Science Advances
AAAS

Title

Community factors associated with local epidemic timing of respiratory syncytial virus: a spatiotemporal modeling study

Authors

Zhe Zheng,[1]* Virginia E. Pitzer,[1] Joshua L. Warren,[2] Daniel M. Weinberger[1]

Affiliations

[1]Department of Epidemiology of Microbial Diseases and the Public Health Modeling Unit, Yale School of Public Health, New Haven, CT 06520, USA.

[2]Department of Biostatistics and the Public Health Modeling Unit, Yale School of Public Health, New Haven, CT 06520, USA

* Corresponding Author:

Zhe Zheng,

Email: zhe.zheng@yale.edu

Abstract

Respiratory syncytial virus (RSV) causes a large burden of morbidity in young children and the elderly. Spatial variability in the timing of RSV epidemics provides an opportunity to probe the factors driving transmission of the virus, including factors that influence epidemic seeding and growth rates. Using hospitalization data from Connecticut, New Jersey, and New York, we estimated epidemic timing at the ZIP-code level using harmonic regression and then used a Bayesian meta-regression model to evaluate correlates of epidemic timing. Earlier epidemics were associated with larger household size and greater population density. Nearby localities had similar epidemic timing. Our results suggest RSV epidemics grow faster in areas with more local contact opportunities, and that epidemic spread follows a spatial diffusion process based on geographic proximity. Our findings can inform the timing of delivery of RSV extended half-life prophylaxis and maternal vaccines and guide future studies on the transmission dynamics of RSV.

MAIN TEXT

Introduction

Respiratory syncytial virus (RSV) causes over 2 million outpatient visits annually among children in the United States (US) (*1*). It is one of the leading causes of hospitalization for lower respiratory infection among young children and the elderly (*2*, *3*). Developing an effective vaccine against RSV is a high priority, as prophylaxis against RSV is

prohibitively expensive and complicated to administer (*4-7*). The timing of the delivery of maternal RSV vaccines and extended half-life prophylaxis relies heavily on the timing of seasonal RSV epidemics. Understanding the community factors associated with the seasonality of RSV can help to optimize implementation of extended-life prophylaxis and maternal vaccines in the near future (*8*).

The timing of RSV epidemics varies markedly over space, both over large geographic scales and within a region (*9, 10*). Across the US, epidemics start earliest in the southeast and later spread to the north and west (*11*). Environmental and climatic drivers affect the timing of the seasonal epidemic by providing a favorable environment for the virus and by affecting host immune defense and/or behavior (e.g., indoor gatherings in the cold or rainy season) (*12-14*). However, while environmental factors can explain some of the broad spatial differences in the timing of RSV epidemics, they cannot fully explain finer-scale spatial variations (*12, 15*). At smaller spatial scales, epidemics tend to start earlier and last longer in urban areas than in surrounding suburbs and rural areas (*7*).

A number of factors can influence the timing of an RSV epidemic in a community. Epidemic timing is broadly determined by the growth rate of the epidemic and, if the virus does not persist year-round, by when the virus is re-introduced into the community ("epidemic seeding") (Fig. S1). Epidemic growth rates affect the shape of the epidemic curve, while epidemic seeding influences when the epidemic starts. Growth rates can be influenced by a number of demographic factors, such as the population age distribution, susceptibility and contact rates (*16*) (Fig. S1). Contact rates are linked to factors like

household size and population density (*10, 17*). Susceptibility to infection is linked to socioeconomic status (SES) (*10, 17*). Less is known about the factors that drive epidemic seeding for RSV. Research on other respiratory viruses has suggested that virus importation from surrounding areas, importation from other regions (e.g., transmission through long-distance air travel or commuting patterns), and transmission within and between schools may play a role in epidemic seeding (*18-21*).

In this study, we developed a Bayesian statistical model to probe the association between the timing of seasonal epidemics of RSV and potential explanatory factors that influence epidemic growth rates and seeding, including area-based measures of human mobility, demographic variables, geographic proximity, and school districts. The tri-state area that includes New York, New Jersey, and Connecticut is ideal to study the fine-scale spatial variation in RSV epidemics because of the demographic diversity, high volume of population movement, and relatively similar climate across the region. This area includes New York City, the most populous city in the US, as well as rural areas like the North Country region of New York state and Litchfield County in Connecticut. In addition, variation in community factors exist at small spatial scales (*22*). The strong commuting ties between New York City and surrounding areas offer valuable opportunities to compare the impacts of commuting flows and geographic distance on RSV epidemic patterns. This study aims to gain a better understanding of the factors influencing local epidemic timing.

Results

Characteristics of the model and data

RSV-specific hospitalization data for children <2 years were obtained from State Inpatient Databases from the Connecticut Department of Public Health (CT-DPH) and the Healthcare Cost and Utilization Project maintained by the Agency for Healthcare Research and Quality (*23*). In total, the databases captured 67,244 RSV hospitalizations across 2,612 ZIP codes in New York, New Jersey, and Connecticut. The number of hospitalizations, commuters, and the length of study period varied among states (Table 1), as did the sociodemographic characteristics (Table S1, Fig. S2-S3).

Spatiotemporal pattern of RSV epidemics

RSV activity begins between late fall and early winter in the tri-state area (Fig. 1). The epidemics peaked between late winter and spring (Fig. 1). The estimated peak timing among ZIP codes ranged from late December to mid-March based on the best-fit model (Fig. 2). Visually, the local epidemics peaked earliest in urban areas (e.g., the New York metropolitan area) and then extended to less populous areas like upstate New York and eastern Connecticut. Epidemic peaks were generally earlier in New Jersey compared to the other states.

Major correlates of local variation in RSV epidemic timing

We fit statistical models that included several demographic covariates while also characterizing the structure of the unexplained variability in the data (see Methods). Earlier epidemic timing was associated with higher population density in all three states

(Table 2). There was an approximately 1.5- to 2-week difference in epidemic timing between the top and bottom deciles of population density. There was also an association between earlier epidemic timing and larger average household size in Connecticut and New York; the difference in epidemic timing between the bottom to top deciles of average household size was 0.7-1.3 weeks. In New Jersey, ZIP codes with higher income were associated with earlier epidemics. There was no association between the total number of people in the ZIP code and epidemic timing.

In all three states, there was substantial residual variability in the data after adjusting for the covariates. The covariates captured 41%, 30% and 51% of the variability in New Jersey, New York, and Connecticut, respectively. The best-fitting model assumed that the residual variation was correlated based on geographic proximity. Alternative models that assumed that the residual variation was related to commuter connectivity or that assumed no spatial structure fit less well (Table 3). In Connecticut, the geographic proximity model only fit marginally better than the commuting model. The residual variability in the geographic proximity model was further partitioned into variation based on geographic adjacency and variability based on school district boundaries. Geographic adjacency explained 89%, 94% and 69% of the total residual variability in New Jersey, New York, and Connecticut, respectively. School district effects contributed relatively little to the variations in epidemic timing (Table 4). The proportion of variability explained by population density, household size, school districts, and adjacency varied between states; however, their relative importance was consistent across states (Table 2 and Table 4).

Discussion

Understanding the timing of RSV epidemics is essential for certain interventions against RSV, including the administration of prophylactic antibodies and for assessing when to immunize pregnant women. We sought to better understand the underlying community factors associated with the timing of RSV epidemics. We find that the timing of RSV epidemics is highly spatially structured and demonstrate considerable variation, beginning in large urban areas and radially spreading to rural places over a 2.7-month period in the tri-state area. Epidemics generally peaked earlier in ZIP codes with higher population density and larger average household size, and epidemic timing in one location was correlated with timing in neighboring areas.

The covariates that were most strongly associated with earlier epidemic timing (population density and household size) could be related to the frequency of contact opportunities that lead to transmission (Fig. S1). Frequent contacts lead to rapid viral spread and epidemic growth, consistent with a positive correlation between population density and estimated RSV transmission rates for different US states (*12*). The strong connectivity between neighboring areas could reflect epidemic seeding (Fig. S1). Having stronger connections between proximate communities could provide more opportunities for introduction of the virus, eventually forming successful local chains of infection (*24*). This is supported by the localized and radially diffusive epidemic pattern. Other factors, such as gathering within schools and commuter flows, might also contribute to RSV transmission, but our analyses suggest these are not the major drivers of the observed spatial patterns in this population.

By mid-January, the tri-state area is usually well into its annual RSV season, but so far in late January 2021, the level of RSV activity remains low (*25*). The positivity rate of RSV detection in this area has declined since non-pharmaceutical interventions (NPIs) to mitigate the COVID-19 pandemic, such as travel bans, mask wearing and social distancing, were introduced in March 2020 (*25*). The measures taken to slow COVID-19 transmission are likely also effective in controlling RSV epidemics. Studies in Western Australia found that RSV in children dropped 98% through their winter of 2020, even though schools were open (*26*). This natural experiment suggests that gathering within schools may not be the major driver of RSV epidemics (or at least is not sufficient in light of other measures).

We found that higher median income is associated with earlier epidemics in New Jersey but not in the other states. The positive correlation between median income and the observed RSV epidemic timing in New Jersey could reflect a greater delay between infection and hospitalization in poorer neighborhoods (Fig. S1). It is also possible that the lack of an association in Connecticut and New York may result from the bidirectional effect of socioeconomic status on RSV epidemic timing. Low socioeconomic status may also increase susceptibility to RSV, thereby increasing the epidemic growth rate, although this effect may be partially mediated through differences in household size (Fig. S1).

The results from our study have important implications for planning clinical trials and developing transmission dynamic models for RSV. First, since household size may be associated with the risk of exposure and transmission of RSV, vaccine trials should

ensure that household size is balanced between vaccinees and non-vaccinees. Second, transmission dynamic models should account for spatial heterogeneity as well as spatial correlations in the force of infection (e.g., using a meta-population model). The spatial patterns of RSV epidemic timing from the best-fit model show considerable discrepancies between urban and rural areas, highlighting the need to consider these as distinct spatial units in RSV transmission models.

