**Spatio-temporal Patterns and Diffusion of the 1918 Influenza Pandemic in British India**

Olivia Reyes, Elizabeth C. Lee, Pratha Sah, Cécile Viboud, Siddharth Chandra, and Shweta Bansal

Correspondence to Dr. Shweta Bansal, Department of Biology, Georgetown University, 37th and O Streets NW, Washington, DC 20057, (email: shweta.bansal@georgetown.edu)

Author affiliations: Department of Biology, Georgetown University, Washington, District of Columbia (Olivia Reyes, Elizabeth C. Lee, Pratha Sah, and Shweta Bansal); Fogarty International Center, National Institutes of Health, Bethesda, Maryland (Cécile Viboud and Shweta Bansal); and Asian Studies Center, James Madison College, and the Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, Michigan (Siddharth Chandra)

Funding: National Science Foundation grant # BCS-1244796; National Institutes of Health grant # 1-R21-DA025917; and the RAPIDD Program of the Science & Technology Directorate, Department of Homeland Security and the Fogarty International Center, National Institutes of Health.

Conflict of interest: None declared.

Running head: Spatio-temporal Diffusion of 1918 Pandemic in India

Abbreviations:
US = United States
UK = United Kingdom
INoDS = Inferring Network of infectious Disease Spread

**Abstract**

The factors that drive spatial heterogeneity and diffusion of pandemic influenza remain debated. Here, we characterize the spatio-temporal mortality patterns of the 1918 influenza pandemic in British India and study the role of demographic factors, environmental variables, and mobility processes on the observed patterns of spread. We analyze fever and all-cause excess mortality across 206 districts in India during the period of January 1916 to December 1920, and control for variation in seasonality particular to India. Our analysis reveals that the 1918 autumn wave in India matches signature features of influenza pandemics with high disease burden among young adults, (moderate) spatial heterogeneity in burden, and highly synchronized outbreaks across the country deviating from annual seasonality. Importantly, we also find that population density and rainfall explain the spatial variation in excess mortality, and that long-distance travel via railroad is predictive of the observed spatial diffusion of disease. Our work integrates a spatio-temporal analysis of mortality patterns during the 1918 influenza pandemic in India with data on underlying factors and processes to reveal transmission mechanisms in a large, intensely connected setting with significant climatic variability. The characterization of such heterogeneity during historical pandemics is crucial to our ability to prepare for future pandemics.

**Keywords:** Influenza, pandemic, spatial heterogeneity, diffusion, transportation, mobility, environment, tropics

The 1918 influenza pandemic has left an indelible mark on human history. Significant increases in respiratory and fever mortality were first observed in the United States in March 1918. By autumn, the H1N1 influenza had spread to the rest of the globe, facilitated in that newly globalized era by steamship travel and the intense movement of World War I troops [1, 2]. While pandemic mortality estimates remain disputed, recent analysis places the global toll at 50 million deaths [3].

The 1918 pandemic has been well-described in the US and Europe [4–7] and recent studies have characterized other parts of the Americas [8–11]. This work of the last two decades has established "signature" features of influenza pandemics: a shift in the virus subtype, an age shift in mortality to young adults, successive pandemic waves, high transmissibility and spatial heterogeneity in burden [12]. However, our understanding of historical pandemics in Asia remains limited and focuses primarily on the estimation of mortality impact [13–19], with few exceptions [20]. Characterizing the environmental,

socio-demographic, and evolutionary factors underlying epidemics is crucial to our ability to develop public health countermeasures and implement effective mitigation plans, and requires diverse case studies across climatic and socio-economic strata. Here, we contribute a case study of the 1918 pandemic in India, a nation which was made up of a largely rural but intensely connected population spread out over diverse climatic regions.

Of the 50 million deaths, British India was thought to have endured 8 million deaths at the time [21], a number which has recently been estimated to be closer to 14 million deaths [17, 22]. This means that 1 in every 23 Indians died during 1918-1919 and that 1 in every 3.5 global pandemic deaths was an Indian, both of which are underestimates as they only include the areas of India under British rule. It is understood that influenza first hit the province of Bombay in September 1918, and proceeded to spread north and east in a wave-like pattern that slowed and attenuated in severity as it traveled further from its origin [20, 22, 23].

The study of socio-demographic factors underlying influenza pandemics has received attention in past work, with the most focus on age-specific mortality risk. Numerous past studies have found that young adults experienced a disproportionately high mortality risk during the 1918 pandemic, while older adults had a relative decreased risk[4, 8, 9, 12, 22, 24–26]. Studies of other demographic features, however, have been few but include the work of Chowell et al. on the role of urbanization in predicting mortality burden, with both high population density and very low population density being associated with high death rates [5].

Past work on the impact of environmental factors on historical pandemics includes work on prediction of pandemic emergence based on El Nino cycles [27] and the association between disease and temperature [28] or latitude as a proxy for climate[11]. Additionally, recent studies on seasonal outbreaks highlight the importance of environmental processes by inferring a U-shaped effect of absolute humidity on seasonal influenza prevalence, mediated by temperature [29, 30]. That is, both low humidity/low temperature environments (as found in temperate regions of the world in winter months) and high humidity/high temperature environments (as found in tropical regions of the world) are hypothesized to increase influenza risk. Additionally, rainfall has been found to be associated with influenza epidemics in the tropics (e.g. [31–34]). Some hypotheses supporting this pattern include increased indoor crowding facilitating airborne and droplet transmission, and decreased vitamin D intake depressing innate immune responses [35]. These mechanisms may support a contemporaneous or asynchronous association between precipitation and influenza burden [29].

