RESEARCH

# Health inequities in influenza transmission and surveillance

Casey M. Zipfel* and Shweta Bansal

*Correspondence:
cmz27@georgetown.edu
Department of Biology,
Georgetown University,
Washington DC, USA
Full list of author information is available at the end of the article

**Abstract**

**Motivation:** The lower an individual's socio-economic position, the higher their risk of poor health in low-, middle-, and high-income settings alike. As health inequities grow, it is imperative that we develop an empirically-driven mechanistic understanding of the determinants of health disparities, and capture disease burden in at-risk populations to prevent exacerbation of disparities. Past work has been limited in data or scope and has thus fallen short of generating generalizable insights.

**Approach & Results** Here, we integrate empirical data from observational studies and large-scale healthcare data with models to characterize the dynamics and spatial heterogeneity of health disparities in an infectious disease case study: influenza. We find that variation in social, behavioral, and physiological determinants exacerbates influenza epidemics, and that low SES individuals disproportionately bear the burden of infection. We also identify geographical hotspots of disproportionate influenza burden in low SES populations, and find that these differences are most predicted by variation in healthcare utilization and susceptibility.

**Conclusion** The negative association between health and socio-economic prosperity has a long history in the epidemiological literature. Addressing health inequities in respiratory infectious disease burden is an important step towards social justice in public health, and ignoring them promises to pose a serious threat to the entire population. Our results highlight that the effect of overlapping behavioral social, and physiological factors is synergistic and that reducing this intersectionality can significantly reduce inequities. Additionally, health disparities are expressed geographically, as targeting public health efforts spatially may be an efficient use of resources to abate inequities.

**Keywords:** Health disparities, influenza, network epidemiology, spatial heterogeneity

## Introduction

Health disparities are differences in health outcomes between social groups, and they persist in all modern public health settings. Health disparities may be the result of health inequalities, which are caused by biological or cultural variations, or by health inequities, which are driven by unfair factors and are avoidable with policy action [1]. There is extensive evidence that social factors, including education, employment, income, and ethnicity have a distinct influence on how healthy a person is: the lower an individual's socio-economic position, the higher their risk of poor health for both chronic and infectious diseases in low-, middle-, and high-income settings alike [2]. There is also a role played by geographic context in which the spatial distribution of disparity in health cannot be explained by variation in social factors alone [3]. As

the divide in health disparities grows wider across the world and within countries, it is imperative that we continue to understand how social determinants impact health at the population-level, and how this is reflected geographically [4]. Here, we integrate empirical insights from past studies to characterize the impact of social determinants on the dynamics and spatial heterogeneity in an infectious disease case study, influenza.

Influenza is a respiratory infectious disease that occurs in annual epidemics in temperate regions that can have severe outcomes, especially in young children and elderly [5]. Several studies have demonstrated social differences in influenza morbidity and mortality [6, 7, 8, 9, 10, 11]. For severe influenza, the most impoverished areas have been shown to experience twice the influenza hospitalizations compared to regions with the lowest rates of poverty [12], and low education has been shown to be positively associated with influenza hospitalization rates [13]. Past work has even shown that socio-economic factors played a significant role in the morbidity and mortality caused by the 1918 influenza pandemic [14, 15, 16]. The proposed determinants of disparities in influenza burden include a number of physiological and socio-behavioral dimensions [17, 18]. In particular, influenza vaccine coverage and healthcare access are higher in areas with increased levels of education and household income [19, 20]. Additionally, low socio-economic status (SES) individuals have been shown to experience increased susceptibility to respiratory infections due to increased stress [21, 22] and have less access to paid sick leave, resulting in less school and workplace absenteeism when ill [23, 24]. Lastly, it has been proposed that the social patterns of low SES populations affect their influenza risk: Larger household sizes and higher population density may lead to higher infection risk [25, 26], while a less robust social network might result to decreased exposure, but also less support during recovery if infected [18].

Mathematical modeling studies of social disparities in influenza burden have used a simulation approach [27, 28, 29] and have focused on the effects of material deprivation (i.e. lack of access from income, education, employment) or social deprivation (i.e. lack of social cohesion and support due to small household sizes, single parenting, divorce or widowing). Such studies are important in uncovering the mechanistic explanations of influenza disparities, but have been limited in their geographical extent, or by the use of proxy measures. For example, [27, 29] consider phenomenological variation in social contact rates without empirical evidence linking vulnerable groups to that variation, thus limiting insights on the mechanisms that lead influenza disparities; [28, 29] focus on dynamics within specific cities, limiting generalizability.

Surveillance-based statistical studies of influenza disparities have been spatial in nature and have highlighted the challenges of disease surveillance under these disparities. Surveillance systems gather the data that shapes our understanding of influenza dynamics, and in the US and most European countries, influenza-like illness (ILI) surveillance occurs through reporting by sentinel healthcare providers. Such sentinel surveillance systems have been resource-efficient means of collecting high quality data, but they do not reliably capture data for all populations, since they are dependent on health care accessibility, health care seeking behavior, and other reporting issues [30, 31]. As a result, studies that rely on healthcare data for

characterizing rates of ILI sometimes find decreasing rates of disease with increasing social deprivation [18]. While this negative association may be the result of lower exposure in impoverished areas (as suggested by [18]), it is likely that there exist spatial and social heterogeneities in surveillance caused by healthcare utilization. Indeed, Scarpino et al. have shown that the most impoverished areas are blindspots in the US influenza sentinel surveillance system, ILINet, and models based on these data make the best predictions in affluent areas while making the worst predictions in impoverished locations [32]. To better understand and respond to influenza epidemics and pandemics, we must improve our capability to detect and monitor outbreaks in at-risk populations.

In this work, we (a) develop data-driven epidemiological models to assess how social, behavioral and physiological determinants impact population-level influenza transmission in a controlled manner; and (b) develop statistical ecological models from large-scale disease data to estimate latent influenza burden in vulnerable populations in the United States. We hypothesize that low SES populations bear a disproportionate burden of influenza infection, and that a combination of social, economic and health factors cause this disparity. We aim to identify geographic areas where burden is highest in low SES populations to provide hotspots for additional surveillance. As health disparities widen, it is imperative that we develop an empirically-driven mechanistic understanding of the determinants of health disparities, and capture disease burden in at-risk populations. Such insights can allow for improved influenza forecasting, resource allocation and target intervention design.