Our findings are consistent with previous genomic analyses of RSV, which found that household transmission is common, and viruses from nearby households share similar phylogenetic origins (*27, 28*). However, due to the small sample size and study design, genomic analyses were unable to compare the different transmission environments and their relative role in local RSV transmission (*27, 28*). Using empirical epidemiologic, demographic and commuting flow data, our research has found evidence to fill this knowledge gap. Previous epidemiological studies also suggested that household crowding and/or a larger number of siblings are associated with increased risk of severe RSV lower respiratory tract infection (*17*). These findings together provide a more complete picture of the major drivers of RSV transmission and how local risk factors affect regional patterns of RSV epidemic spread.

Other factors might also contribute to RSV spatial transmission and local seasonality (Fig. S1). There are limits to what can be mechanistically tested with the statistical model we used in this study. As the demographic variables tested in our model may represent other underlying mechanisms, it is difficult to pinpoint the exact drivers of RSV dynamics with great certainty. For example, our results will not be able to confirm who

drives transmission within households, as household size is correlated with the share of school-age children per household and the probability of having very young children. Mechanistic transmission dynamic models (e.g., meta-population models) that explicitly capture transmission and host immunity are needed to further study these issues.

Notably, the spatial spread pattern of RSV epidemics differs from that of influenza (*21, 24, 29*), which suggest different age groups drive transmission of RSV and influenza (*16*). As a result, the level of indirect protection that might be generated by vaccinating different age groups is expect to vary for the two pathogens (*30, 31*). Spatial studies suggested that commuting flows drive the spread of seasonal influenza epidemics, while school-age children may have been major drivers of the 2009 H1N1 pandemic in the US (*21, 24, 29*). However, our analysis suggests that short-range modes of transmission, especially transmission within and between local communities, predominates for RSV. This, together with the fact that children under five have the highest overall risk of RSV infection (*16*), suggests that it is possible that infants and pre-school aged children drive RSV transmission. This could potentially be explained by the build-up of partial immunity against RSV due to previous exposure as age increases (*32*).

Our study and interpretations have several caveats. First, we performed an ecological analysis with aggregate data on sociodemographic characteristics at the ZIP-code level. Thus, we did not assess the role of individual household size on risk of RSV transmission. Likewise, there could be substantial heterogeneity within ZIP codes in the demographic characteristics that we evaluated. Second, the aggregation of ZIP-code level cases to dominant school districts may have led to some misclassification and does not

account for the multiple primary schools typically found within a school district. This could have affected the data-fitting process and resulted in a lower percentage of variability that was explained by school districts. However, since school calendars and school bus routes are shared within the same school district, using school districts can help us evaluate the transmission within and between schools. Third, monthly time series may disguise some detailed variation compared to weekly time series. However, in a sensitivity analysis, we estimated the peak epidemic timing in Connecticut using both weekly data and data aggregated to the monthly level and found that the phase estimates were nearly identical. Thus, the availability of finely-resolved temporal surveillance data is not essential to capture broad patterns in epidemic timing. Fourth, while we draw inference about the potential mechanisms that drive RSV transmission, the pattern of spatial diffusion that we observed may result from a wide range of possibilities. Thus, the interpretation of our study results needs to be further explored with epidemiologic studies, genomic analyses, and transmission dynamic models. More detailed data, such as the proportion of young children attending daycare in each ZIP code, could help to address remaining gaps in the knowledge of this system.

In conclusion, our results reveal significant variation in local RSV epidemic timing. We find evidence that the timing of RSV epidemics is associated with average household size, population density, median income, and epidemic timing in neighboring areas. In general, RSV epidemics take off earlier in urban areas and spread to rural places with low population density and average household size. These findings highlight the need for infection control within households and communities to protect high-risk populations. Our results also offer additional insights that can be used to inform the development of

transmission dynamic models and provide guidance on future vaccine target populations and clinical trial design.

Materials and Methods

Data source

RSV-specific hospitalization data for children <2 years from 1997 to 2013 were obtained from the Connecticut State Inpatient Database through the Connecticut Department of Public Health (CT-DPH). This dataset included the week and month of admission and the ZIP code of residence. Similar data for New York state and New Jersey were obtained from State Inpatient Databases of the Healthcare Cost and Utilization Project maintained by the Agency for Healthcare Research and Quality; these comprehensive databases contain all hospital discharge records from community hospitals in participating states (*23*). Datasets include the month of hospitalization (for 2005-2014) and ZIP code of residence. Cases were defined as any child <2 years old whose hospital discharge diagnostic codes included 079.6 (RSV), 466.11 (bronchiolitis due to RSV), or 480.1 (pneumonia due to RSV), based on the International Classification of Disease Ninth Revision [ICD-9]. The analysis of the data was approved by the Human Investigation Committees at Yale University and the CT-DPH. The authors assume full responsibility for analyses and interpretation of data obtained from the CT-DPH.

Geospatial data at the ZIP-code level were obtained from the US Census Bureau's Geography program (*33*). Information about population size <5 years old, average family type of household size and median income in each ZIP code area were obtained from the

US Census Bureau's American Community Survey (*34-36*). School district information was obtained from state geographic information systems, education department databases, and national center for education statistics (*37-40*). Countrywide commuting patterns at the ZIP-code level were obtained from the US Census Bureau's Center of Economic Studies (*41*). All geographic and demographic data except median income are from the 2010 census, since annual inter-censual estimates were not available for all variables during this time period; this corresponds to the midpoint of the hospitalization data from New York and New Jersey. We also performed sensitivity analyses using data from 2000 in Connecticut and 2014 in New Jersey to evaluate whether changes in demographic variables would influence the estimates. Ideally, we would perform sensitivity analysis using covariate data from each year in each state; however, due to the computational cost, we chose to perform sensitivity analysis using the data from two ends of the time-series spectrum.

Model structures

Two-stage hierarchical Bayesian model

We estimated the factors associated with spatial variability in the timing of RSV epidemics at the ZIP-code level using a two-stage modeling approach. In the first stage, we obtained an estimate of epidemic timing (phase shift) for each ZIP code, and a corresponding measure of uncertainty, using harmonic regression. In the second stage, we fit hierarchical Bayesian spatial models with multiple covariates and different correlation structures. Model comparison techniques were used to evaluate the different models.

First stage model

The first stage consists of a harmonic Poisson regression model that estimates the amplitude and phase of seasonal RSV epidemics (18). This model was fit separately to the monthly time series of observed RSV hospitalizations from each ZIP code and was specified as:

$$Y_{it} \sim \text{Poisson}(\lambda_{it})$$

$$\ln(\lambda_{it}) = \ln(pop5_i) + \alpha_i + \xi_i \cos\left(\frac{2\pi t}{12} - T_i\right) + \varphi_{it}.$$

The observed number of RSV cases in ZIP code *i* during month *t* is denoted as Y_{it}, phase T_i captures the shift in peak timing (with the 12-month period starting in early July), ξ_i is the amplitude of seasonal epidemics, $\ln(pop5_i)$ is the offset term for the population under 5, α_i is the intercept, and φ_{it} is an observation-level random effect that accounts for any unexplained variability in the data (i.e., overdispersion). In this first stage, all model parameters were estimated separately for each ZIP code. This model was fitted in the Bayesian setting using the rjags package with weakly informative prior distributions specified for all model parameters (*42*). Full details are provided in the Supplement.

Second stage model

Based on estimates of T_i (posterior means) and their uncertainty (posterior standard deviations) obtained from the first stage, we used a Bayesian meta-regression approach in the second stage to explain spatial variability in the epidemic peak timing of RSV (*43-47*). Specifically, we first transformed the peak timing parameters, whose values are confined to [0, 2π], to have support on the real line (useful for the second stage regression modeling) such that $\theta_i = \ln\left(\frac{T_i}{2\pi - T_i}\right) \in \mathbb{R}$. Next, we obtained posterior means, $\hat{\theta}_i$, and

posterior standard deviations, $\hat{\sigma}_i$, from the first stage and specify the second stage model as $\hat{\theta}_i \sim \mathrm{N}(\theta_i, \hat{\sigma}_i^2)$ where $\theta_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \eta_{d(i)} + \phi_i$. We used stepwise forward selection to choose candidate variables from the causal diagram for use in our second stage models (Fig. S1). We performed variable selection using a simplified version of the regression model that excluded the spatial random effects. We began with the two variables (average household size and population density) that decreased the residual variance the most. We then included the variables representing other mechanisms, provided they did not affect the posterior estimates of the previously selected variables to avoid issues related to multicollinearity.

We evaluated several alternative correlation structures to describe the residual variation in the data after adjusting for the ZIP-code level covariates (average household size, population density, population size, and median income). All covariates were standardized before entering the models. The structure of the second stage model was specified as:

$$\theta_i = \beta_0 + \beta_1 * (\text{household size}_i) + \beta_2 * (\log \text{pop den}_i) + \beta_3 * (\log \text{pop size}_i) + \beta_4 * (\log \text{median income}_i) + \eta_{d(i)} + \phi_i$$

where $\eta_{d(i)}$ and ϕ_i are random effects for school district *d* and ZIP code *i*. The school district random effects are distributed as $\eta_{d(i)} \sim \mathrm{N}(0, \sigma_\eta^2)$, where *d(i)* is the school district that includes ZIP code *i*. The alternative structures for ϕ_i were:

1) *No residual spatial correlation in RSV epidemic timing*: $\phi_i \sim \mathrm{N}(0, \sigma_\phi^2)$.

2) *RSV epidemics have similar timing in neighboring geographic locations*: $\boldsymbol{\phi} \sim \mathrm{CAR}(\rho, \sigma_{\phi}^{2})$ (*43*). Here, spatial proximity was defined by ZIP codes with adjacent borders.

3) *RSV epidemics have similar timing in locations connected strongly by commuting patterns*: $\boldsymbol{\phi} \sim \mathrm{CAR}(\rho, \sigma_{\phi}^{2})$(*43*). Here, spatial proximity was defined by commuting patterns across ZIP codes.

We estimated the proportion of residual variability explained by the spatially correlated random effects (ϕ_i) versus the school district random effects ($\eta_{d(i)}$) (see Supplement). We also calculated the difference in epidemic timing between the top and bottom deciles for each variable to measure the relative importance of input variables (see Supplement). More details about hyperprior distributions for these different models are given in the Supplement.