In addition to factors that drive individual-level transmission, population-level processes such as mobility are also crucial to the spatial dynamics of disease. The movement of human hosts provides the

scaffolding over which pathogens traverse great spatial distances, and has been implicated from the diffusion of plague in pre-industrial Europe along silk trade routes [36] to the spread of Ebola via regional connectivity [37] and Zika via air travel [38]. In rare explanatory studies about spatial diffusion during the 1918 pandemic, Palmer et al. examined the impact of boat and rail traffic on the spread of influenza in Newfoundland through qualitative methods [39] and Eggo et al. tested whether assorted mobility models predict pandemic dynamics in the UK and US [7]. Valleron et al even implicate surface travel in the spread of the 1889 influenza pandemic through Europe, though are unable to test this hypothesis due to unavailable data [40]. In India, the railway network began carrying unprecedented numbers of people further, faster, and more frequently leading up to the twentieth century (with annual passenger numbers growing from 0.5 million in 1854 to 176 million by 1900 [41]). Simultaneously, infectious disease outbreaks of cholera, plague, malaria and smallpox were ravaging the country, and Indian rail travel entered the global debate linking human travel to public health. During the year 1918, over 459 million passengers traveled the Indian railway, and the Sanitary Commissioner of India believed that this played an important role in pandemic influenza spread [20, 42].

Here, we characterize the spatial dynamics of excess mortality during the 1918 influenza pandemic in British India with respect to spatial age-structure, heterogeneity and synchrony. Using key covariate data, we also analyze the underlying environmental factors and social processes that may have contributed to the observed spatial variation and diffusion. In our study of the 1918 influenza pandemic in India, we aim to (a) tease apart the impact of population density, seasonality, rainfall and temperature on the observed east-west gradient; (b) differentiate the seasonal dynamics and environmental drivers of inter-pandemic and pandemic influenza; (c) understand the role of host mobility on the propagation of pandemic influenza. We suggest that the characterization of such heterogeneity during historical pandemics is crucial to our ability to prepare against future pandemics.

## Methods

### Defining Pandemic Mortality

Mortality data were obtained from sanitary reports published annually for 206 districts in the provinces or presidencies of Assam, Bengal, Bihar & Orissa, Bombay, Central Province & Berar, Madras, Northwest Frontier Province, Punjab, and the United Provinces [43, 44]. Monthly fever deaths were compiled for all years between 1916 and 1920 at the district-level and covered areas of India under British rule representing approximately 70% of the total population of 318 million in India [45]. Additionally, we compiled fever, respiratory and (age-specific and total) all-cause mortality at the province-level for 1916-1920 [43]. The primary source indicated that pandemic deaths were preferentially coded as fever rather

than respiratory causes. (In the Appendix, we compare these cause-specific and all-cause mortality data, Web Figure 1.) Consequently, we estimated monthly excess fever mortality (above a seasonal baseline) to identify the number of deaths attributed to influenza using a seasonal regression model, controlled for differences in regional seasonality. To provide a finer temporal resolution, we also re-sampled and interpolated the monthly excess fever mortality to produced weekly excess fever mortality time series. We used district-specific, weekly excess fever mortality for most of our analysis, with two exceptions: for our analysis on age-specific mortality patterns, we use total all-cause mortality as age-specific fever mortality data was not available; for our analysis on environmental drivers of disease burden, we use district-level, monthly excess fever mortality as environmental variables were only available at the monthly level. More details on our data and these procedures can be found in Web Appendix 1.

### Defining Covariates

Population size data were collected from the 1911 (decennial) Census of India [46] for each district in our dataset. Monthly rainfall and monthly minimum temperature data were compiled for 25 districts across all 9 provinces for 1916-1920 from the Sanitary Commissioner's Annual Report [43].

Travel data were collected on the number of passengers traveling annually on 59 local railway lines in India [42]. Based on these data, we constructed a railway travel network, where nodes were districts and an edge existed between two districts if there was one or more railway lines connecting them. Only railway line origins and final destinations were available. Travel was assumed to be bidirectional on each railway line, and each edge was weighted with the number of annual passengers traveling on the line, if available. After eliminating nodes with no disease data, the railway network consists of 52 nodes and 41 edges (with edge weight data available on 16 edges).

We also constructed a local travel network where nodes were districts and an edge existed between two districts if they shared a physical border. This network represents unobserved travel via roads or waterways. After eliminating nodes with missing disease data, the local travel network consists of 197 nodes and 382 edges.

### Measuring Spatial Heterogeneity & Synchrony

We examined spatial heterogeneity of excess fever mortality with the *Lorenz curve*, which compares the cumulative distribution of excess deaths to the cumulative population size among districts (ranked smallest to largest) [47]. The farther the Lorenz curve is from our expectation (the main diagonal), the greater the spatial heterogeneity in death rates. We further quantify this through the *Gini coefficient*, which is close to 1 when there is high spatial heterogeneity in death rates; and close to zero when death rates are directly proportional to population size.

To estimate the seasonality of pandemic and non-pandemic seasons, we detected the timing of epidemic peaks in each district by performing a continuous wavelet transformation on the time-series of excess fever mortality [48, 49]. Details on these methods can be found in Web Appendix 1.

### Examining Environmental Drivers of Disease Burden

Among the 25 districts for which rainfall and temperature data were collected, we used two time series generalized linear mixed models to examine the association between excess death rates and environmental factors for months before and during the pandemic period (January 1916 through July 1918 and August 1918 through March 1919, respectively). Excess death rates were transformed as log of the excess death rates plus one; environmental data were centered and standardized; district was included as a group (random) effect. For the two periods, we compared models where the rainfall predictor had 0 through 2 month lags to examine synchronous and asynchronous relationships between disease and rainfall.

### Explaining Spatial Diffusion of Disease

To test hypotheses about spatial diffusion, we considered associations between observed travel networks and the observed infection data using a likelihood-based approach called *INoDS*[50]. We apply this method on alternative travel networks assuming infection timing for each district coincides with pandemic onset. We defined a pandemic onset date for each district by using weekly excess death data and specifying onset as the first week when the excess death rate was greater than 1 per 1000 population. Using this pandemic onset date for each district and three network hypotheses (the local travel network and the railway travel network, unadjusted or weighted by passenger fluxes; each described above), we used a likelihood approach to estimate the predictive power of each empirical network to explain the observed patterns of pandemic spatial spread [50]. We infer transmission parameters for network spread and non-network spread, and measure predictive power by comparing each empirical network to a set of null networks. Further details on this methodology can be found in Web Appendix 1.