## Results

Here, we have evaluated the impact of social, behavioral, and physiological mechanisms on driving influenza disparities. We achieved this through epidemiological model experiments in a population network with realistic SES-based contact patterns. This increases our understanding of the role that SES-driven variation plays in determining influenza dynamics. This also allows us to disentangle the effects of multiple proposed drivers of disproportionate burden in low SES. We have also assessed the impacts of low SES on influenza at the population level. We estimated low SES ILI at the county-level across the United States, accounting for variation in social, economic and health factors, as well as measurement biases. This provides estimates of low SES ILI burden at a fine spatial scale, identifying areas which are likely currently overlooked by influenza surveillance systems. These findings also provide an understanding of SES-based factors associated with disproportionate burden at the population level, which could guide future public health efforts to reduce socioeconomic health disparities.

### Contact patterns vary by socioeconomic status

We used an egocentric exponential random graph model (ERGM) to simulate networks with realistic social contact patterns based on socioeconomic status (SES) (measured by education level, [33]) from the POLYMOD social contact survey. (Additional model details can be found in Methods). The fitted network model is consistent with the contact heterogeneity in the data (Figure 1A), and all individual-level attributes (i.e. age, sex, contact location, and education level) are significant in predicting contact structure (Table S1). Additionally, our method to vary population

SES composition (details in Methods) results in consistency in network structure based on degree and assortative degree by SES-status (Figure 1B). Thus, networks with increased representation of low SES individuals maintain the same SES-based contact patterns as the POLYMOD data. Importantly, the network model captures variation in contact structure by SES. In particular, low-SES individuals have lower mean degree and variation in degree (Figure 1C), but have higher SES-assortative degree compared to those of higher SES (Figure 1D).

### Inequities increase low SES influenza transmission

The contact patterns of low SES individuals appear to differ from the rest of the population, thus it is important to consider how this network structure impacts epidemiological dynamics. To assess the role of these individuals and their behavioral and physiological differences, we integrated into an epidemiological network model of influenza transmission five key hypothesized drivers of disparities in influenza burden: a) social contact differences, or fewer social contacts and higher assortativity (as represented in our empirically-informed contact network model); b) low vaccine uptake; c) low healthcare utilization, which results in less access to influenza antivirals; d) high susceptibility, which results from stressful environmental factors; and e) low absenteeism from school or work. Figure 2A shows the low SES infection burden (i.e. the ratio of the number low SES infections and the number of all infections) in the presence of each factor, combined with social cohesion. The results can be compared against a positive control in which there is no SES-based heterogeneity in that factor. Each factor results in a significant increase in the low SES infection burden in the presence of SES-based heterogeneity, and the effect is most pronounced when all the factors occur simultaneously. In contrast, the epidemic size (i.e. the ratio of the number of infections and the population size) for the positive control is larger than the treatments, for all treatments (with the exception of the increased stress treatment) (Figure S33).

This combination of results can be explained by the role that low SES individuals play in the network. On the one hand, low SES individuals have lower mean degree (Figure 1C), resulting in a smaller epidemic size when these low degree individuals experience transmission-increasing health disparities, compared to when higher degree, higher SES individuals experience them. Thus, when SES-driven processes that increase transmission effect low SES individuals, it results in a smaller overall epidemic.On the other hand, low SES individuals have high assortativity with other low SES individuals (Figure 1D). Thus, when health disparities increase transmission of low SES individuals, they are more likely to infect other low SES individuals that are also experiencing these mechanisms, resulting in increased spread among this assortative group.

Next, we consider how the low SES infection burden scales with an increasingly large low SES population. We find that epidemic size increases with an increasing proportion of low SES individuals, and this effect appears to be driven by increasing infection of low SES individuals as they make up a larger component of the network (Figure S32). Indeed, low SES individuals experience a disproportionately large infection burden when all SES-based behavioral and physiological factors occur (Figure 2B). Additionally, high SES individuals experience a disproportionately small infection burden in the presence of the same factors.

### Low SES populations experience disproportionate influenza burden

Our results thus far characterize the mechanistic role that social, behavioral and physiological factors play on influenza burden in low-SES populations in data-driven controlled experiments. Here, we aim to characterize how macroscopic factors impact influenza dynamics in low-SES populations, integrating our theoretical findings with population-level data. For population-level influenza data, we used medical claims of ILI at the county level across the United States. This data stream has been demonstrated to provide enhanced surveillance opportunities for influenza-like illness [31, 34]. However, we find that these data suggest that ILI burden decreases with increasing low SES representation (measured by proportion of low education individuals) (Figure S34). This pattern is counter to our previous mechanistic model findings and to past small scale studies, suggesting that there may be measurement biases in these surveillance data.

To better estimate influenza burden in low SES populations, we fit a Bayesian spatial hierarchical model that accounts for measurement biases and borrows information from spatial covariates (details in Methods). Our model estimates of low SES ILI from this model show a positive relationship with the low SES population size (Figure S36) , and allow us to consider spatial disparities in influenza burden. Figure 3A shows the county-level map of low SES ILI relative risk. In counties where the relative risk is greater than one, low SES populations bear a disproportionate burden of ILI, compared to expected levels. This map highlights areas of high relative low SES ILI cases in the southeastern United States, which is a region where low SES levels are high. Other states like California and Colorado also appear to have high levels of low SES ILI relative to the size of the low SES population.

To validate our findings, we consider data on poor health among low SES individuals as reported in the CDC's Behavioral Risk Factor Surveillance System (BRFSS), a health behavior telephone survey conducted with over 400,000 adult participants annually [35]. We compare our model estimates of ILI among low SES populations to the BRFSS data on poor health in low SES populations (Figures S43 and S44), and find a moderate Pearson's correlation S42 and a Jaccard similarity index of 0.66.