Parameter estimation was carried out separately for each of the three states via Markov chain Monte Carlo simulation with an initial burn-in period of 10,000 iterations and a subsequent set of 50,000 posterior samples collected for all parameters in the first stage (*48*). A burn-in of 5,000 iterations and a subsequent set of 20,000 posterior samples were collected in the second stage. Convergence was assessed by Gelman-Robin diagnostics and examining individual parameter trace plots, with no obvious signs of non-convergence observed. Deviance information criterion (DIC) was calculated for each model to compare the performance; a lower DIC indicates a model has an improved

balance of fit and complexity (*49*). Equations for the calculation of residual variance can be found in the Supplement. All analyses were performed using R v3.5.3.

References

1. C. B. Hall, G. A. Weinberg, M. K. Iwane, A. K. Blumkin, K. M. Edwards, M. A. Staat, P. Auinger, M. R. Griffin, K. A. Poehling, D. Erdman, C. G. Grijalva, Y. Zhu, P. Szilagyi, The Burden of Respiratory Syncytial Virus Infection in Young Children. *New England Journal of Medicine* **360**, 588-598 (2009).
2. T. Shi, D. A. McAllister, K. L. O'Brien, E. A. F. Simoes, S. A. Madhi, B. D. Gessner, F. P. Polack, E. Balsells, S. Acacio, C. Aguayo, I. Alassani, A. Ali, M. Antonio, S. Awasthi, J. O. Awori, E. Azziz-Baumgartner, H. C. Baggett, V. L. Baillie, A. Balmaseda, A. Barahona, S. Basnet, Q. Bassat, W. Basualdo, G. Bigogo, L. Bont, R. F. Breiman, W. A. Brooks, S. Broor, N. Bruce, D. Bruden, P. Buchy, S. Campbell, P. Carosone-Link, M. Chadha, J. Chipeta, M. Chou, W. Clara, C. Cohen, E. de Cuellar, D. A. Dang, B. Dash-Yandag, M. Deloria-Knoll, M. Dherani, T. Eap, B. E. Ebruke, M. Echavarria, C. C. de Freitas Lazaro Emediato, R. A. Fasce, D. R. Feikin, L. Feng, A. Gentile, A. Gordon, D. Goswami, S. Goyet, M. Groome, N. Halasa, S. Hirve, N. Homaira, S. R. C. Howie, J. Jara, I. Jroundi, C. B. Kartasasmita, N. Khuri-Bulos, K. L. Kotloff, A. Krishnan, R. Libster, O. Lopez, M. G. Lucero, F. Lucion, S. P. Lupisan, D. N. Marcone, J. P. McCracken, M. Mejia, J. C. Moisi, J. M. Montgomery, D. P. Moore, C. Moraleda, J. Moyes, P. Munywoki, K. Mutyara, M. P. Nicol, D. J. Nokes, P. Nymadawa, M. T. da Costa Oliveira, H. Oshitani, N. Pandey, G. Paranhos-Baccala, L. N. Phillips, V. S. Picot, M. Rahman, M. Rakoto-Andrianarivelo, Z. A. Rasmussen, B. A. Rath, A. Robinson, C. Romero, G. Russomando, V. Salimi, P. Sawatwong, N. Scheltema, B. Schweiger, J. A. G. Scott, P. Seidenberg, K. Shen, R. Singleton, V. Sotomayor, T. A. Strand, A. Sutanto, M. Sylla, M. D. Tapia, S. Thamthitiwat, E. D. Thomas, R. Tokarz, C. Turner, M. Venter, S. Waicharoen, J. Wang, W. Watthanaworawit, L. M. Yoshida, H. Yu, H. J. Zar, H. Campbell, H. Nair, R. S. V. G. E. Network, Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: a systematic review and modelling study. *Lancet* **390**, 946-958 (2017).
3. T. Shi, A. Denouel, A. K. Tietjen, I. Campbell, E. Moran, X. Li, H. Campbell, C. Demont, B. O. Nyawanda, H. Y. Chu, S. K. Stoszek, A. Krishnan, P. Openshaw, A. R. Falsey, H. Nair, R. Investigators, Global Disease Burden Estimates of Respiratory Syncytial Virus-Associated Acute Respiratory Infection in Older Adults in 2015: A Systematic Review and Meta-Analysis. *The Journal of infectious diseases*, (2019).
4. D. Isaacs, Should respiratory care in preterm infants include prophylaxis against respiratory syncytial virus? The case against. *Paediatric respiratory reviews* **14**, 128-129 (2013).
5. H. Nair, V. R. Verma, E. Theodoratou, L. Zgaga, T. Huda, E. A. Simoes, P. F. Wright, I. Rudan, H. Campbell, An evaluation of the emerging interventions against Respiratory Syncytial Virus (RSV)-associated acute lower respiratory infections in children. *BMC Public Health* **11 Suppl 3**, S30 (2011).

6. "RSV Vaccine Research and Development Technology Roadmap. Priority activities for development, testing, licensure and global use of RSV vaccines, with a specific focus on the medical need for young children in low- and middle-income countries," (Geneva: World Health Organization, 2017;https://apps.who.int/iris/handle/10665/258706). [the easiest access to this source is via the URL]
7. D. M. Weinberger, J. L. Warren, C. A. Steiner, V. Charu, C. Viboud, V. E. Pitzer, Reduced-Dose Schedule of Prophylaxis Based on Local Data Provides Near-Optimal Protection Against Respiratory Syncytial Virus. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **61**, 506-514 (2015).
8. S. B. Drysdale, C. J. Sande, C. A. Green, A. J. Pollard, RSV vaccine use--the missing data. *Expert Rev Vaccines* **15**, 149-152 (2016).
9. K. Bloom-Feshbach, W. J. Alonso, V. Charu, J. Tamerius, L. Simonsen, M. A. Miller, C. Viboud, Latitudinal variations in seasonal activity of influenza and respiratory syncytial virus (RSV): a global comparative review. *PloS one* **8**, e54445 (2013).
10. D. B. Noveroske, J. L. Warren, V. E. Pitzer, D. M. Weinberger, Local variations in the timing of RSV epidemics. *BMC infectious diseases* **16**, 674 (2016).
11. E. B. Rose, A. Wheatley, G. Langley, S. Gerber, A. Haynes, Respiratory Syncytial Virus Seasonality - United States, 2014-2017. *MMWR Morb Mortal Wkly Rep* **67**, 71-76 (2018).
12. V. E. Pitzer, C. Viboud, W. J. Alonso, T. Wilcox, C. J. Metcalf, C. A. Steiner, A. K. Haynes, B. T. Grenfell, Environmental Drivers of the Spatiotemporal Dynamics of Respiratory Syncytial Virus in the United States. *PLOS Pathogens* **11**, e1004591 (2015).
13. M. Moriyama, W. J. Hugentobler, A. Iwasaki, Seasonality of Respiratory Viral Infections. *Annu Rev Virol* **7**, 83-101 (2020).
14. N. Pica, N. M. Bouvier, Environmental factors affecting the transmission of respiratory viruses. *Curr Opin Virol* **2**, 90-95 (2012).
15. G. Almogy, L. Stone, B. A. Bernevig, D. G. Wolf, M. Dorozko, A. E. Moses, R. Nir-Paz, Analysis of Influenza and RSV dynamics in the community using a 'Local Transmission Zone' approach. *Sci Rep* **7**, 42012 (2017).
16. E. Goldstein, H. H. Nguyen, P. Liu, C. Viboud, C. A. Steiner, C. J. Worby, M. Lipsitch, On the Relative Role of Different Age Groups During Epidemics Associated With Respiratory Syncytial Virus. *The Journal of infectious diseases* **217**, 238-244 (2018).
17. E. A. Simoes, Environmental and demographic risk factors for respiratory syncytial virus lower respiratory tract disease. *The Journal of pediatrics* **143**, S118-126 (2003).
18. M. G. Hancean, M. Perc, J. Lerner, Early spread of COVID-19 in Romania: imported cases from Italy and human-to-human transmission networks. *R Soc Open Sci* **7**, 200780 (2020).
19. N. Hens, G. M. Ayele, N. Goeyvaerts, M. Aerts, J. Mossong, J. W. Edmunds, P. Beutels, Estimating the impact of school closure on social mixing behaviour and the transmission of close contact infections in eight European countries. *BMC infectious diseases* **9**, 187 (2009).
20. S. Cauchemez, A. J. Valleron, P. Y. Boelle, A. Flahault, N. M. Ferguson, Estimating the impact of school closure on influenza transmission from Sentinel data. *Nature* **452**, 750-754 (2008).