## Results

We use historical reports to estimate excess mortality for the autumn wave of the 1918 influenza pandemic in British India. We aim to (a) characterize the spatio-temporal patterns and age structure of excess mortality; (b) explain the spatial variation in excess mortality patterns with demographic and environmental factors; and (c) understand the role of short- and long-distance travel on spatial diffusion during the outbreak.

## Spatial Dynamics & Age Structure

We focus on analyzing the spatial and temporal dynamics of the autumn wave of the 1918 pandemic measured through fever deaths in 206 districts of India. The autumn wave of the pandemic in India started during the first week of September 1918 with shipping traffic into the Bombay port seeding infection [20, 23], and lasted through March 1919. While this wave was concentrated, our data shows significant heterogeneity in the temporal dynamics of the disease (Figure 1A). All districts had pandemic onset by November 1918, and cases lasted in each district from 2 to 13 weeks. The northern and central parts of the country (particularly parts of the Central Province & Berar and the Northwest Frontier) experienced the highest mortality burden, while the southern and eastern districts had less pronounced mortality waves (Figure 1B). The spatial diffusion of the pandemic resembled a wave-like pattern starting from the western coast and spreading eastward (Figure 3A), as has been demonstrated previously [20].

In Figure 2A, we compare all-cause mortality across age groups during the autumn wave of the pandemic compared to the influenza-relative excess mortality during 1917 (Figure 2A). Mortality rates are higher for all age groups compared to a recent seasonal outbreak. In particular, the pandemic impacted the young with death rates in individuals aged 20-30 years being 4-5 fold seasonal mortality rates in the western, central and northern provinces ; the pattern is weak but still detectable in the eastern provinces (including Madras, Bengal, Assam, and Bihar & Orissa) where burden was low. (See also Web Figure 2A in the Web Appendix for relative comparison). Additionally, we highlight that India largely did not experience the elderly sparing observed in other settings (Web Figure 2B).

## Spatial Heterogeneity & Synchrony

Our spatially-resolved dataset gives us an opportunity to further consider heterogeneity in the patterns of the 1918 influenza pandemic. In particular, we consider the spatial distribution and the temporal dynamics of the outbreak. In Figure 2B, we illustrate the Lorenz curve which highlights that larger populations are disproportionately responsible for disease burden. The Gini coefficient for the district-level data is moderate (0.27).

In Figure 2C, we consider the seasonality of influenza during the pandemic and during non-pandemic seasons using a wavelet analysis. We find that non-pandemic influenza-relative excess mortality is characterized by two different seasonality profiles (with peaks occurring during the summer or winter) for different geographic locations. On the other hand, the 1918 pandemic was highly synchronous across the country regardless of geography and non-pandemic seasonality.

### Environmental Drivers of Spatial Variation in Disease Burden

We examined associations between environmental drivers and peak influenza activity. First, we validate existing hypotheses about inter-pandemic seasons that suggest that high rainfall is associated with high excess mortality burden in tropical regions and low temperature (correlated with low humidity) is associated with excess mortality disease burden in temperate regions [51]. Next, we explore possible environmental associations with pandemic influenza [27, 52, 53]. Because the mechanisms behind these hypotheses lead to a synchronous or asynchronous association with rainfall, we examined six models that had one of 0 to 2 month lags for the rainfall predictor (but no lag for the minimum temperature predictor) for the pandemic and non-pandemic period.

We found that during the non-pandemic period, rainfall was positively predictive of spatial variation in excess mortality burden at all lags, while minimum temperature was largely insignificant or demonstrated a small (negative) effect size (Table 1). In contrast, influenza burden during the pandemic period is negatively predicted by rainfall at no lag or a lag of 1 month (Table 2). The 2 month lag model provided the best fit for the non-pandemic period, while the no lag model provided the best fit for the pandemic; however, all models had comparable AICs. We note that all models suffered from heteroscedascity, despite log transformation of the response data.

### Human Mobility & Spatial Diffusion of Disease

Our analysis of the spatio-temporal patterns of the 1918 pandemic in India suggests two hypotheses about the spatial diffusion of influenza during this outbreak: (a) The wave-like pattern observed in Figure 1A, Figure 3A, as well as [20], support spread via local (e.g. road, waterway) travel. (b) The spatial heterogeneity and spatial synchrony we observe in Figure 2C instead support spread via long-distance (e.g. railway) travel.

We tested these hypotheses through the *INoDS* method [50] by testing the ability of each travel network (local (Figure 3B), rail (Figure 3C), and weighted rail (Web Figure 3) networks) to predict the observed spatial progression of disease through the country (Figure 3A). We find that all three networks are significant in explaining the observed disease patterns when compared against the null (See significance in Table 3), conditional on the network transmission parameter, $\beta$, and non-network transmission parameter, $\epsilon$. In Web Appendix 2, we test the sensitivity of these results to our assumptions and find that they are robust (Web Table 1).

## Discussion

We have presented an analysis of the spatio-temporal spread of the autumn wave of the 1918 influenza pandemic between districts of British India. Our findings demonstrate that the spread of the 1918 H1N1

influenza virus was rapid and synchronous across the country, but resulted in varying disease burden across regions along an east-west gradient. We show that the spatial variation in infection burden is explained by environmental drivers, and that spatial diffusion of disease is predicted by long-distance mobility patterns.

The historical mortality data presented here are subject to limitations, notably in the coding for cause of death. While our use of fever deaths is validated by the primary source, fever deaths also include other seasonal infectious disease causes such as malaria. The seasonality of malaria has been identified as May-September during that era, so this would be a confounding factor for those provinces in our study that have been found to have monsoon influenza seasonality. Of these, a few are known to be hyper-endemic areas (e.g. parts of Central Province & Berar, Bihar & Orissa, Assam) [54]. However, given our focus on the autumn wave of the 1918 pandemic and because historical records report low malaria burden during 1918, we expect this to have limited impact on our findings [21].