### Healthcare utilization and susceptibility differences are the strongest drivers of ILI in low SES populations

Figure 3B shows the coefficient estimates and credible intervals resulting from the Bayesian spatial hierarchical model. First, the model shows that there is a positive relationship between low SES ILI and the number of physicians in the medical claims database, which represents the measurement submodel. It is expected that counties with more reporting will be positively associated with increased ILI data. The remaining covariates represent the process submodel, incorporating markers for different social and health processes believed to be associated with health disparities. Susceptibility, represented by rates of poor health and pollution levels, is positively associated with higher rates of low SES ILI. Social contact, expected to be high for larger household size and to be low for a higher proportion of single parent households, is positively associated with low SES ILI. Absenteeism, represented by chronically absent students (10 or more days of absenteeism), is negatively associated with low SES ILI. Influenza vaccine uptake, represented by adult influenza

vaccinations, is also negatively associated with low SES ILI. Additionally, healthcare utilization, represented by insurance rate and number of primary care providers, is negatively associated with low SES ILI. One environmental factor, represented by specific humidity, was also included in the model as it has previously been found to be a strong predictor of spatial variation in ILI [31]. As expected, humidity is strongly negatively related with low SES ILI.

## Discussion

Increased infectious disease prevalence among lower socio-economic status populations has been observed in many settings. What has been missing, however, is a better understanding of the mechanisms that drive this disparity. We used a mechanistic epidemiological network model which allowed us to assess the impacts of SES-based behavioral and physiological differences on influenza in controlled experiments. This highlighted the role played by all mechanisms in tandem to produce disproportionate disease burden in low SES populations. To address the gap that exists in our surveillance of ILI and to estimate the spatial distribution of influenza disparity, we then used a Bayesian spatial hierarchical model to estimate population-level low SES ILI at a fine spatial scale across the United States, accounting for disproportionate infection of low SES individuals, measurement biases, and county-level factors hypothesized to be associated with influenza and SES. Our results shine light on the spatial distribution of respiratory disease health disparities.

In our epidemiological model, disease transmission occurs over the contact network structure, which accounts for heterogeneity in contact patterns by SES. While past work has integrated contact heterogeneity by other socio-demographic characteristics such as age and occupation [36, 37], SES-based contact heterogeneity has not been integrated into contact network models for epidemiological purposes. Epidemiological simulations on the SES-heterogeneous network reveals that each hypothesized behavioral and physiological factor leads to increased infection of low SES individuals. Additionally, we find that communities with larger low SES populations experience larger epidemics, which is in agreement with small scale studies [11, 10] and with the largest observational study of influenza in low SES conducted by the CDC, though the measures evaluated are not directly comparable [12]. The proposed drivers are not mutually exclusive, so this reveals potential effects that could not be identified in past studies that investigate the impact of a single SES-based mechanism or impacts that might be aggregated in observational studies. We note that these experiments also include SES-based variation in social cohesion (i.e. SES-based contact heterogeneity in the population model), so the effect shown in Figure 2 is the result of both mechanisms combined. In Figure S33, we also illustrate the impact of each mechanism independent of social cohesion.

Our efforts to consider the impacts of low SES on influenza spatial heterogeneity generated county-level maps of disproportionate burden of ILI in low SES populations. Our findings identify pockets of disproportionate ILI burden on low SES populations across the United States, and represent a first step in filling the gap that exists in all healthcare-based surveillance. The model also produced a set of estimates for the effect of each hypothesized ecological measure. We find that all proposed factors, i.e. lack of healthcare utilization (limited access to physicians and

insurance), low vaccination coverage, low absenteeism, high social contact (large household size), and susceptibility (poor health and pollution), are significantly associated with influenza in low SES populations. This supports our previous finding that all of the mechanisms combined result in disproportionate low SES influenza burden. To validate our findings, we considered independent health data in low SES populations, and found our model estimates to be correlated and spatially similar. However, SES-stratified influenza data would be important to ground truth our model estimates.

Our work has several limitations. The network structure of our epidemiological model is based on one social survey from 2007 in Europe, and may be less representative of the United States today. Additionally, survey data was not collected for the SES of the contacts of survey participants so required us to make assumptions which could affect our results about SES assortativity. Additional social contact data collection across the United States that accounts for SES heterogeneity would be useful for future studies given the large socio-economic inequality in the country [38, 39]. In our spatial ecological model, we assume that disproportionate burden in low SES populations remains constant over influenza seasons. While this is a reasonable first assumption based on social and healthcare processes being consistent over our study period, there may be variation in the impact of ILI on low SES populations due to strain distribution and environmental features that do vary across seasons. Future work could focus on temporal variation in low SES ILI dynamics.

As the divide in health disparities grows wider across the United States, we propose the use of infectious disease case studies to improve our understanding of this challenging problem. We suggest that we move beyond studies based on proxy measures such as income and education which may provide an incomplete picture [3], and dig into the mechanisms that may be at the root of inequities. Furthermore, we advocate for the prioritization of capabilities to detect and monitor outbreaks in at-risk populations so that we may prevent exacerbation of health disparities. Addressing health inequities in respiratory infectious disease burden is an important step towards social justice in public health, and ignoring them promises to pose a serious threat to the entire population. Our results suggest that (a) the effect of overlapping behavioral and social factors is synergistic and reducing this intersectionality can significantly reduce inequities; and (b) health disparities are expressed geographically and targeting public health efforts spatially may be an efficient use of resources to abate inequities. Further attention to the mechanisms and processes that lead to health inequities, and specifically health inequities that may be overlooked by our currently surveillance systems, will be important to identifying actionable steps to mitigate negative health outcomes in the future.

## Methods

In this study, we use (1) a mechanistic network epidemiological model to assess influenza transmission in the presence of individual-level socioeconomic status (SES)-based behavioral and physiological variation; and (2) an inferential spatial model to geographically localize influenza-like illness (ILI) burden among low-SES populations in the presence of population-level variation in social and health indicators.

## Modeling Impact of Individual-Based SES factors on Disease Burden

To achieve the mechanistic understanding, we (a) fitted a contact network model from empirical contact data that includes contact heterogeneity stratified by age, sex, contact location, and socioeconomic status; and (b) performed epidemiological simulations on these networked populations integrating epidemiological differences based on SES, parameterized by empirical studies.