21. V. Charu, S. Zeger, J. Gog, O. N. Bjornstad, S. Kissler, L. Simonsen, B. T. Grenfell, C. Viboud, Human mobility and the spatial transmission of influenza in the United States. *PLoS Comput Biol* **13**, e1005382 (2017).
22. N. Y. C. D. o. H. a. M. Hygiene, "Health Disparities in New York City, Health Disparities in Life Expectancy and Death," (2010).
23. "Healthcare Cost and Utilization Project State Inpatient Databases," (Agency for Healthcare Research and Quality, 2014; http://www.hcup-us.ahrq.gov/sidoverview.jsp).
24. J. R. Gog, S. Ballesteros, C. Viboud, L. Simonsen, O. N. Bjornstad, J. Shaman, D. L. Chao, F. Khan, B. T. Grenfell, Spatial Transmission of 2009 Pandemic Influenza in the US. *PLoS Comput Biol* **10**, e1003635 (2014).
25. "RSV Census Regional Trends," (Centers for Disease Control and Prevention, 2021;https://www.cdc.gov/surveillance/nrevss/rsv/region.html#north).
26. D. K. Yeoh, D. A. Foley, C. A. Minney-Smith, A. C. Martin, A. O. Mace, C. T. Sikazwe, H. Le, A. Levy, C. C. Blyth, H. C. Moore, The impact of COVID-19 public health measures on detections of influenza and respiratory syncytial virus in children during the 2020 Australian winter. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, (2020).
27. C. N. Agoti, P. K. Munywoki, M. V. T. Phan, J. R. Otieno, E. Kamau, A. Bett, I. Kombe, G. Githinji, G. F. Medley, P. A. Cane, P. Kellam, M. Cotten, D. J. Nokes, Transmission patterns and evolution of respiratory syncytial virus in a community outbreak identified by genomic analysis. *Virus Evol* **3**, vex006 (2017).
28. C. N. Agoti, M. V. T. Phan, P. K. Munywoki, G. Githinji, G. F. Medley, P. A. Cane, P. Kellam, M. Cotten, D. J. Nokes, Genomic analysis of respiratory syncytial virus infections in households and utility in inferring who infects the infant. *Sci Rep* **9**, 10076 (2019).
29. C. Viboud, O. N. Bjornstad, D. L. Smith, L. Simonsen, M. A. Miller, B. T. Grenfell, Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* **312**, 447-451 (2006).
30. J. Medlock, A. P. Galvani, Optimizing influenza vaccine distribution. *Science* **325**, 1705-1708 (2009).
31. D. Yamin, F. K. Jones, J. P. DeVincenzo, S. Gertler, O. Kobiler, J. P. Townsend, A. P. Galvani, Vaccination strategies against respiratory syncytial virus. *Proc Natl Acad Sci U S A* **113**, 13239-13244 (2016).
32. F. W. Henderson, A. M. Collier, W. A. Clyde, Jr., F. W. Denny, Respiratory-syncytial-virus infections, reinfections and immunity. A prospective, longitudinal study in young children. *The New England journal of medicine* **300**, 530-534 (1979).
33. "TIGER/Line Shapefiles," (United States Census Bureau, 2010;https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.2010.html).
34. "American Community Survey 1-Year Data (2005-2019)," (United States Census Bureau, 2010;https://www.census.gov/data/developers/data-sets/acs-1year.2010.html).
35. "American Community Survey 5-Year Data (2009-2019)," (United States Census Bureau, 2011;https://www.census.gov/data/developers/data-sets/acs-5year.2011.html).
36. "Decennial Census (2010, 2000)," (United States Census Bureau, 2000;https://www.census.gov/data/developers/data-sets/decennial-census.html).

37. "Connecticut GIS Data," (Map and Geographic Information Center, UCONN, 2010;http://magic.lib.uconn.edu/connecticut_data.html#education).
38. "School District Boundaries," *Eduction Demographic and Geographic Estimates* (National Center for Education Statistics, 2010;https://nces.ed.gov/programs/edge/Geographic/DistrictBoundaries).
39. "NYS Schools and School District Boundaries," (NYS Education Department, GIS.NY.GOV;https://gis.ny.gov/gisdata/inventories/details.cfm?DSID=1326).
40. "School Districts - Unified for New Jersey," (NJGIN Open Data;https://njogis-newjersey.opendata.arcgis.com/datasets/school-districts-unified-for-new-jersey?geometry=-85.178%2C38.667%2C-64.271%2C41.606).
41. "Longitudinal Employer-Household Dynamics," (Center for Economic Studies, United States Census Bureau;https://lehd.ces.census.gov/data/).
42. M. Plummer, A. Stukalov, M. m.-j. s. n. Denwood, "rjags: bayesian graphical models using MCMC (Version 4-10)," (2019;http://mcmc-jags.sourceforge.net/).
43. B. G. Leroux, X. Lei, N. Breslow, in *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, M. E. Halloran, D. Berry, Eds. (Springer New York, New York, NY, 2000), pp. 179-191.
44. D. Lee, A comparison of conditional autoregressive models used in Bayesian disease mapping. *Spat Spatiotemporal Epidemiol* **2**, 79-89 (2011).
45. Y. C. MacNab, Hierarchical Bayesian modeling of spatially correlated health service outcome and utilization rates. *Biometrics* **59**, 305-316 (2003).
46. M. L. Bell, A. McDermott, S. L. Zeger, J. M. Samet, F. Dominici, Ozone and short-term mortality in 95 US urban communities, 1987-2000. *JAMA* **292**, 2372-2378 (2004).
47. J. Besag, J. York, A. Mollié, Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* **43**, 1-20 (1991).
48. S. Banerjee, B. Carlin, A. Gelfand, *Hierarchical Modeling and Analysis of Spatial Data*. (2004), vol. 101.
49. D. J. Spiegelhalter, N. G. Best, B. P. Carlin, A. Van Der Linde, Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 583-639 (2002).
50. C. Prener, "Areal Weighted Interpolation," (05/12/2020;https://slu-opengis.github.io/areal/articles/areal-weighted-interpolation.html).
51. C. G. Prener, C. K. Revord, areal: An R package for areal weighted interpolation. *Journal of Open Source Software* **4**, 1221 (2019).

Acknowledgments

Funding: DMW, VEP, JLW acknowledge support from grant R01AI137093 from the National Institute of Allergy and Infectious Diseases/National Institutes of Health. ZZ acknowledges support from the China Scholarship Council (CSC) scholarship #201806380027. The content is solely the responsibility of the authors.

Author contributions: ZZ implemented the study, analyzed the data and drafted the article. DMW conceptualized and designed the study. VEP helped revise the study design and manuscript for important intellectual content. JLW designed the study's analytic strategy and helped prepare the Methods sections of the text.

Competing interests: VEP has received reimbursement from Merck and Pfizer for travel expenses to Scientific Input Engagements on respiratory syncytial virus. DMW has received consulting fees from Pfizer, Merck, GSK, and Affinivax for topics unrelated to this manuscript and is Principal Investigator on a research grant from Pfizer on an unrelated topic. All other authors report no relevant conflicts.

Data and materials availability: The demographic and geographic data that support the findings of this study are publicly available from the Geography program of the US Census Bureau, the American Community Survey of the US Census Bureau, the Center of Economic Studies of the US Census Bureau, National Center for Education Statistics, state geographic information system, and education department databases. The hospitalization data are not available publicly, but can be obtained from the State Inpatient Database with the permission of the Connecticut Department of Public Health (CT-DPH) and upon signing a data use agreement with the Agency for Healthcare Research and Quality. The R code for this study can be found in the GitHub repository: https://github.com/weinbergerlab/RSV_spatiotemporal.

Figures and Tables

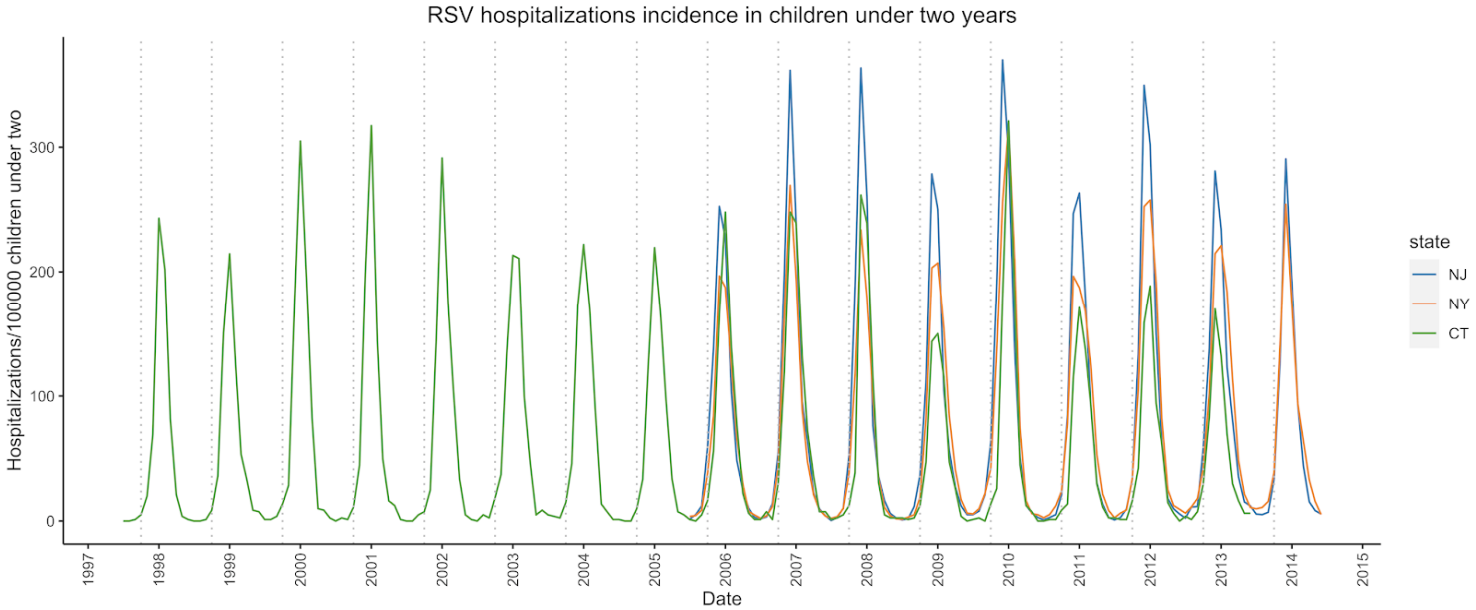


Fig. 1. RSV hospitalization incidence in children under two. The solid color lines show the time series of RSV hospitalizations in children <2 years in Connecticut (CT), New Jersey (NJ), and New York (NY). The vertical dotted line indicates October of each year.