Influenza seasonality remains poorly characterized, particularly in low-income countries and in the tropics [29, 55–57]. The distinct seasonality that we observed in India during non-pandemic excess mortality activity in 1916-1920 coincides with the climate zones of India based on the Köppen classification [58], with the northeastern region classified as "humid semitropical" and distinct from surrounding regions. Our seasonality findings are also largely consistent with recent studies of contemporary influenza seasonality in India [59–61] and other countries with mixed climates [56]. However, we note that our methods are unable to disentangle excess mortality caused by influenza from other pathogens with similar symptoms (e.g. malaria), and this may affect our understanding of non-pandemic seasonality.

We observe the signature "W" pattern of 1918 age-specific mortality among the provinces of India, with the highest mortality rates among infants, followed by older adults and adults. This pattern is similar to what has been found in other countries outside Europe and North America during the 1918 pandemic, including Colombia [10], Mexico [8], and Brazil [62]. (Comparison of age-specific mortality data from other countries is available in Web Figure 2 in the Appendix.) High-income countries have reported relatively low mortality rates among the elderly, but this was not observed in Indian populations, thus suggesting that they may not have been exposed to the 1830s global pandemic virus or its descendants [4, 26]. We also observe similarity in the age-specific mortality curves among provinces with the same seasonal influenza patterns (following the east-west gradient), where western districts with temperate-region seasonal influenza patterns tended to have greater mortality. We note that our findings may be limited due to our use of all-cause mortality data.

While we could not examine absolute and relative humidity [30], we considered the role of rainfall in predicting mortality burden during the pandemic compared with non-pandemic years. India experiences

complex seasonal influenza dynamics due to its size and climatic diversity. During non-pandemic periods, we found distinct seasonal patterns according to regional climatic profiles, and rainfall was positively associated with the magnitude of excess mortality, thus providing evidence for the increased crowding or decreased micronutrient hypothesis [35]. On the other hand, the 1918 pandemic in India had highly synchronous dynamics that supplanted distinct seasonality in different regions of the country. We thus hypothesize that socio-demographic and immunological factors may have dominated environmental ones to synchronize the timing of the autumn pandemic wave, as has been observed for the recent 2009 H1N1 pandemic (e.g. [63]). The magnitude of the autumn pandemic wave, however, was still modulated by environmental factors, with pandemic disease burden being inversely proportional to rainfall. We speculate that this surprising result can be explained by the link between environment and nutrition. The year 1918 brought one of the most severe droughts of the twentieth century to India, except in the northeastern region which received excess rain during the monsoon season (June-August) [64, 65]. These dry conditions, while beneficial for depressing other infectious diseases such as malaria and plague, led to food and milk shortages, thus increasing susceptibility to infection [21, 66, 67]. Our findings support this hypothesis in the models with an asynchronous association between rainfall and influenza burden. Our results also show a synchronous inverse relationship between precipitation and disease (in the lag-0 pandemic models) suggesting a correlation between rainfall and absolute humidity but would require additional data to test.

Beyond environmental factors, we also sought to identify demographic and social processes that could explain the observed spatial dynamics. First, by constructing a Lorenz curve, we identified spatial heterogeneity in the burden of the pandemic in British India and found that this burden was nonlinearly associated with population size. Our finding of a Gini coefficient of 0.27 is comparable to those found in rural areas of England & Wales for the 1918 pandemic [5]. Second, we focus on the impact of host movement dynamics on the spatial spread of disease. Past studies have identified two classes of spatial dynamics for influenza: (a) local and radially diffusive wave of spread as observed in [63]; and (b) hierarchical spread starting at populous centers (connected by long-distance travel) with subsequent spread to smaller areas [7, 68, 69]. Disentangling the hypotheses of wave-like versus hierarchical spread is key to our understanding of transmission mechanisms and to targeting control measures. Our spatio-temporal descriptive analysis and past work of the Indian pandemic suggest preliminary support for both hypotheses, thus we use a data-driven statistical approach to test them. Our findings provide significant evidence for long-distance travel (via the rail network) and for short-distance travel (via a local travel network), thus supporting the hierarchical spread hypothesis. Other modes of transportation (such as shipping traffic, WWI troop movements) may have also contributed to host mobility and

infection seeding (particularly in the port cities of Madras and Calcutta). However, our findings provide a parsimonious explanation for the observed spread without these alternative modes.

While the intense connectivity provided by rail travel may have been a key player in the propagation of the pandemic, the railways were also a focus of public health monitoring and biosecurity in India. Motivated by devastating plague outbreaks, the railways were part of an extensive and coordinated entry and exit screening medical surveillance system starting in 1897 [41]. Additionally, railway carriages were disinfected intensely. Modern outbreaks of SARS and the 2009 H1N1 influenza have brought into focus the limited impact of travel restrictions and travel surveillance given the fact that the reductions necessary to significantly affect spatial spread are not feasible in practice [70, 71]. Our findings about the role of travel in the pandemic spread (particularly the support of the railway network weighted by passengers) either confirm that travel surveillance was also not very effective in 1918, or suggest that the pandemic burden would have been worse in the absence of these public health efforts.

We limit our current study to the autumn wave of 1918 as it was the largest wave in India. Recent work has highlighted the importance of the "herald waves" that have been documented in North America and Europe ahead of the autumn wave [72]. Our data show limited evidence of high influenza mortality during April-May 1918, particularly in the districts of the United Provinces (UP) of India (Figure 1). While this epidemic may have been a herald wave, perhaps explaining UP's relatively mild autumn pandemic wave, UP's year-round excess mortality activity makes it difficult to distinguish from a seasonal influenza outbreak.

Our study contributes to our understanding of spatial variation and diffusion during the 1918 influenza pandemic. India of 1918 represents a unique case study with a highly rural population in a climatically diverse setting, intra-connected via railways and inter-connected with the rest of the globe through shipping traffic. The lack of elderly sparing and a largely missing herald wave place the 1918 pandemic in India with other rural and isolated populations; while the early, large, and fast autumn wave make it similar to the pandemic dynamics of large, connected locations. Our findings provide a parsimonious explanation of the spatial dynamics of the pandemic in India via environmental and social processes. In particular, our work highlights the role of rainfall in emerging infectious disease dynamics. As our society moves into an increasingly water-stressed future, we advocate that pandemic planning should better integrate an understanding of environmental extremes and how they feed into social, agricultural and economic processes affecting disease transmission.