### *Contact Network Model*

In a contact network model, nodes represent individuals, and edges represent epidemiologically-relevant interactions between individuals. The degree of a node is the number of edges, or contacts, of the node, and the degree distribution of a network is the frequency distribution of node degrees within the population. To generate realistic contact networks to evaluate epidemic outcomes, we used an egocentric exponential random graph model (ERGM) [40]. An egocentric ERGM allows for the construction of sociocentric networks based on egocentrically sampled data, in which participants (or *egos*) report the identity of their contacts (or *alters*), who may or may not be study participants. Our egocentric ERGM model was based on the POLYMOD dataset, a large, egocentric contact survey that took place across several countries in Europe to identify close interactions of over 7000 individuals across eight European countries [41].

Nodes in the network had the following attributes: (a) age, grouped as infants-toddlers (age 0-4), school-aged children (age 5-18), adults (age 19-64), and elderly (age 65-100); (b) sex, classified as male or female; (c) contact location, in which a node can have known home contacts and known school or work contacts; (d) education level as a proxy for socioeconomic status [33], grouped as low education (less than a high school education), medium education (high school or vocational school education), or high education (any university education or beyond). Age and sex were available in the data for egos and alters, while education level was only provided for egos. Therefore, it was assumed that an ego's work contacts had the same education level based on their occupation, and that an ego's home contacts had the same education level as an indicator of household socioeconomic status. To represent communities with different SES compositions, we resampled additional low education egos from the low education sample in the POLYMOD dataset. We produce networks composed of approximately 20-60% low education individuals (Table S3).

The model was fit using the ERGM package [42, 43]. The best model was selected based on collinearity criteria and goodness of fit to the POLYMOD data. From the best fit ERGM model, we simulated 5 networks. Additional model details, including model terms (Table S1), collinearity (Table S2), model diagnostics (Figure S4), and goodness of fit (Figure S5 - Figure S29) can be found in the Supplement.

Random regular networks of the same size and mean degree were also generated as null networks to evaluate the effect of contact heterogeneity. We used the Networkx package for network generation and analysis [44].

### *SES-based Epidemiological Model*

Chain binomial SEIR (Susceptible-Exposed-Infected-Recovered) simulations were performed on the networks generated by the egocentric ERGM model and the control networks to examine the spread of a respiratory infection, like influenza, through

a naive population. Model parameters pertinent to seasonal influenza spread were selected from literature (Table S4) [45, 46].

Five hypothesized drivers for increased influenza in low SES populations were integrated into the epidemiological simulations. Each hypothesized driver represents a health behavior or physiological factor, and is represented by a single parameter, the value of which was selected from pertinent literature (Table S4). Social contact differences represents the SES-based social contact rates of individuals, and thus is represented by the ERGM-generated networks. The remaining factors are:

- Low vaccine uptake: Individuals may be vaccinated before the start of the season with a perfectly efficacious vaccine. Vaccinated nodes were randomly selected and removed from the network. Vaccination coverage is parameterized by $\delta$ and $\delta_{low}$ in high- and low-SES individuals, respectively. The value of delta was based on a US population survey of vaccine coverage related to education level [19].
- High susceptibility: Those who experience a more stressful environment are more susceptible to infection, and thus have a greater probability of becoming infected upon contact with an infected individual. Susceptibility is parameterized by $\beta$ and $\beta_{low}$ in high- and low-SES individuals, respectively. This is based on an immune challenge experiment that found that those of high SES were about half as likely to become infected with a cold compared to those of low SES [22].
- Low healthcare utilization: Infected individuals who do not seek healthcare and receive antivirals have a longer infectious period, based on a model of within-host and population-level dynamics [47]. The proportion of the infected population seeking healthcare is parameterized by $\gamma$ and $\gamma_{low}$ in high- and low-SES individuals, respectively.
- Low absenteeism: Infected individuals may exhibit absenteeism from school or work if they have access to leave and care at home. Those exhibiting absenteeism remove school or work contacts. Access to absenteeism is parameterized by $\rho$ and $\rho_{low}$ in high- and low-SES individuals, respectively. These values are based on rates of paid sick leave by education level in a survey across the US [48].

For our experimental design, each SES-based factor was tested separately and together on each network. Disease outbreaks for each treatment were simulated 200 times on each network, with 5 replicate networks. We also considered two controls to compare our experimental results: a) a homogeneous control, in which each factor was randomly distributed across a random regular network; b) a heterogeneous control, in which each factor was randomly distributed across the ERGM-generated networks.

### Modelling Impact of Disease in Low-SES Populations

To achieve an inferential understanding, we (a) integrated the network model findings with empirical ILI data for an estimate of ILI burden among low-SES individuals; and (b) fitted a spatial Bayesian hierarchical model with population-level covariates to account for measurement biases and improve our estimate of low-SES ILI burden at the population-level.

*Spatial Inferential Model*

We used a Bayesian spatial hierarchical model to estimate latent ILI cases among low SES individuals, accounting for measurement biases and county-level factors associated ILI in low SES populations. We assume that that the proportion of ILI cases occurring in low-SES populations is constant over time from 2002-2008. We modeled low-SES ILI ($Y_{it}$) in county $i$ in flu season $t$ as:

$$Y_{it}|N_i \sim Binomial(N_i, p_{i,t})$$

where $p_{i,t}$ is the probability of detecting low-SES ILI cases, and $N_i$ is the true ILI cases among low-SES individuals.

We modeled the probability of detection $p_i$ as:

$$logit(p_{i,t}) = \alpha_0 + \sum_{1}^{k} \alpha_k z_{i,t,k} + \nu_c + \nu_s$$

where $\alpha_0$ is the intercept, $\alpha_k$ represents the measurement process predictor variables, and $\nu_c$ and $\nu_s$ are group effects for county and state, respectively.

We modeled the latent low-SES ILI cases as:

$$N_i \sim NegBin(\lambda_i, \theta)$$

where the negative binomial distribution is parameterized by probability $\lambda_i$ and size $\theta$.