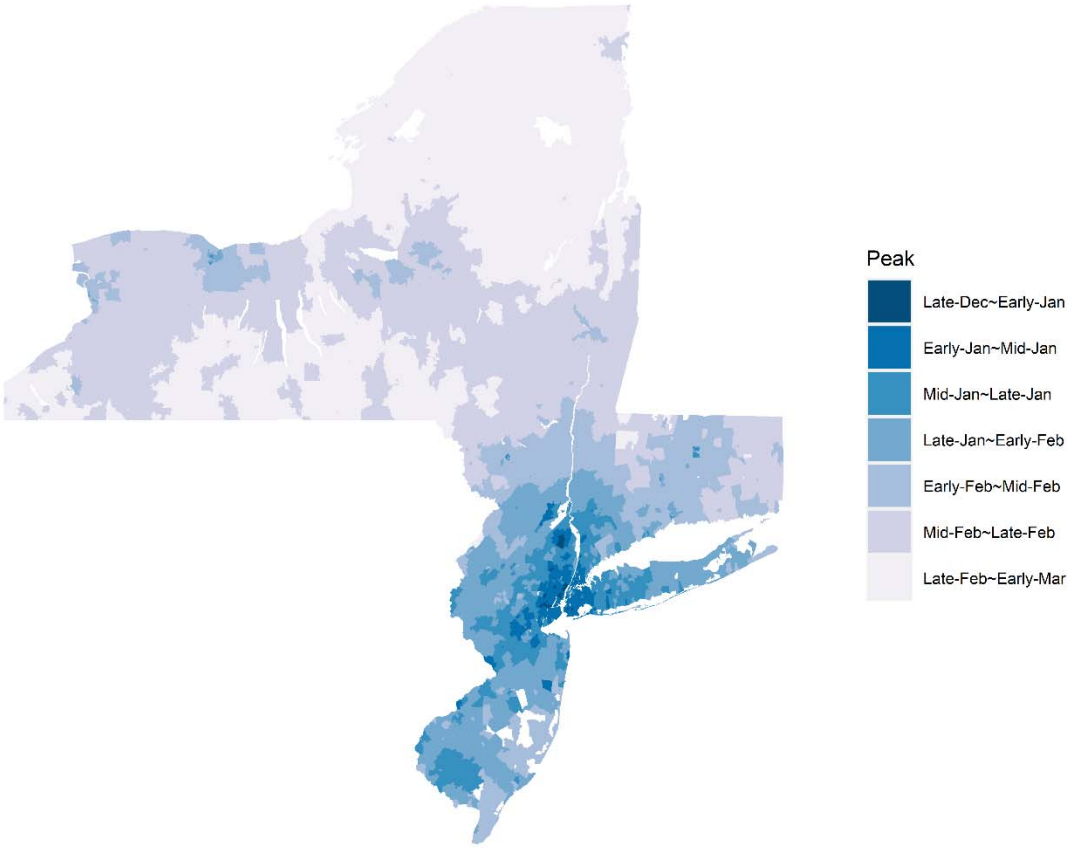


Fig. 2. Estimated peak timing of RSV epidemics by ZIP code from the best-fit model. This model accounted for average household size, population density, population size, median income, school district, and geographic proximity. The study periods were July 1997 - June 2013 in Connecticut and July 2005 - June 2014 in New York and New Jersey.

Table 1. Number of hospitalizations, ZIP codes, school districts, commuters, and study period in New Jersey, New York, and Connecticut.

	New Jersey	New York	Connecticut
RSV Hospitalizations	19,708	38,376	9,160
ZIP codes	592	1,745	275
School districts*	337	934	156
Commuters	4,295,718	8,689,118	1,725,973
Study period	July 2005 - June 2014	July 2005 - June 2014	July 1997 - June 2013

* We assigned the ZIP codes that do not belong to any school district as a single school district (in total 40) in the analysis.

Table 2. The difference in epidemic timing (in weeks) between the top and bottom deciles for each variable and coefficients of spatial correlation in New Jersey, New York, and Connecticut

	New Jersey		New York		Connecticut	
	Mean	95% CrI	Mean	95% CrI	Mean	95% CrI
	Difference in epidemic timing (weeks)					
Household Size	-0.13	(-0.48, 0.22)	-0.65	(-0.96, -0.35)	-1.26	(-1.74, -0.78)
Population Density[+]	-1.48	(-2.35, -0.56)	-2.13	(-3.04, -1.13)	-1.61	(-2.78, -0.43)
Population Size[+]	-1.00	(-2.17, 0.17)	0.35	(-0.61, 1.43)	0.16	(-0.74, 1.30)
Income[+]	0.91	(0.35, 1.43)	0	(-0.35, 0.35)	0.22	(-0.78, 0.35)
Spatial autocorrelation (ρ)*						
	0.97	(0.87, 1.00)	1	(0.99, 1.00)	0.76	(0.27, 0.99)

[+] Population density, population size and income (household median income) were log transformed in the model.

***** The posterior mean of the spatial autocorrelation parameter, ρ, was ≥ 0.75 across the three states, suggesting that the residual variability was spatially structured.

Table 3. DIC scores of competing models in New Jersey, New York, and Connecticut.

	New Jersey	New York	Connecticut
No Spatial Correlation	1365	7098	601
Geographic Proximity	1128	6347	472
Commuting Flows	1151	6449	476

Table 4. Percent of random effect variation attributable to school districts versus geographic proximity for models fitted to RSV hospitalizations in New Jersey, New York, and Connecticut. Posterior means and 95% credible intervals are displayed.

	New Jersey	New York	Connecticut
School Districts	6% (3%, 11%)	6% (3%, 11%)	31% (12%, 56%)
Geographic Proximity	94% (89%, 97%)	94% (89%, 97%)	69% (44%, 88%)

Supplementary Materials

Contents

Supplementary Materials and Methods 2

Two-stage model to estimate epidemic peak timing 2

First Stage 2

Second Stage 2

(1) Version 1 2

(2) Version 2. 3

(3) Version 3 4

Percentage of variability explained by fixed effects: 5

Percentage of variability explained by school district: 5

Percentage of variability explained by spatial proximity: 5

Difference in epidemic timing between the top and bottom deciles for each variable: 6

School District Interpolation 6

Supplementary Figures 7

Fig. S1. Proposed causal diagram of putative drivers of RSV epidemic timing 7

Fig. S2 Histogram of RSV cases at the ZIP code level. 9

Fig. S3. Maps of the covariates. 10

Fig. S4. Estimated peak timing of RSV epidemics by ZIP code from the competing models.. 11

Fig. S5. Epidemic peak timing estimation and log-precision of competing models. 12

Fig. S6. Maps comparing the predicted value of epidemic peak timing of the best-fit spatial proximity model and the marginal predicted value (prediction based on fixed effect only). 14

Supplementary Tables 15

Table S1 Descriptive statistics and sociodemographic factors by ZIP code in New Jersey, New York, and Connecticut, 2010. 15

Additional Analysis 16

Sensitivity Analysis on Time Series Data 16

Sensitivity Analysis on Socio-demographic Data 17

Supplementary Materials and Methods

Two-stage model to estimate epidemic peak timing

First Stage

The priors of the first stage model were as follows:

$$\alpha_i, \theta_i, \ln(\xi_i) \overset{\text{iid}}{\sim} \mathrm{N}(0, 1000)$$

$$\varphi_{it} \overset{\text{iid}}{\sim} \mathrm{N}(0, \tau^2)$$

$$\tau^2 \sim \mathrm{Gamma}(0.01, 0.01)$$

Second Stage

The priors of the coefficients of the input variables were:

$$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4 \overset{\text{iid}}{\sim} \mathrm{N}(0, 1000)$$

The prior of the variance $\boldsymbol{\sigma_\eta^2}$ of the school district random effects $\boldsymbol{\eta_{d(i)}}$ was:

$$\boldsymbol{\sigma_\eta^2 \sim \mathrm{Gamma}(0.01, 0.01)}$$

The details of the structures of $\boldsymbol{\phi_i}$ and the corresponding priors for the alternative versions of $\boldsymbol{\phi_i}$ were:

(1) Version 1: Assumes there is no spatial correlation in RSV epidemic timing within the three states after adjusting for ZIP code-level covariates and shared variability within school districts. In order to obtain desired levels of statistical precision and avoid extreme rates based on small local sample sizes, the non-spatial model uses an *i.i.d.* random intercept to pool information across all small areas, then the model pulls estimates toward

a global mean without regard to their relative spatial locations. The priors of this structure were as follows:

$$\boldsymbol{\phi_i} \overset{\textbf{iid}}{\sim} \mathbf{N}(\mathbf{0}, \boldsymbol{\sigma_\phi^2})$$

$$\boldsymbol{\sigma_\phi^2} \sim \mathbf{Gamma}(\mathbf{0.01}, \mathbf{0.01})$$

The model variables are defined as:

- $\boldsymbol{\phi_i}$: ZIP-code-level random effect
- $\boldsymbol{\sigma_\phi^2}$: variance of ZIP code random effect

(2) Version 2: Assumes that RSV epidemics have similar timing in neighboring geographic locations after adjusting for ZIP-code-level covariates and shared variability within the same school district. In this structure, we adopted a conditional autoregressive (CAR) prior model with a Leroux formulation for smoothing the spatial random effect term, which allowed the models to have a single random intercept accounting for both spatially structured and unstructured variation (*43*). The degree of smoothing in spatial and non-spatial components of the random effect was determined by the empirical data. Spatial correlation was estimated by ρ. The neighborhood matrix contained information about spatial proximity between two locations. Here, spatial proximity was defined as ZIP codes with adjacent borders. The model is given as:

$$\phi_i | \phi_{-i}, \tau^2, \rho \sim \mathrm{N}\left(\frac{\rho \Sigma_{j=1}^{n} w_{ij} \phi_j}{\rho \Sigma_{j=1}^{n} w_{ij} + 1 - \rho}, \frac{\sigma_\phi^2}{\rho \Sigma_{j=1}^{n} w_{ij} + 1 - \rho}\right)$$

$$\boldsymbol{\rho} \sim \mathbf{Uniform}(\mathbf{0}, \mathbf{1})$$

$$\boldsymbol{\sigma_\phi^2} \sim \mathbf{Gamma}(\mathbf{0.01}, \mathbf{0.01})$$

- $\boldsymbol{n}$: total number of ZIP code areas.
- $\mathbf{w}_{\boldsymbol{ij}}$: equal to one if regions *i* and *j* share a common border, equal to zero otherwise ($\mathbf{w}_{\boldsymbol{ii}}$ = 0 for all *i*).
- $\boldsymbol{\rho}$: spatial autocorrelation parameter; $\boldsymbol{\rho} \in [\mathbf{0}, \mathbf{1})$, where $\boldsymbol{\rho} = \mathbf{0}$ corresponds to spatial independence, simplifying the model to an independent random effects model, and $\boldsymbol{\rho}$ near 1 means strong spatial autocorrelation. However, $\boldsymbol{\rho}$ should not equal to 1, since this corresponds to an improper intrinsic autoregressive (IAR) model(*44*, *47*).
- $\boldsymbol{\sigma}_{\boldsymbol{\phi}}^{\mathbf{2}}$: variance of spatial proximity random effect.