## Acknowledgements

We are grateful for support from National Science Foundation grant # BCS-1244796 and National Institutes of Health (NIH) grant # 1-R21-DA025917 for the mortality data collection; and from the RAPIDD Program of the Science & Technology Directorate, Department of Homeland Security and the Fogarty International Center, NIH. We thank Angela Wong for assistance with covariate data digitization and Sarah Kramer for data collection and analysis early in the life of this work. We also acknowledge Dr. Patrick Manning and the Collaborative for Historical Information and Analysis at the University of Pittsburgh for collaboration on data collection.





**References**

1. Niall Johnson and Juergen Mueller. Updating the accounts: global mortality of the 1918-1920 "Spanish" influenza pandemic. *Bulletin of the History of Medicine*, 76(1):105–115, 2002.
2. John M Barry. *The great influenza: The story of the deadliest pandemic in history*. Penguin, New York, NY, 2005.
3. Jeffery K Taubenberger and David M Morens. 1918 influenza: the mother of all pandemics. *Emerging Infectious Diseases*, 12(1):15, 2006.
4. Cécile Viboud, Jana Eisenstein, Ann H Reid, Thomas A Janczewski, David M Morens, and Jeffery K Taubenberger. Age-and sex-specific mortality associated with the 1918–1919 influenza pandemic in Kentucky. *Journal of Infectious Diseases*, 207(5):721–729, 2013.
5. Gerardo Chowell, Luís MA Bettencourt, Niall Johnson, Wladimir J Alonso, and Cécile Viboud. The 1918–1919 influenza pandemic in England and Wales: spatial patterns in transmissibility and mortality impact. *Proceedings of the Royal Society of London B: Biological Sciences*, 275(1634):501–509, 2008.
6. Baltazar Nunes, Cécile Viboud, Ausenda Machado, Corinne Ringholz, Helena Rebelo-de Andrade, Paulo Nogueira, and Mark Miller. Excess mortality associated with influenza epidemics in Portugal, 1980 to 2004. *PLoS One*, 6(6):e20661, 2011.
7. Rosalind M Eggo, Simon Cauchemez, and Neil M Ferguson. Spatial dynamics of the 1918 influenza pandemic in England, Wales and the United States. *Journal of The Royal Society Interface*, 8(55):233–243, 2011.
8. Gerardo Chowell, Cécile Viboud, Lone Simonsen, Mark A Miller, and Rodolfo Acuna-Soto. Mortality patterns associated with the 1918 influenza pandemic in Mexico: evidence for a spring herald wave and lack of preexisting immunity in older populations. *Journal of Infectious Diseases*, 202(4):567–575, 2010.
9. Gerardo Chowell, Cécile Viboud, L Simonsen, MA Miller, J Hurtado, G Soto, R Vargas, MA Guzman, M Ulloa, and CV Munayco. The 1918–1920 influenza pandemic in Peru. *Vaccine*, 29:B21–B26, 2011.
10. Gerardo Chowell, Cécile Viboud, Lone Simonsen, Mark A Miller, Rodolfo Acuna-Soto, Juan M Ospina Díaz, and Abel Fernando Martínez-Martín. The 1918–19 influenza pandemic in Boyaca, Colombia. *Emerging Infectious Diseases*, 18(1):48, 2012.
11. Gerardo Chowell, Lone Simonsen, Jose Flores, Mark A Miller, and Cécile Viboud. Death patterns during the 1918 influenza pandemic in Chile. *Emerging Infectious Diseases*, 20(11):1803, 2014.
12. Mark A Miller, Cécile Viboud, Marta Balinska, and Lone Simonsen. The signature features of influenza pandemics—implications for policy. *New England Journal of Medicine*, 360(25):2595–2598, 2009.
13. Siddharth Chandra. Mortality from the influenza pandemic of 1918–19 in indonesia. *Population Studies*, 67(2):185–193, 2013.
14. Siddharth Chandra. Deaths associated with influenza pandemic of 1918–19, japan. *Emerging Infectious Diseases*, 19(4):616, 2013.
15. Siddharth Chandra and Dilshani Sarathchandra. The influenza pandemic of 1918–1919 in sri lanka: its demographic cost, timing, and propagation. *Influenza and other respiratory viruses*, 8(3):267–273, 2014.
16. Vernon J Lee, Mark I Chen, Siew Pang Chan, Chia Siong Wong, Jeffery Cutter, Kee Tai Goh, and Paul Anath Tambyah. Influenza pandemics in Singapore, a tropical, globally connected city. *Emerging Infectious Diseases*, 13(7):1052, 2007.
17. Siddharth Chandra, Goran Kuljanin, and Jennifer Wray. Mortality from the influenza pandemic of 1918–1919: the case of India. *Demography*, 49(3):857–865, 2012.
18. Ying-Hen Hsieh. Excess deaths and immunoprotection during 1918–1920 influenza pandemic, Taiwan. *Emerging Infectious Diseases*, 15(10):1617, 2009.
19. SA Richard, N Sugaya, L Simonsen, MA Miller, and C Viboud. A comparative study of the 1918–1920 influenza pandemic in Japan, USA and UK: mortality impact and implications for pandemic planning. *Epidemiology & Infection*, 137(8):1062–1072, 2009.
20. Siddharth Chandra and Eva Kassens-Noor. The evolution of pandemic influenza: evidence from India, 1918–19. *BMC Infectious Diseases*, 14(1):1, 2014.
21. E. Selby Phipson. The pandemic of influenza in India in the year 1918. *The Indian Medical Gazette*, 58(11):509, 1923.
22. Christopher JL Murray, Alan D Lopez, Brian Chin, Dennis Feehan, and Kenneth H Hill. Estimation of potential global pandemic influenza mortality on the basis of vital registry data from the 1918–20 pandemic: a quantitative analysis. *The Lancet*, 368(9554):2211–2218, 2006.
23. Kenneth Hill. Influenza in India 1918: excess mortality reassessed. *Genus*, 67(2):9–29, 2011.
24. Lone Simonsen, Matthew J Clarke, Lawrence B Schonberger, Nancy H Arden, Nancy J Cox, and Keiji Fukuda. Pandemic versus epidemic influenza mortality: a pattern of changing age distribution. *Journal of infectious diseases*, 178(1):53–60, 1998.
25. Shweta Bansal, Babak Pourbohloul, Nathaniel Hupert, Bryan Grenfell, and Lauren Ancel Meyers. The shifting demographic landscape of pandemic influenza. *PloS One*, 5(2):e9360, jan 2010.
26. Viggo Andreasen, Cécile Viboud, and Lone Simonsen. Epidemiologic characterization of the 1918 influenza pandemic summer wave in copenhagen: implications for pandemic control strategies. *Journal of Infectious Diseases*, 197(2):270–278, 2008.
27. Jeffrey Shaman and Marc Lipsitch. The El Niño–Southern Oscillation (ENSO)–pandemic Influenza connection: Coincident or causal? *Proceedings of the National Academy of Sciences*, 110(Supplement 1):3689–3691, 2013.