The $\lambda_i$ is modeled by:

$$log(\lambda_i) = \beta_0 + \sum_{1}^{j} \beta_j x_{i,j} + \mu_c + \mu_s$$

where $\beta_0$ is the intercept, $\beta_j$ represents coefficient estimates for low-SES ILI process covariates, and $\mu_c$ and $\mu_s$ represent county-level and state-level group effects, respectively. We performed approximate Bayesian inference using Integrated Nested Laplace Approximations (INLA) with the R-INLA package [49]. INLA has demonstrated computational efficiency for latent Gaussian models, produced similar estimates for fixed parameters as established implementations of Markov Chain Monte Carlo (MCMC) methods for Bayesian inference, and been applied to disease mapping and spatial ecology questions. We evaluated DIC, WAIC, model residuals and compared modeled and observed outcomes in order to assess model fit. Additional model details can be found in Figures S35, S37, S38, S39, S40, & S41.

*Response data*

We define the response in our model to be the observed influenza-like illness (ILI) burden in low-SES populations. In particular, we use influenza-like illness reports from a medical claims database from across the United States collected during 2002-2008 (Figure S30). [Additional details on the dataset can be found in [31, 34]]. To normalize these observed counts, we divide ILI visits by visits for any diagnosis

during the influenza season. These data are at the county-level but are not stratified by SES. To produce a county-level estimate of ILI in low-SES populations for our spatial model, we use the observed ILI burden in the total population and scale this by the proportion expected among low-SES individuals as predicted by the epidemiological model from the first part of our study (as summarized in Figure 2). (The response data can be seen in Figure S31).

*Covariate data*

All covariate data are at the county level. All covariate data were centered and standardized. Covariate data were included as a marker for each hypothesized driver of low SES influenza, based on what data were available at the county level. Covariate data was averaged across 2002-2008 where possible; otherwise the most appropriate temporal timeframe available was selected. For the measurement model, the number of reporting physicians was reported by the medical claims database. For a measure of susceptibility, reports of poor health and pollution levels were collected from County Health Rankings [50]. For a measure of social cohesion, mean household size and number of single parent households were collected from County Health Rankings. To measure access to healthcare, insurance rates and number of primary care physicians were collected from County Health Rankings. To measure absenteeism, the number of student who were absent for more than 10 days, noted as "chronic absenteeism" was collected from the US Department of Education [51]. To measure vaccination, reports of adult vaccination were collected from the Behavioral Risk Factor Surveillance System (BRFSS) from the CDC [35]. Humidity levels were collected from the National Oceanic and Atmospheric Administration (NOAA) to account for the spatial distribution of influenza cases in the US, as humidity was identified as the strongest driver of spatial dynamics in a prior model with medical claims ILI data [52, 31]. See supplement table for covariate data details Table S5.

*Validation*

To validate our model findings, we compared the model estimates of ILI in low SES populations with reports of poor health in low SES populations at the county-level from BRFSS. We calculated the Pearson's correlation coefficient of low SES ILI estimates and low SES poor health estimates. We also calculated the Jaccard index, based on relative risk classified as 0 or 1, where 0 represents relative risk below 1 and 1 represents relative risk above 1, for each county in both datasets.

**Competing interests**

The authors declare that they have no competing interests.

**Acknowledgements**

Research reported in this publication was supported by the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number R01GM123007. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We also acknowledge support from the PhRMA Foundation and the Chateaubriand Fellowship Program. We thank Håvard Rue for his development of and assistance with the R-INLA package.

**References**

1. Penman-Aguilar, A., Talih, M., Huang, D., Moonesinghe, R., Bouye, K., Beckles, G.: Measurement of health disparities, health inequities, and social determinants of health to support the advancement of health equity. J Public Helath Manag Pract **22**(Suppl 1), 33–42 (2016)
2. Adler, N.E., Newman, K.: Socioeconomic disparities in health: Pathways and policies. Health Affairs **21**(2), 60–76 (2002). doi:10.1377/hlthaff.21.2.60

