracts was 1.09–1.54, which is on the lower end of previous estimates for the 1918 pandemic (35–38). However, basic reproduction numbers ($R_0$) greater than three have been reported on small spatial scales, such as military bases (39), prisons and ships (26), and even in some cities (25). The lower estimates found in this study could be the result of population immunity after a spring outbreak of this pandemic virus in Chicago or the impact of control measures (24, 34).

**Fig. 3.** Effective reproduction estimates by week for (A) illiteracy and (B) population density quartiles. Lines represent mean $R$ at each week for 124 census tracts in each quartile, and shaded regions represent 95% bootstrapped CIs.

**Fig. 4.** Short-term spatial dependence of influenza and pneumonia deaths occurring within the same week, $\phi(d_1, d_2)$. Spatial window $d_2 - d_1$ was held constant at 100 m. Estimates are plotted for the midpoint of each spatial window. The shaded regions represents 95% bootstrapped CIs.
This study observed a significant association between transmissibility and population density at the census tract level, which has not been observed at greater spatial scales (4, 7). The association between \( R \) and population density may be a result of an increased number of effective contacts between individuals in dense population tracts, but the relationship is confounded by the strong associations between population density and other sociodemographic factors and the largely nonsignificant association between population density and \( R \) estimates (Tables S7 and S8). Moreover, the increased transmissibility of influenza in more densely populated census tracts did not translate into significantly higher mortality risk, although population density has been linked to pandemic influenza mortality at greater spatial scales (7, 11, 20).

It is important to note that this study uses influenza and pneumonia mortality data as a proxy for influenza incidence when estimating reproduction numbers. When making inferences about transmissibility, it is preferable to use data on all influenza cases, not just those that resulted in death, to eliminate possible errors from heterogeneity or secular changes in case fatality rates, but this information was not present in the available data. Our transmissibility estimates may be more precise in census tracts with lower socioeconomic status, where there may have been more deaths per influenza infection because of worse preexisting health status and access to healthcare expected in those census tracts. Additionally, it has been shown that temperature and relative humidity, can have a substantial impact on transmissibility (40). September of 1918 was unseasonably cold throughout Illinois, which may have helped facilitate the rapid spread of influenza in the latter half of the month (41). There are no data on humidity from this time period, but it is unlikely that climate conditions would have varied significantly enough by census tract to cause noticeable differences in transmission at such fine spatial scales.

This study characterized the spatial dependence of influenza deaths at the level of individual cases. The spatial analysis indicates that influenza transmission and mortality were highly localized at small spatiotemporal scales. Deaths were more likely to occur within the same week and a few hundred meters than expected given the underlying spatial and temporal clustering, speaking to the importance of neighborhood-level transmission even in a mobile, urban population. Previous studies found significant autocorrelation in influenza mortality across county boroughs in England and Wales (42) and provinces in Spain (20), but this study is unique in showing a spatial dependence in the spread of individual influenza cases during the 1918 pandemic.

There are several limitations to this study. This study focuses on only one wave of pandemic influenza in Chicago. Other studies have shown significant differences in transmissibility and mortality patterns of different waves of pandemic influenza in 1918 (7, 21). The social and spatial dependencies observed in this study are likely to be greater if studied on a state- or country-wide scale. Previous studies have shown that the outbreak spread in a spatiotemporally dependent manner. The results from this study will help city public health authorities decide how to optimally visualize their limited resources to improve control strategies to minimize transmission and mortality.

**Methods**

**Data Collection and Outcome Definitions.** Census tract-level data on demographic characteristics in 1920 were collected from the National Historical Geographic Information System website (22). Influenza mortality data were obtained from maps published in a report by the City of Chicago Department of Health after the pandemic (23). Each paper map showed the point location of the home address of individuals who died from influenza or pneumonia in Chicago each week from September 29, 1918 to November 16, 1918. The maps were scanned and georeferenced to raster files using ArcGIS. Locations of the influenza deaths in 1918 were transferred from the digitized maps to a current vector-based shapefile of Chicago. The report included the location of 7,971 influenza and pneumonia deaths in 496 census tracts in Chicago. There were an additional 59 deaths in three census tracts for which there was no demographic information available that were excluded from analysis. Spatial and statistical analyses were performed using ArcGIS, version 10.1 and R (43).

Association Between Sociodemographic Factors and Mortality. The covariates included in the Poisson model were age, population density (total population divided by total area in acres of each census tract), literacy rate (number of people over 10 y of age who were illiterate divided by the total population in each census tract), proportion of homeowners (number of people who owned houses free and clear, with encumbrance, and with encumbrance status unknown divided by total population in each census tract), and unemployment rate (number of individuals not in the nine most common professions included in the census survey—agriculture/forestry, mining, factory, transportation, trade, public service, professionals, domestic servants, and clerics—divided by total population in each census tract). Age was categorized as in the original census data: less than 1, 1–4, 5–10, 11–16, 16–20, 21–44, and 45 y or older.

Univariate analysis was first performed to determine the association of each covariate independently with mortality. A forward stepwise selection approach was used to identify significant predictors in the multivariate model. Population density and age were included in the multivariate model, regardless of significance, because they are very strong confounders of the relationship between sociodemographic factors and mortality. All \( P \) values were two sided based on 95% significance levels.

**Estimating Transmissibility and Social Determinants of \( R \).** \( R \) was estimated for each census tract by substituting an initial exponential growth rate, \( \lambda \), in an expression derived from the SIR compartmental model

\[
R = 1 + \lambda \frac{1}{7}
\]

We assumed that influenza incidence grew exponentially for 3 wk after the first influenza or pneumonia death was reported and calculated the initial growth.
rate, λ, from the epidemic curve of each census tract. The average infectious period (1/λ, where λ is the mean recovery rate) was assumed to be 2.7 d (47). The 95% CIs for reproduction number were created by creating 1,000 bootstrapped resampled datasets of deaths in each census tract, repeating the estimation procedure, and identifying the 2.5 and 97.5 percentiles of these estimates. The 95% CIs on the regression coefficients of sociodemographic covariates with reproduction numbers were created in a similar manner. The correlation between reproduction number and sociodemographic variables in each census tract was explored using Spearman rank correlation coefficients (7).

Analysis of Spatial Dependence. The ϕ clustering statistic [ϕ(d1, d2)] is defined as the probability of two deaths occurring in a defined spatial and temporal window relative to the independent probabilities of observing two deaths within the spatial window or within the temporal window (29, 30).