28. Daihai He, Jonathan Dushoff, Troy Day, Junling Ma, and David JD Earn. Inferring the causes of the three waves of the 1918 influenza pandemic in England and Wales. *Proceedings of the Royal Society of London B: Biological Sciences*, 280(1766):20131345, 2013.
29. James D Tamerius, Jeffrey Shaman, Wladmir J Alonso, Kimberly Bloom-Feshbach, Christopher K Uejio, Andrew Comrie, and Cécile Viboud. Environmental predictors of seasonal influenza epidemics across temperate and tropical climates. *PLoS Pathogens*, 9(3):e1003194, 2013.
30. Ethan R Deyle, M Cyrus Maher, Ryan D Hernandez, Sanjay Basu, and George Sugihara. Global environmental drivers of influenza. *Proceedings of the National Academy of Sciences*, 113(46):13081–13086, 2016.
31. Thierry Nyatanyi, Richard Nkunda, Joseph Rukelibuga, Rakhee Palekar, Marie Aimée Muhimpundu, Adeline Kabeja, Alice Kabanda, David Lowrance, Stefano Tempia, Jean Baptiste Koama, et al. Influenza sentinel surveillance in Rwanda, 2008–2010. *The Journal of Infectious Diseases*, 206(suppl_1):S74–S79, 2012.
32. Slinporn Prachayangprecha, Jarika Makkoch, Kamol Suwannakarn, Preeyaporn Vichaiwattana, Sumeth Korkong, Apiradee Theamboonlers, and Yong Poovorawan. Epidemiology of seasonal influenza in Bangkok between 2009 and 2012. *The Journal of Infection in Developing Countries*, 7(10):734–740, 2013.
33. Radina P Soebiyanto, Wilfrido Clara, Jorge Jara, Leticia Castillo, Oscar Rene Sorto, Sidia Marinero, María E Barnett de Antinori, John P McCracken, Marc-Alain Widdowson, Eduardo Azziz-Baumgartner, et al. The role of temperature and humidity on seasonal influenza in tropical areas: Guatemala, El Salvador and Panama, 2008–2013. *PLoS One*, 9(6):e100659, 2014.
34. Fernanda EA Moura, Anne CB Perdigão, and Marilda M Siqueira. Seasonality of influenza in the tropics: a distinct pattern in northeastern Brazil. *The American Journal of Tropical Medicine and Hygiene*, 81(1):180–183, 2009.
35. James Tamerius, Martha I Nelson, Steven Z Zhou, Cécile Viboud, Mark A Miller, and Wladimir J Alonso. Global influenza seasonality: reconciling patterns across temperate and tropical regions. *Environmental Health Perspectives*, 119(4):439, 2011.
36. Ricci PH Yue, Harry F Lee, and Connor YH Wu. Trade routes and plague transmission in pre-industrial Europe. *Scientific Reports*, 7(1):12973, 2017.
37. Andrew M Kramer, J Tomlin Pulliam, Laura W Alexander, Andrew W Park, Pejman Rohani, and John M Drake. Spatial spread of the West Africa Ebola epidemic. *Royal Society Open Science*, 3(8):160294, 2016.
38. Qian Zhang, Kaiyuan Sun, Matteo Chinazzi, Ana Pastore y Piontti, Natalie E Dean, Diana Patricia Rojas, Stefano Merler, Dina Mistry, Piero Poletti, Luca Rossi, et al. Spread of zika virus in the americas. *Proceedings of the National Academy of Sciences*, page 201620161, 2017.
39. Craig T Palmer, Lisa Sattenspiel, and Chris Cassidy. Boats, trains, and immunity: The spread of the spanish flu on the island of newfoundland. *Newfoundland and Labrador Studies*, 22(2), 2007.
40. Alain-Jacques Valleron, Anne Cori, Sophie Valtat, Sofia Meurisse, Fabrice Carrat, and Pierre-Yves Boëlle. Transmissibility and geographic spread of the 1889 influenza pandemic. *Proceedings of the National Academy of Sciences*, page 201000886, 2010.
41. Ritika Prasad. Contagion and control: Managing disease, epidemics, and mobility. In *Tracks of change: railways and everyday life in colonial India*, chapter 5. Taylor & Francis, Delhi, India, 2016.
42. Statistical Abstract Relating to British India from 1910-1911 to 1919-1920, 1922.
43. India Sanitary Commissioner. Annual Report of the Sanitary Commissioner with the Government of India, 1910-1919.
44. India Sanitary Commissioner. Annual Reports of the Sanitary Commissioners of Assam, Bihar and Orissa, Bengal, Bombay, Central Provinces, Madras, North West Frontier Provinces, Punjab and United Provinces of Agra and Oudh, 1916-1920.
45. India Census Commissioner. Census of India, 1921: Volumne 1, 1923. Online at http://www.censusindia.gov.in/Census_And_You/old_report/Census_1921.aspx. Accessed July 24, 2018.
46. India Census Commissioner. Census of India, 1911: Volumne 1, 1913. Online at http://www.censusindia.gov.in/Census_And_You/old_report/Census_1911.aspx. Accessed July 24, 2018.
47. Wen-Chung Lee. Characterizing exposure–disease association in human populations using the lorenz curve and gini index. *Statistics in Medicine*, 16(7):729–739, 1997.
48. Pan Du, Warren A Kibbe, and Simon M Lin. Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics*, 22(17):2059–2065, 2006.
49. Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. Online at http://www.scipy.org/. Accessed July 24, 2018.
50. Pratha Sah, Michael Otterstatter, Stephan T. Leu, Sivan Leviyang, and Shweta Bansal. Revealing mechanisms of infectious disease transmission through empirical contact networks. *bioRxiv*, 10.1101/169573, 2018.
51. Natalie Pica and Nicole M Bouvier. Environmental factors affecting the transmission of respiratory viruses. *Current Opinion in Virology*, 2(1):90–95, 2012.
52. James Tamerius, Cécile Viboud, Jeffrey Shaman, and Gerardo Chowell. Impact of School Cycles and Environmental Forcing on the Timing of Pandemic Influenza Activity in Mexican States, May-December 2009. *PLoS Computational Biology*, 11(8), 2015.
53. Spencer J. Fox, Joel C. Miller, and Lauren Ancel Meyers. Seasonality in risk of pandemic influenza emergence. *PLoS Computational Biology*, 13(10):1–23, 2017.
54. Patrick Hehir. *Malaria in India*. Oxford Medical Publications, London, United Kingdom, 1927.
55. Eduardo Azziz Baumgartner, Christine N Dao, Sharifa Nasreen, Mejbah Uddin Bhuiyan, Syeda Mah-E-Muneer, Abdullah Al Mamun, MA Yushuf Sharker, Rashid Uz Zaman, Po-Yung Cheng, Alexander I Klimov, et al. Seasonality, timing, and climate drivers of influenza activity worldwide. *The Journal of Infectious Diseases*, 206(6):838–846, 2012.
56. Cécile Viboud, Wladimir J Alonso, and Lone Simonsen. Influenza in tropical regions. *PLoS medicine*, 3(4):e89, 2006.
57. Eric Lofgren, Nina H Fefferman, Yuri N Naumov, Jack Gorski, and Elena N Naumova. Influenza seasonality: underlying causes and modeling theories. *Journal of Virology*, 81(11):5429–5436, 2007.
58. Murray C Peel, Brian L Finlayson, and Thomas A McMahon. Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences Discussions*, 4(2):439–473, 2007.
59. Parvaiz A Koul, Shobha Broor, Siddhartha Saha, John Barnes, Catherine Smith, Michael Shaw, Mandeep Chadha, and Renu B Lal. Differences in influenza seasonality by latitude, northern India. *Emerging Infectious Diseases*, 20(10):1723, 2014.
60. Mandeep S Chadha, Shobha Broor, Palani Gunasekaran, Varsha A Potdar, Anand Krishnan, Mamta Chawla-Sarkar, Dipankar Biswas, Asha M Abraham, Suresh V Jalgaonkar, Harpreet Kaur, et al. Multisite virological influenza surveillance in India: 2004–2008. *Influenza and other respiratory viruses*, 6(3):196–203, 2012.
61. Mandeep S Chadha, Varsha A Potdar, Siddhartha Saha, Parvaiz A Koul, Shobha Broor, Lalit Dar, Mamta Chawla-Sarkar, Dipankar Biswas, Palani Gunasekaran, Asha Mary Abraham, et al. Dynamics of influenza seasonality at sub-regional levels in India and implications for vaccination timing. *PLoS One*, 10(5):e0124122, 2015.
62. Wladimir J Alonso, Francielle C Nascimento, Rodolfo Acuna-Soto, Cynthia Schuck-Paim, and Mark A Miller. The 1918 influenza pandemic in Florianopolis: a subtropical city in Brazil. *Vaccine*, 29:B16–B20, 2011.