(3) Version 3: Assumes RSV epidemic timing is explained by commuting flows after adjusting for ZIP-code-level covariates and school districts. The model structure was similar to the second hypothesis, but instead of defining the neighborhood matrix based on shared geographic borders, spatial proximity in the commuting model (W) was defined by the frequency of commuters between two locations. The overall model structure was the same as the model structure of the geographic proximity model. The only difference is the definition of w_{ij} matrix which is given as:

$$w_{ij} = \begin{cases} \sum \nu_{ij} & \text{if } i = j \\ \nu_{ij} & \text{if } i \sim j \\ 0 & \text{otherwise} \end{cases}$$

where $i \sim j$ suggests that there are more than 5 commuters between ZIP codes *i* and *j*; the element ν_{ij} is defined by the maximum commuting frequency between ZIP codes *i* and *j*.

We used the following equations to calculate the percentage of variability in RSV epidemic timing explained by the various components of the best-fit model:

Percentage of variability explained by fixed effects:

$$\frac{\left(\sigma_{\eta}^{2}+\sigma_{\phi}^{2}\right)-\left(\sigma_{\eta}^{\prime 2}+\sigma^{\prime 2}{}_{\phi}\right)}{\sigma^{\prime 2}{}_{\eta}+\sigma_{\phi}^{\prime 2}} * 100\%$$

- $\boldsymbol{\sigma_{\eta}^{2}}$: variance of school district effect in the spatial correlated model with fixed effects
- $\boldsymbol{\sigma_{\phi}^{2}}$: variance of spatial proximity random effect in the spatial correlated model with fixed effects
- $\boldsymbol{\sigma^{\prime 2}{}_{\eta}}$: variance of school district effect in the spatial correlated model without fixed effects
- $\boldsymbol{\sigma^{\prime 2}{}_{\phi}}$: variance of spatial proximity random effect in the spatial correlated model without fixed effects

Percentage of variability explained by school district:

$$\frac{\boldsymbol{\sigma_{\eta}^{2}}}{\boldsymbol{\sigma_{\eta}^{2}+\sigma_{\phi}^{2}}}$$

Percentage of variability explained by spatial proximity:

$$\frac{\boldsymbol{\sigma_{\phi}^{2}}}{\boldsymbol{\sigma_{\eta}^{2}+\sigma_{\phi}^{2}}}$$

The difference in epidemic timing between the top and bottom decile for each variable (in months) was calculated as:

$$\Delta\mathbf{month} = \frac{\mathbf{12e^{\theta_{i90th}}}}{\mathbf{1 + e^{\theta_{i90th}}}} - \frac{\mathbf{12e^{\theta_{i10th}}}}{\mathbf{1 + e^{\theta_{i10th}}}}$$

$$\boldsymbol{\theta}_{\mathbf{i90th}} = \boldsymbol{\beta}_{\mathbf{0}} + \boldsymbol{\beta}_{i}\boldsymbol{X}_{i90th}$$

$$\boldsymbol{\theta}_{\mathbf{i10th}} = \boldsymbol{\beta}_{\mathbf{0}} + \boldsymbol{\beta}_{i}\boldsymbol{X}_{i10th}$$

School District Interpolation

RSV hospitalizations were reported at the ZIP-code level. We applied an areal interpolation process to derive estimates at the school-district level from the ZIP-code level data. First, we calculated the intersection between ZIP codes and school districts. Then we measured their areal weight and multiplied the value by the weight to get an estimate of hospitalizations for each intersection. We then aggregated these estimated values to get an estimate for each school district (*50*). The analysis was performed using the areal package version 0.1.6 (*51*).

Supplementary Figures

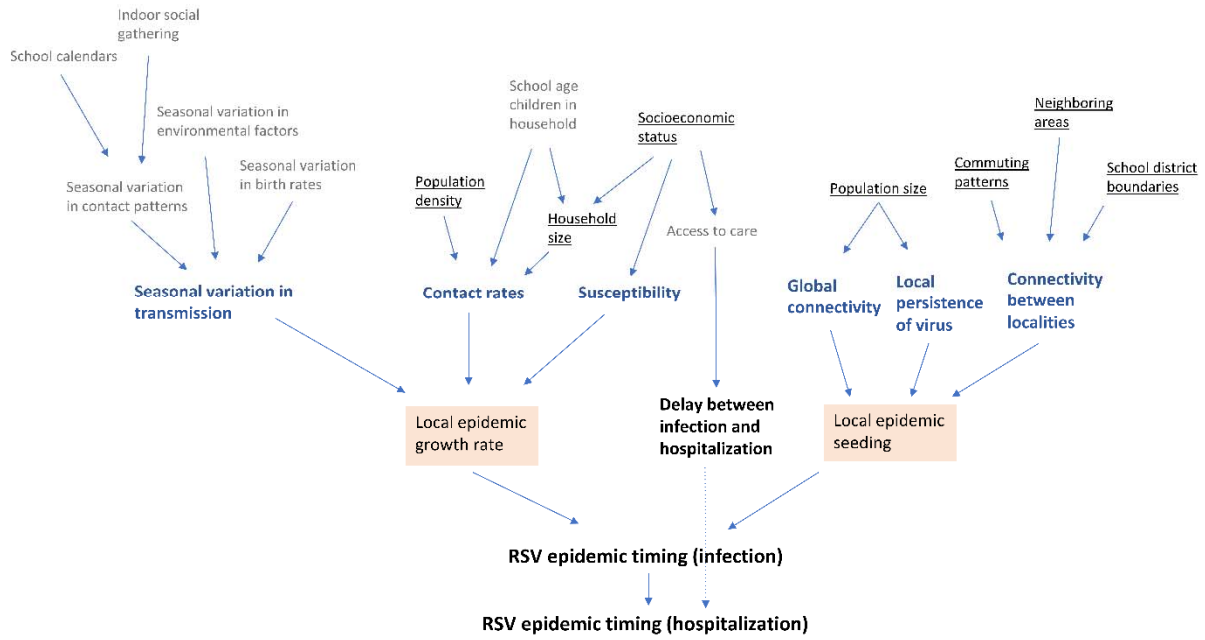


Fig. S1. Proposed causal diagram of putative drivers of RSV epidemic timing

There are three levels of factors that influence the RSV epidemic timing. The most direct factors are highlighted in the first level. The factors that may contribute to epidemic growth and epidemic seeding are in the second level with bold font. The third level is the individual variables that affect the factors in the second level. We underlined the third-level factors that are included in our analysis. The reasons for choosing these factors are as follows:

While school calendars are likely important, school start dates do not vary widely across this region and so are unlikely to affect the timing of seasonal RSV epidemics. We also did not include environmental factors because we consider finer-scale spatial variation. Seasonal variation in birth rates could influence the size of susceptible populations in the RSV season. However, we did not have this information available at the monthly ZIP-code level. The share of school-age children in a household may affect possible importation of RSV into the household

from schools, but its effect was non-significant in models that also included household size, and it is essentially capturing the same underlying process: that higher contact opportunities lead to rapid epidemic growth. Therefore, we only included the household size as a covariate. We examined the correlation with the remaining covariates and included all available factors.

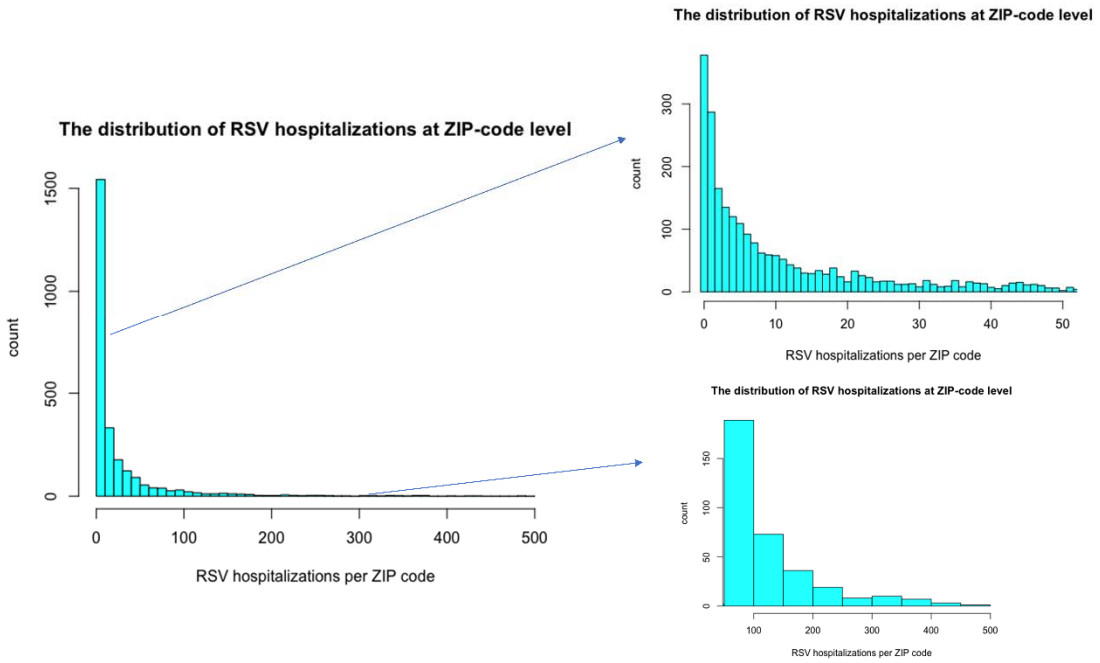


Fig. S2 Histogram of RSV cases at the ZIP code level. The cut-off point for the subunit plots is 50 RSV hospitalizations in a ZIP code.