3. Murray, C.J., Kulkarni, S.C., Michaud, C., Tomijima, N., Bulzacchelli, M.T., Iandiorio, T.J., Ezzati, M.: Eight americas: investigating mortality disparities across races, counties, and race-counties in the united states. PLoS medicine **3**(9) (2006)
4. Bosworth, B.: Increasing Disparities in Mortality by Socioeconomic Status. Annual Review of Public Health **39**(1), 237–251 (2018). doi:10.1146/annurev-publhealth-040617-014615
5. Centers for Disease Control and Prevention: Influenza (2019)
6. Biggerstaff, M., Jhung, M.A., Reed, C., Garg, S., Balluz, L., Fry, A.M., Finelli, L., Branch, P., Division, I., Disease, R., Service, I.: Impact of medical and behavioural factors on influenza-like Illness, healthcare-seeking, and antiviral treatment during the 2009 H1N1 pandemic — United States, 2009–2010. Epidemiol Infect **142**(1), 114–125 (2014). doi:10.1017/S0950268813000654.Impact
7. Lowcock, E.C., Rosella, L.C., Foisy, J., McGeer, A., Crowcroft, N.: The social determinants of health and pandemic h1n1 2009 influenza severity. American Journal of Public Health **102**(8), 51–58 (2012). doi:10.2105/AJPH.2012.300814
8. Rutter, P.D., Mytton, O.T., Mak, M., Donaldson, L.J.: Socio-economic disparities in mortality due to pandemic influenza in England. International Journal of Public Health **57**(4), 745–750 (2012). doi:10.1007/s00038-012-0337-1
9. Galvin, J.R., Cartter, M.L., Sosa, L.: Neighborhood Socioeconomic Status Among Children Hospitalized With Influenza: New Haven County, Connecticut, 2003-2010. Connecticut Epidemiologist **30**(6), 21–24 (2010)
10. Yousey-Hindes, K.M., Hadler, J.L.: Neighborhood socioeconomic status and influenza hospitalizations among children: New Haven County, Connecticut, 2003-2010. American Journal of Public Health **101**(9), 1785–1789 (2011). doi:10.2105/AJPH.2011.300224
11. Tam, K., Yousey-hindes, K., Hadler, L.: Influenza-related hospitalization of adults associated with low census tract socioeconomic status and female sex in New Haven County , Connecticut , 2007-2011. Influenza and other Respiratory Viruses **8**(3), 274–281 (2014). doi:10.1111/irv.12231
12. Hadler, J.L., Yousey-hindes, K., Pérez, A., Anderson, E.J., Bargsten, M.: Influenza-Related Hospitalizations and Poverty Levels — United States , 2010 – 2012. CDC Morbidity and Mortality Weekly Report **65**(5), 101–105 (2016)
13. Crighton, E.J., Elliott, S.J., Moineddin, R., Kanaroglou, P., Upshur, R.: A spatial analysis of the determinants of pneumonia and influenza hospitalizations in Ontario (1992-2001). Social Science and Medicine **64**(8), 1636–1650 (2007). doi:10.1016/j.socscimed.2006.12.001
14. Grantz, K.H., Rane, M.S., Salje, H., Glass, G.E., Schachterle, S.E.: Disparities in influenza mortality and transmission related to sociodemographic factors within Chicago in the pandemic of 1918. PNAS (2016). doi:10.1073/pnas.1612838113
15. Murray, C.J., Lopez, A.D., Chin, B., Feehan, D., Hill, K.H.: 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)
16. Mamelund, S.-E.: 1918 pandemic morbidity: The first wave hits the poor, the second wave hits the rich. Influenza and other respiratory viruses **12**(3), 307–313 (2018)
17. Cordoba, E., Aiello, A.E.: Social Determinants of Influenza Illness and Outbreaks in the United States. NCMJ **77**(5), 341–345 (2016)
18. Charland, K.M., Brownstein, J.S., Verma, A., Brien, S., Buckeridge, D.L.: Socio-economic disparities in the burden of seasonal influenza: The effect of social and material deprivation on rates of influenza infection. PLoS ONE **6**(2), 1–5 (2011). doi:10.1371/journal.pone.0017207
19. Linn, S.T., Guralnik, J.M., Patel, K.K.: Disparities in Influenza Vaccine Coverage in the United States, 2008. J Am Geriatric Soc **58**(7), 1333–1340 (2010). doi:10.1038/jid.2014.371. NIHMS150003
20. Derigne, L., Stoddard-dare, P., Quinn, L.: Workers without paid sick leave less likely to take time off for illness or injury compared to those with paid sick leave. Health Affairs **35**(3), 520–527 (2016)
21. Cohen, F., Kemeny, M.E., Zegans, L.S., Johnson, P., Kearney, K.A., Stites, D.P.: Immune function declines with unemployment and recovers after stressor termination. Psychosomatic Medicine **69**(3), 225–234 (2007). doi:10.1097/PSY.0b013e31803139a6
22. Cohen, S., Adler, N., Alper, C.M., Doyle, W.J., Treanor, J.J., Turner, R.B.: Objective and Subjective Socioeconomic Status and Susceptibility to the Common Cold. Health Psychology **27**(2), 268–274 (2008). doi:10.1037/0278-6133.27.2.268
23. Berendes, D., Andujar, A., Barrios, L.C., Hill, V.: Associations Among School Absenteeism , Gastrointestinal and Respiratory Illness , and Income — United States , 2010 – 2016. CDC Morbidity and Mortality Weekly Report **68**(9) (2019)
24. Blendon, R.J., Koonin, L.M., Benson, J.M., Cetron, M.S., Pollard, W.E., Mitchell, E.W., Weldon, K.J., Herrmann, M.J.: Public Response to Community Mitigation Measures for Pandemic Influenza. Emerging Infectious Diseases **14**(5) (2008)
25. Cardoso, M.R.A., Cousens, S.N., De Góes Siqueira, L.F., Alves, F.M., D'Angelo, L.A.V.: Crowding: Risk factor or protective factor for lower respiratory disease in young children? BMC Public Health **4**, 1–8 (2004). doi:10.1186/1471-2458-4-19
26. Sloan, C., Chandrasekhar, R., Mitchel, E., Schaffner, W., Lindegren, M.L.: Socioeconomic Disparities and Influenza Hospitalizations, Tennessee, USA. Emerging Infectious Diseases **21**(9) (2015)
27. Munday, J.D., Van Hoek, A.J., Edmunds, W.J., Atkins, K.E.: Quantifying the impact of social groups and vaccination on inequalities in infectious diseases using a mathematical model. BMC Medicine **16**(1), 1–12 (2018). doi:10.1186/s12916-018-1152-1
28. Hyder, A., Leung, B.: Social deprivation and burden of influenza: Testing hypotheses and gaining insights from a simulation model for the spread of influenza. Epidemics **11**, 71–79 (2015). doi:10.1016/j.epidem.2015.03.004
29. Kumar, S., Piper, K., Galloway, D.D., Hadler, J.L., Grefenstette, J.J.: Is population structure sufficient to generate area-level inequalities in influenza rates? An examination using agent-based models. BMC Public Health **15**(1), 1–12 (2015). doi:10.1186/s12889-015-2284-2