63. Julia R Gog, Sébastien Ballesteros, Cécile Viboud, Lone Simonsen, Ottar N Bjornstad, Jeffrey Shaman, Dennis L Chao, Farid Khan, and Bryan T Grenfell. Spatial transmission of 2009 pandemic influenza in the US. *PLoS Computational Biology*, 10(6):e1003635, 2014.
64. ON Dhar, AK Kulkarni, and GC Ghose. Rainfall distribution over Indian subdivisions during the wettest and the driest monsoons of the period 1901–1960. *Hydrological Sciences Journal*, 23(2):213–221, 1978.
65. B Parthasarathy, AA Munot, and DR Kothawale. All-India monthly and seasonal rainfall series: 1871–1993. *Theoretical and Applied Climatology*, 49(4):217–224, 1994.
66. Peter Katona and Judit Katona-Apte. The interaction between nutrition and infection. *Clinical Infectious Diseases*, 46(10):1582–1588, 2008.
67. Carla Stanke, Marko Kerac, Christel Prudhomme, Jolyon Medlock, and Virginia Murray. Health effects of drought: a systematic review of the evidence. *PLoS Currents*, 5, 2013.
68. Cécile Viboud, Ottar N Bjørnstad, David L Smith, Lone Simonsen, Mark A Miller, and Bryan T Grenfell. Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science*, 312(5772):447–451, 2006.
69. Duygu Balcan, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J Ramasco, and Alessandro Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, 106(51):21484–21489, 2009.
70. T Déirdre Hollingsworth, Neil M Ferguson, and Roy M Anderson. Will travel restrictions control the international spread of pandemic influenza? *Nature Medicine*, 12(5):497, 2006.
71. Sandro Meloni, Nicola Perra, Alex Arenas, Sergio Gómez, Yamir Moreno, and Alessandro Vespignani. Modeling human mobility responses to the large-scale spreading of infectious diseases. *Scientific Reports*, 1:62, 2011.
72. Lone Simonsen, Gerardo Chowell, Viggo Andreasen, Robert Gaffey, John Barry, Don Olson, and Cécile Viboud. A review of the 1918 herald pandemic wave: importance for contemporary pandemic response strategies. *Annals of Epidemiology*, 2018.
73. Baltazar Nunes, Isabel Natário, and M Lucília Carvalho. Time series methods for obtaining excess mortality attributable to influenza epidemics. *Statistical Methods in Medical Research*, 2010.