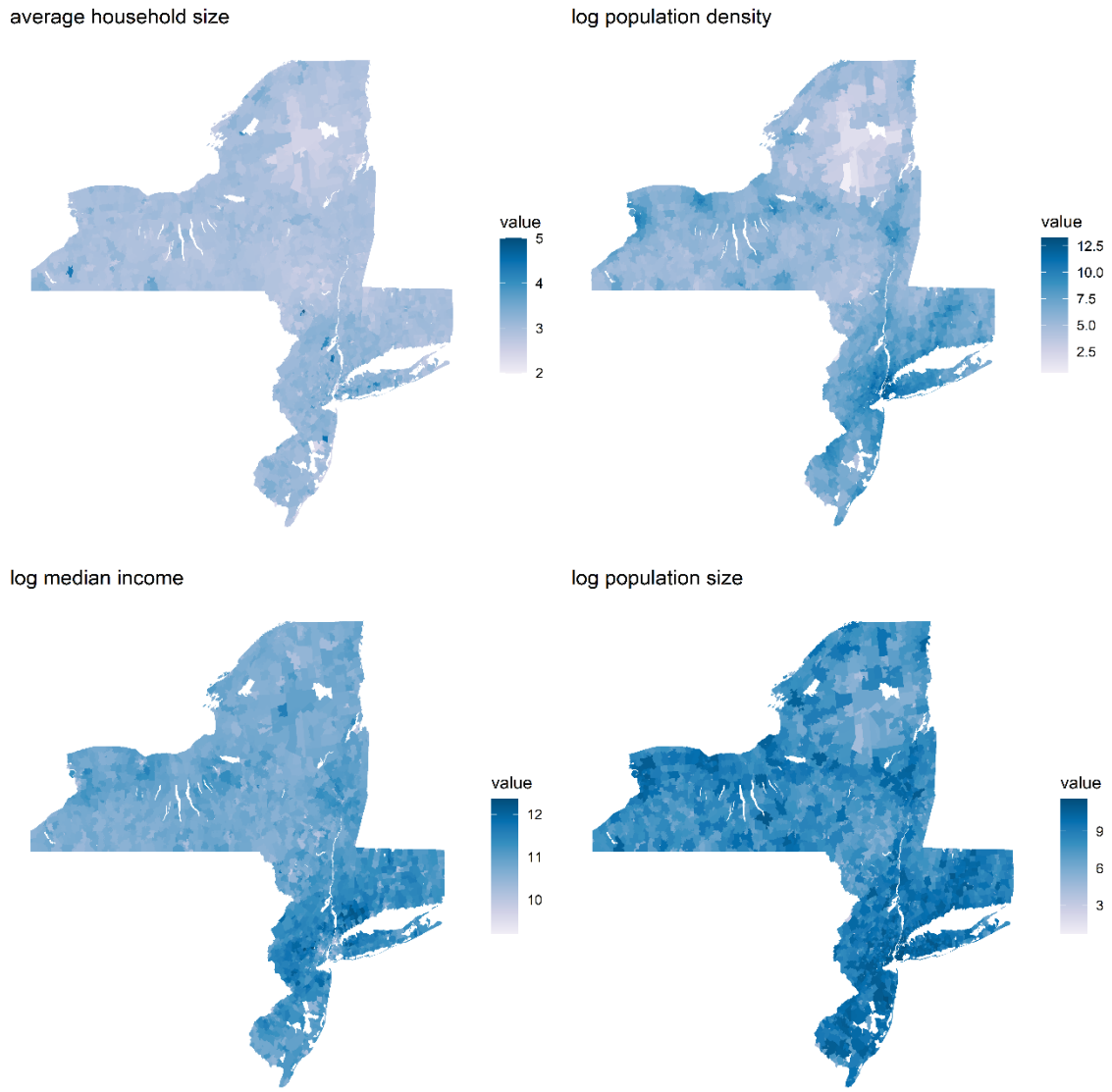


Fig. S3. Maps of the covariates. The color bars are used to indicate the value of the covariate in each ZIP code, with darker colors indicating higher values. The white spots in the map are locations that are not associated with RSV hospitalizations such as water bodies, forests and large facilities.

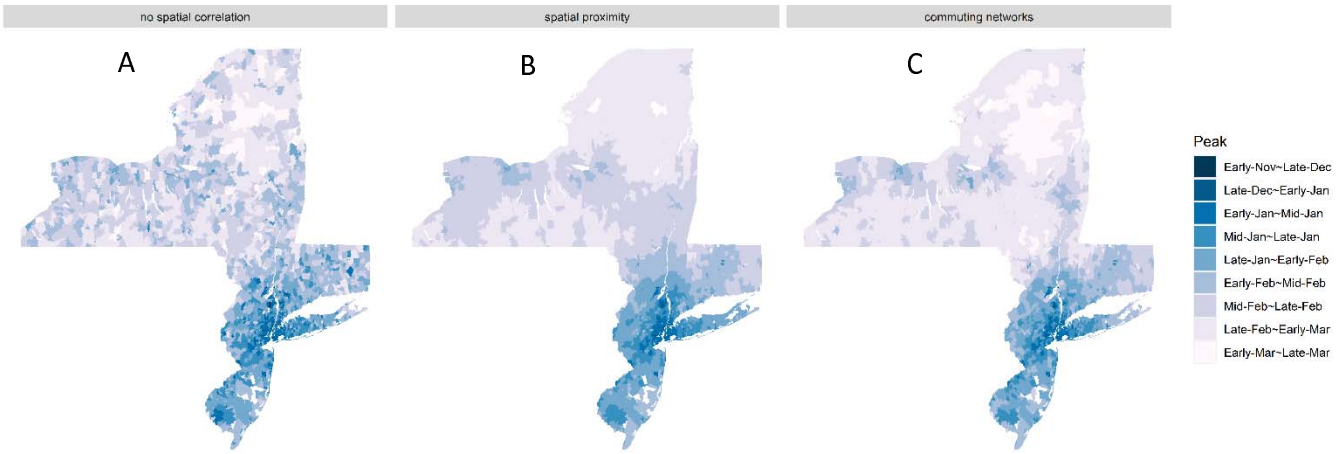
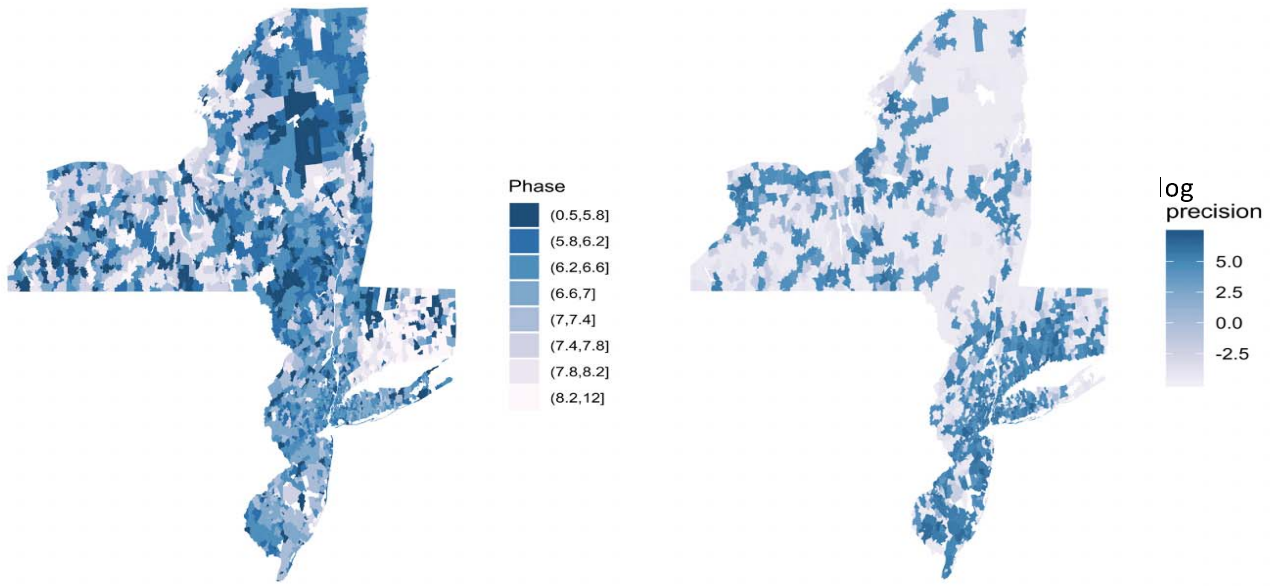


Fig. S4. Estimated peak timing of RSV epidemics by ZIP code from the competing models. **(A)** No spatial correlation: This model assumed no spatial correlation while accounting for average household size, population density, population size, median income, and school district. **(B)** The best-fit model: This model accounted for average household size, population density, population size, median income, school district, and geographic proximity. **(C)** Commuting network model: This model accounted for average household size, population density, population size, median income, school district, and commuting flows. The study periods are July 1997 - June 2013 in Connecticut and July 2005 - June 2014 in New York and New Jersey.