30. Who, W.H.O.: WHO global technical consultation: global standards and tools for influenza surveillance (WHO/HSE/GIP/2011.1). Who (March), 2011 (2011). doi:WHO/HSE/GIP/2011.1
31. Lee, E.C., Arab, A., Goldlust, S.M., Grenfell, B.: Deploying digital health data to optimize influenza surveillance at national and local scales. PLoS Computational Biology **14**(3), 1–23 (2018). doi:10.1371/journal.pcbi.1006020
32. Scarpino, S.V., Scott, J.G., Eggo, R.M., Clements, B., Dimitrov, N.B., Meyers, L.A.: Socioeconomic bias in influenza surveillance. arXiv (2014). arXiv:1804.00327v1
33. Thomson, S.: Achievement at school and socioeconomic background—an educational perspective. npj Science of Learning **3**(1), 5 (2018). doi:10.1038/s41539-018-0022-0
34. Viboud, C., Charu, V., Olson, D., Ballesteros, S., Gog, J., Khan, F., Grenfell, B., Simonsen, L.: Demonstrating the use of high-volume electronic medical claims data to monitor local and regional influenza activity in the US. PLoS ONE **9**(7) (2014). doi:10.1371/journal.pone.0102429
35. Centers for Disease Control and Prevention: Behavioral Risk Factor Surveillance System (2012)
36. Bansal, S., Grenfell, B.T., Meyers, L.A.: When individual behaviour matters: Homogeneous and network models in epidemiology. Journal of the Royal Society Interface **4**(16), 879–891 (2007). doi:10.1098/rsif.2007.1100
37. Cauchemez, S., Bhattarai, A., Marchbanks, T.L., Fagan, R.P., Ostroff, S., Ferguson, N.M., Swerdlow, D., Sodha, S.V., Moll, M.E., Angulo, F.J., Palekar, R., Archer, W.R., Finelli, L.: Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. Proceedings of the National Academy of Sciences **108**(7), 2825–2830 (2011). doi:10.1073/pnas.1008895108
38. Kennedy, B.P., Kawachi, I., Glass, R., Prothrow-Stith, D.: Income distribution, socioeconomic status, and self rated health in US : multilevel analysis. British Medical Journal **318**(7195), 1417–1418 (1999). doi:10.1136/bmj.318.7195.1417a
39. Noel, R.: Race, Economics, and Social Status (May), 2–9 (2018)
40. Krivitsky, P.N., Handcock, M.S., Morris, M.: Adjusting for network size and composition effects in exponential-family random graph models. Statistical Methodology **8**(4), 319–339 (2011). doi:10.1016/j.stamet.2011.01.005. NIHMS150003
41. Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G.S., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M., Edmunds, W.J.: Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Medicine **5**(3), 0381–0391 (2008). doi:10.1371/journal.pmed.0050074. arXiv:1011.1669v3
42. Handcock, M.S., Hunter, D.R., Butts, C.T., Goodreau, S.M., Krivitsky, P.N., Morris, M.: ergm: Fit, Simulate and Diagnose Exponential-Family Models for Networks (2019)
43. Hunter, D.R., Handcock, M.S., Butts, C.T., Goodreau, S.M., Morris, M.: ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks. Journal of Statistical Software **24**(3), 1–29 (2008)
44. Hagberg, A.A., Schult, D.A., Swart, P.J.: Exploring network structure, dynamics, and function using NetworkX. Proceedings of the 7th Python in Science Conference (SciPy2008) **G"ael Varoq**, 11–15 (2008)
45. Biggerstaff, M., Jhung, M.A., Reed, C., Fry, A.M., Balluz, L., Finelli, L.: Influenza-like illness, the time to seek healthcare, and influenza antiviral receipt during the 2010–11 influenza season — United States. J Infect Dis **210**(4), 535–544 (2014). doi:10.1111/mec.13536.Application
46. Carrat, F., Vergu, E., Ferguson, N.M., Lemaitre, M., Cauchemez, S., Leach, S., Valleron, A.J.: Time lines of infection and disease in human influenza: A review of volunteer challenge studies. American Journal of Epidemiology **167**(7), 775–785 (2008). doi:10.1093/aje/kwm375
47. Pepin, K.M., Riley, S., Grenfell, B.T.: Effects of Influenza Antivirals on Individual and Population Immunity Over Many Epidemic Waves. Epidemiol Infect **141**(2), 366–376 (2013). doi:10.1016/j.physbeh.2017.03.040
48. Piper, K., Youk, A., James, A.E., Kumar, S.: Paid sick days and stay-At-home behavior for influenza. PLoS ONE **12**(2), 1–13 (2017). doi:10.1371/journal.pone.0170698
49. Rue, H., Martino, S., Chopin., N.: Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). Journal of the Royal Statistical Society, Series B **71**(2), 319–392 (2009)
50. University of Wisconsion Population Health Institute: County Health Rankings and Roadmaps (2019)
51. US Department of Education: Chronic Absenteeism in the Nation's Schools (2019)
52. USA.gov: NOAA National Centers for Environmental Information

**Figures**

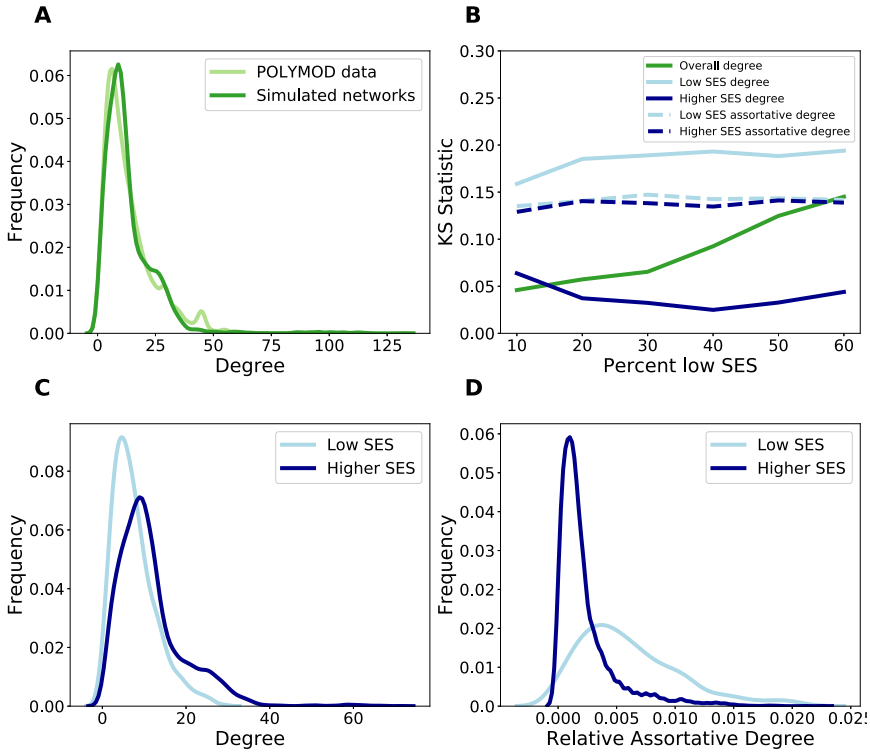


**Figure 1:** The network characteristics of the networks generated from the ERGM model based on POLYMOD data. A: The degree distribution of the POLYMOD data (light green) compared to the simulated networks (dark green). B: The Kolmogorov-Smirnov (KS) statistic to evaluate the dissimilarity of the ERGM-simulated networks to the POLYMOD data as additional low education individuals are added to the network. KS statistics compare the dissimilarity of the overall degree distribution (dark green), the degree distribution of low SES nodes (light blue, solid), the degree distribution of higher SES nodes (dark blue solid), the assortative degree (e.g. the low SES contacts of low SES nodes) for low SES nodes (light blue, dashed), and the assortative degree for higher SES nodes (dark blue, dashed). Low KS values indicate similar distributions. C: The degree distribution of low SES nodes (light blue) and higher SES nodes (dark blue). D: The relative assortative degree distribution (e.g. number of low SES contacts of low SES nodes/number of low SES nodes) of low SES nodes (light blue) and other SES nodes (dark blue).