## Figures:

**Figure 1: Spatio-temporal Dynamics. (A):** Excess fever mortality per 1000 population from April 1918 to April 1919. District time series are illustrated with thin lines and are colored by province. Thicker lines show the province mean excess fever mortality. **(B):** Total excess fever deaths (per 1000 population) during the autumn wave of the 1918 pandemic in India. District borders are colored for locations where mortality data is available according to colors shown in legend.

**Figure 2: Spatial Age-Structure, Heterogeneity & Synchrony.(A):** Age-specific all-cause deaths during the 1918 pandemic autumn wave, by province. For comparison, we consider the age-specific all-cause deaths during the 1917 seasonal influenza epidemic for all of India. Young children: <5y, school children: 5-20y, young adults: 20-30y, adults: 30-50y, older adults: 50+y. **(B):** The Lorenz curve illustrates the distribution of the total number of excess deaths for the 1918-1919 pandemic wave, as a function of cumulative (ascending) population size at the district level. The blue circles show the empirical data which demonstrate moderate heterogeneity (with a Gini coefficient of 0.27) and the gray dashed line shows the null which represents no heterogeneity in death rates. **(C):** A wavelet analysis illustrates the synchrony of peak time of influenza activity during pandemic (1918-1919) and non-pandemic (1916-1917 and 1917-1918) periods. Each violin plot shows the occurrence of excess mortality peaks across districts within a given province between June of the first year and May of the subsequent year.

**Figure 3: Explaining Spatial Diffusion. (A):** A map showing the onset of the 1918 pandemic autumn wave in districts of India. The earliest onset is the week of September 1918 in the province of Bombay on the western coast. Pandemic influenza then spread eastward and northward. Using these onset times, we test whether different travel networks are explanatory of the observed spatial diffusion. **(B):** The local travel network between districts (nodes) of British India. Network edges represent shared district borders. **(C):** The railroad travel network between districts (nodes) of British India. Network edges represent one or more railway lines which originated in one district and terminated in another district. In (B) and (C), we have mortality data for all (blue) nodes shown, and nodes highlighted in red experienced a 1918 autumn wave of high excess mortality.

## Tables

**Table 1: Regression results for non-pandemic excess mortality activity.**

| Model & Predictor[a] | AIC[b] | Estimate[c] | SE[d] | $p$-value |
|---|---|---|---|---|
| Rainfall lag = 0 | 368 | | | |
| intercept | | 0.26 | 0.060 | 1.6E-5 |
| t (month) | | 0.0037 | 0.0012 | 0.0029 |
| rainfall[e] | | 1.07 | 0.50 | 0.034 |
| min. temperature | | $-0.019$ | 0.011 | 0.085 |
| Rainfall lag = 1 | 340 | | | |
| intercept | | 0.30 | 0.061 | 1.6E-6 |
| t (month) | | 0.0035 | 0.0013 | 0.0057 |
| rainfall[e] | | 2.78 | 0.52 | 1.04E-7 |
| min. temperature[e] | | $-0.022$ | 0.011 | 0.039 |
| Rainfall lag = 2 | 334 | | | |
| intercept | | 0.30 | 0.063 | 2.0E-6 |
| t (month) | | 0.0037 | 0.0013 | 0.0061 |
| rainfall[e] | | 3.44 | 0.54 | 2.9E-10 |
| min. temperature | | $-0.019$ | 0.011 | 0.079 |

[a] Models are shown with 0-2 month lags for the rainfall predictor.
[b] Akaike Information Criterion
[c] Estimates for the district group effects are excluded from this table.
[d] Standard error.
[e] Significant predictor.

**Table 2: Regression results for non-pandemic excess mortality activity**.

| Model (with rainfall lag)[a] | AIC[b] | Estimate[c] | SE[d] | $p$-value |
|---|---|---|---|---|
| Rainfall lag = 0 | 610 | | | |
| intercept | | 1.66 | 0.34 | 2.3E-6 |
| t (month) | | $-0.26$ | 0.027 | <2E-16 |
| rainfall[e] | | $-16.67$ | 3.35 | 1.5E-6 |
| min. temperature | | $-0.044$ | 0.054 | 0.42 |
| Rainfall lag = 1 | 632 | | | |
| intercept | | 1.93 | 0.35 | 1.4E-7 |
| t (month) | | $-0.23$ | 0.029 | 2.0E-13 |
| rainfall[e] | | $-6.57$ | 3.24 | 0.044 |
| min. temperature | | $-0.057$ | 0.057 | 0.32 |
| Rainfall lag = 2 | 636 | | | |
| intercept | | 2.12 | 0.36 | 1.1E-8 |
| t (month) | | $-0.19$ | 0.029 | 8.7E-10 |
| rainfall | | 2.90 | 3.15 | 0.36 |
| min. temperature | | $-0.072$ | 0.058 | 0.21 |

[a] Models are shown with 0-2 month lags for the rainfall predictor.
[b] Akaike Information Criterion
[c] Estimates for the district group effects are excluded from this table.
[d] Standard error.
[e] Significant predictor.

**Table 3: Explaining spatial diffusion through travel networks**.

| Travel network | $\beta$[a] | $\epsilon$[b] | Significance[c] |
|---|---|---|---|
| Local | 0.106 | 0.078 | 0.0 |
| Railroad (unweighted) | 0.08 | 0.04 | 0.0 |
| Railroad (weighted) | 0.45 | 0.03 | 0.002 |





[a] $\beta$ captures the transmission rate.

[b] $\epsilon$ captures the error rate.

[c] The significance represents the p-value comparing the empirical travel network to null networks