Stage One



Stage Two Log-Precision Comparison

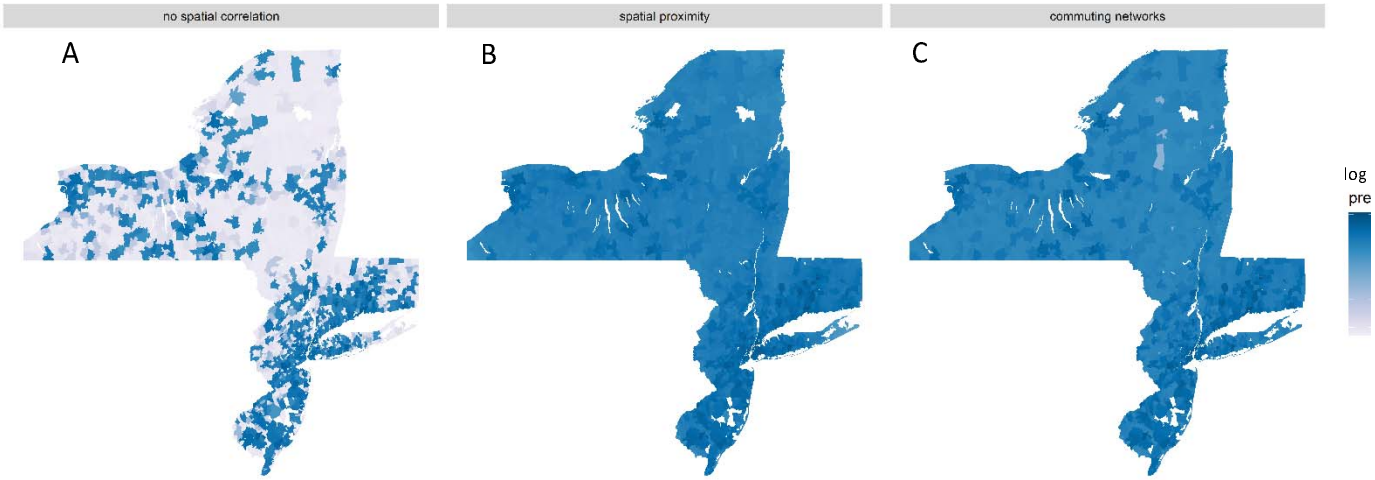


Fig. S5. Epidemic peak timing estimation and log-precision of competing models. In the Stage One panel, the left panel shows the estimated RSV epidemic peak timing and the right panel shows the log-precision of the first stage model. In the Stage Two panel, from left to right are the precision after log transform of the model that assumed no spatial correlation, the spatial

proximity model, and the commuting network model, respectively. Darker colors indicate greater precision in the estimated values. Before applying the second stage model, estimates of peak timing from the first stage model were more variable overall, including a few extreme values. The estimates of many ZIP codes had low precision in the stage one model. The geographic-distance-based spatial correlation structure pooled information across nearby small areas, obtaining desired levels of statistical precision and avoiding extreme values based on small local sample sizes.

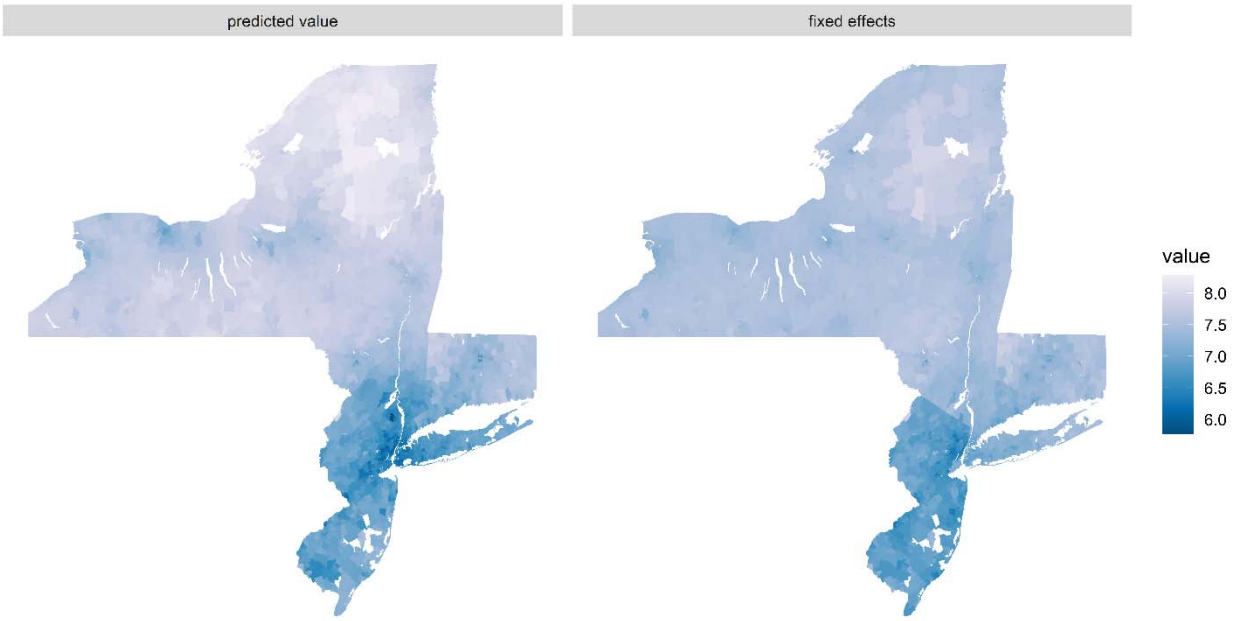


Fig. S6. Maps comparing the predicted value of epidemic peak timing of the best-fit spatial proximity model and the marginal predicted value (prediction based on fixed effect only). The left panel shows the model predictions of epidemic peak timing from the spatial proximity model and the right panel shows the predictions for the fitted model based only on the fixed effects (random effects held to 0).

Supplementary Tables

Table S1 Descriptive statistics and sociodemographic factors by ZIP code in New Jersey, New York, and Connecticut, 2010.

	Mean (Range)
Population Size under Five	713 (1, 16177)
Population Density (persons/square kilometers)*	1781 (2, 57248)
Population Size*	12185 (2, 109931)
Median Income*[]	68197 (9983, 232031)
Average Household Size[§]	3 (2, 5)
Number of Commuters	14 (2, 15598)

* log transformed in the analysis

[] in 2011 inflation-adjusted dollars

§ family type of household

Additional Analyses

Sensitivity Analysis on Time Series Data

To examine the robustness of our analysis, we compared the phase estimates from models fitted to monthly aggregated data to the phase estimates from models fitted to weekly data in Connecticut. We aggregated the original weekly time series data in Connecticut to the monthly level based on the month that the first dates of the week fell into. We then ran the hierarchical Bayesian model without covariates and spatial structures twice with the two types of data. We then measured their similarity by constructing a linear regression and calculating the slope and intercept (Figure S7).

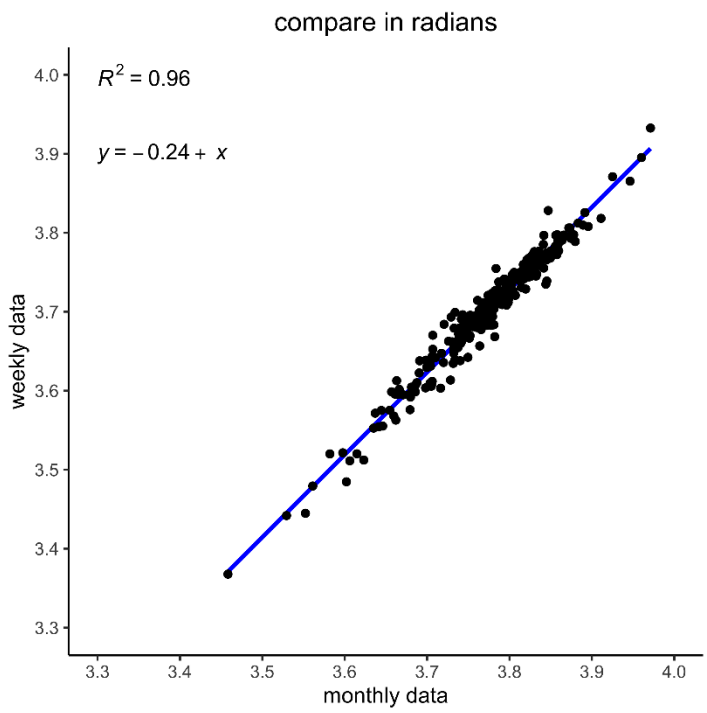


Fig. S7. Posterior estimation using weekly data vs monthly data in Connecticut. The regression coefficient 1.04 showed that phase estimates based on models fitted to aggregated monthly data are consistent with those estimated using weekly data.

Sensitivity Analysis on Socio-demographic Data

We also ran our models using the socio-demographic data from 2000 in Connecticut and 2014 in New Jersey as covariates to ensure the associations observed are robust to potential changes in socio-demographics over time. Due to the lack of data on commuting patterns before 2002, we only tested the relative importance of spatial structure based on the commuting data from 2002 in Connecticut and 2014 in New Jersey. These data covered most of the time period of our study.

The DIC scores in both sensitivity analyses were lowest for the model that assumed a spatial correlation defined by neighboring areas (Table S3), which is consistent with our main findings. The results of standardized coefficient analyses and percentage of random effects explained by spatial proximity showed similar patterns in the sensitivity analysis and our main analysis (Table S4). The fixed effects explained 41% of the total variability in New Jersey and 59% in Connecticut (Table S5). Overall, the choice of socio-demographic data and commuting patterns did not affect the conclusions of our study.

Table S3. DIC scores of competing models fitted to RSV hospitalizations in New Jersey and Connecticut using socio-demographic data from 2014 (New Jersey) and 2000 (Connecticut).

	New Jersey	Connecticut
No Spatial Correlation	1174	494
Geographic Proximity	1133	374
Commuting Flows	1151	378

Table S4. Posterior estimates of standardized fixed-effect coefficients for models fitted to RSV hospitalizations in New Jersey and Connecticut using socio-demographic data from 2014 (New Jersey) and 2000 (Connecticut) versus data from 2010.

	New Jersey				Connecticut			
	2014		2010		2000		2010	
	Mean	95% CrI	Mean	95% CrI	Mean	95% CrI	Mean	95% CrI
Household Size	0.00	(-0.02, 0.02)	-0.01	(-0.02, 0.01)	-0.04	(-0.05, -0.02)	-0.04	(-0.06, -0.03)
Population Density	-0.06	(-0.08, -0.03)	-0.05	(-0.07, -0.02)	-0.10	(-0.13, -0.06)	-0.05	(-0.09, -0.01)
Population Size	-0.02	(-0.05, 0.01)	-0.02	(-0.05, 0.01)	0.01	(-0.02, 0.04)	0.01	(-0.03, 0.05)
Income	0.03	(0.01, 0.05)	0.03	(0.01, 0.04)	-0.02	(-0.04, 0.01)	-0.01	(-0.03, 0.01)

Table S5. Percent of residual variation attributable to each random effect for models fitted to RSV hospitalizations in New Jersey and Connecticut using socio-demographic data from 2014 (New Jersey) and 2000 (Connecticut). Posterior means and 95% credible intervals are displayed.

	New Jersey	Connecticut
School District	11% (4%, 23%)	36% (15%, 63%)
Geographic Proximity	89% (77%, 96%)	64% (37%, 85%)