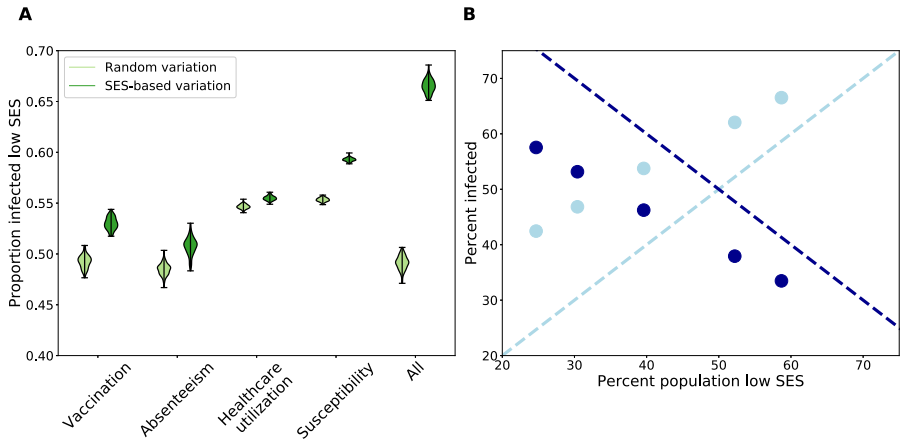


**Figure 2:** Results of epidemiological simulations on ERGM networks with SES-driven behavioral and physiological differences. A) All of the proposed SES-driven behavioral and physiological differences result in an increase in infection of low SES individuals (dark green, right of paired violin plots), compared to simulations where the differences are randomly distributed throughout the population (light green, left of paired violin plots). This difference is most pronounced when all of the mechanisms occur together. These simulations were performed on a network composed of 60% low SES, but the results are consistent across networks with different SES compositions. B) In all networks, when all SES-driven behavioral and physiological differences are present, low SES individuals (mean percent of infected population that is low SES shown in light blue dots) are disproportionately infected, relative to the expectation (light blue dashed line). High SES individuals are disproportionately underinfected compared the expectation (dark blue dots compared to dark blue dashed line).

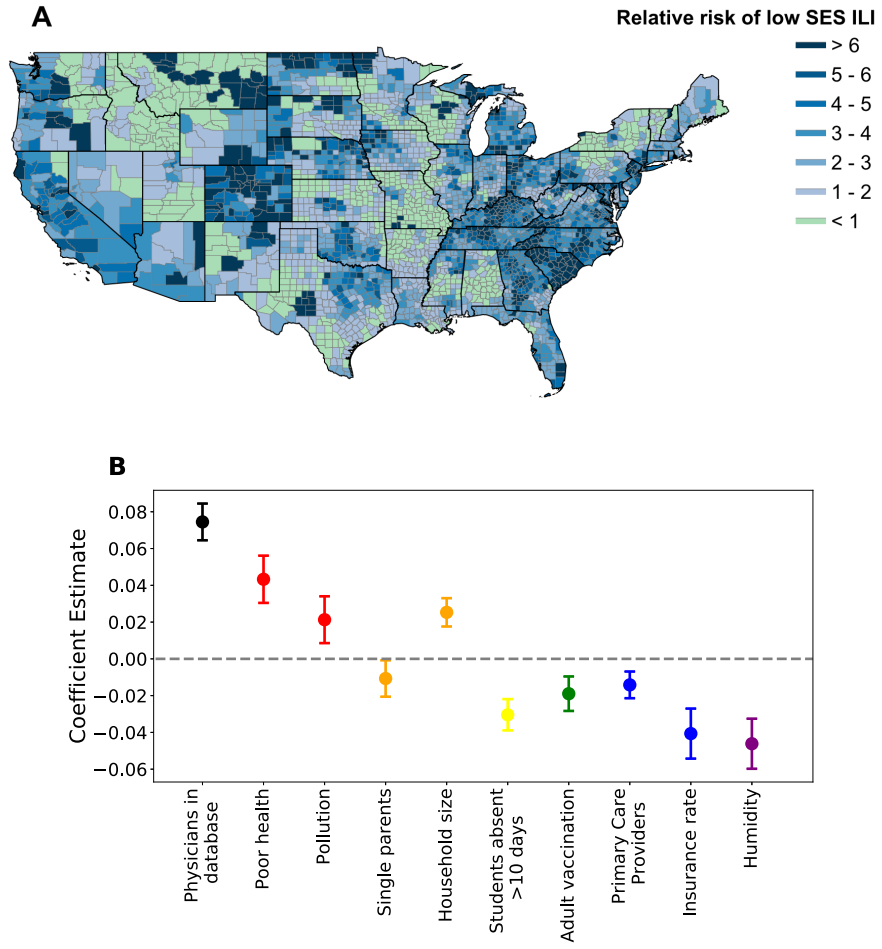


**Figure 3:** Results of spatial Bayesian hierarchical model. A) County-level map of the relative risk of low SES ILI, compared to expected levels, based on low SES population size, average observation rate, and average ILI. Counties with a value less than 1 (light green) have a disproportionately low burden of ILI in low SES population. Counties with a value greater than 1 (light blue for lower values and dark blue for higher values) have a disproportionately high burden of ILI in low SES populations. B) Mean model coefficient estimates and credible intervals. Points are colored by what process each covariate represents (black: measurement bias, red: susceptibility, orange: social contact differences, yellow: absenteeism, green: vaccination, blue: healthcare utilization, purple: environmental factors).