RESEARCH ARTICLE

The sociospatial factors of death: Analyzing effects of geospatially-distributed variables in a Bayesian mortality model for Hong Kong

Thayer Alshaabi[1,2,4]*, **David R. Dewhurst**[1,2,3], **James P. Bagrow**[1,5], **Peter S. Dodds**[1,2,4], **Christopher M. Danforth**[1,2,5]*

1 Vermont Complex Systems Center, University of Vermont, Burlington, VT, United States of America, **2** Computational Story Lab, University of Vermont, Burlington, VT, United States of America, **3** Charles River Analytics, Cambridge, MA, United States of America, **4** Department of Computer Science, University of Vermont, Burlington, VT, United States of America, **5** Department of Mathematics & Statistics, University of Vermont, Burlington, VT, United States of America

* thayer.alshaabi@uvm.edu (TA); chris.danforth@uvm.edu (CMD)

Abstract

Human mortality is in part a function of multiple socioeconomic factors that differ both spatially and temporally. Adjusting for other covariates, the human lifespan is positively associated with household wealth. However, the extent to which mortality in a geographical region is a function of socioeconomic factors in both that region and its neighbors is unclear. There is also little information on the temporal components of this relationship. Using the districts of Hong Kong over multiple census years as a case study, we demonstrate that there are differences in how wealth indicator variables are associated with longevity in (a) areas that are affluent but neighbored by socially deprived districts versus (b) wealthy areas surrounded by similarly wealthy districts. We also show that the inclusion of spatially-distributed variables reduces uncertainty in mortality rate predictions in each census year when compared with a baseline model. Our results suggest that geographic mortality models should incorporate nonlocal information (e.g., spatial neighbors) to lower the variance of their mortality estimates, and point to a more in-depth analysis of sociospatial spillover effects on mortality rates.

1 Introduction

Although Hong Kong is a small island territory, it exhibits significant variation in occupations, income, foreign inhabitant density, and residence status of workers. In this study, we examine the benefits and drawbacks of incorporating nonlocal and spatial information into a mortality model for a limited area with restricted publicly available data. Simulating a realistic scenario with limited spatial resolution, we show heterogeneity of such exogenous factors and investigate nonlocal behavioral interactions of prosperity and deprivation across neighborhoods.

We present an analytical evaluation comparing local and nonlocal models to show the importance of spatial associations for mortality modeling. In particular, we apply a spatial network technique to examine socioeconomic nonlocality among communities. For instance, we investigate how the magnitude of a socially deprived area can consequently have a nonlocal effect on its neighbors' mortality risks. Similarly, we delve into how the spatial spread of

property of an affluent area can spillover to its surrounding areas, and thus affect their longevity. Our work not only reveals the deep influence of these spatial interactions of districts on predicting fatality rates, but also provides a method for investigating systematic inference errors of mortality models.

We structure our paper as follows. We discuss key findings of mortality risk studies in the literature and how they relate to our case examination in the next section. We introduce and analyze our data sources in Sec. 3.1. We summarize the economic and social indicators used in our investigation in S1 File. For our analytical inquiry, we employ a set of Bayesian generalized additive models to predict mortality rates across districts in Hong Kong. We describe our experiments and our exposition of the models in Sec. 3.2. First, we present our local model that does not use any spatial information in Sec. 3.2.1. We compare our Baseline design to two nonlocal spatial models in Sec. 3.3.2. Our first nonlocal model uses spatial features from the nearest neighbours, while the second uses features from all neighbours weighted by their distance to the target area. We will refer to the nonlocal models as SP, and WSP respectively. We show our findings in Sec. 4, highlighting the computational complexity of each method and discussing the benefits and shortcomings of each design. Our evaluation also reveals evidence of sociospatial spillovers of mortality rates. We conclude with some remarks on the limitations of our investigation and potential future work.

2 Related work

2.1 Mortality risks and social deprivation

There are many studies that delve into the temporal dynamics of mortality risks with respect to nation-wide epidemics [1, 2], pollution [3, 4], and life expectancy [5] over the last decade. Researchers have hypothesized and identified several connections of longevity, social deprivation, and socioeconomic discrimination [6–8]. Notably, there are many interpretations of social deprivation. Messer *et al.*'s study [9] offers a well-written overview of socioeconomic deprivation in the literature. The authors highlight the limitations of such definitions and propose an alternative method to calculate and standardize what they call a "neighborhood deprivation index" (NDI). Employing principal components analysis (PCA) on census data from 1995 to 2001, they illustrate the effectiveness of their proposed measurement at capturing socioeconomically deprived counties in the US.

Others have investigated a wide range of socioeconomic, psychological, and behavioral factors of fatality risks [10]. We often examine the notion of disparity in health and mortality risks using population-scale inputs and sensitive individual variables such as age, race, and gender respecting the privacy concerns that emerge from such applications [11]. Ou *et al.* [12] infer socioeconomic status by type of housing, education, and occupation. They find that regions with lower socioeconomic status have higher rates of air pollution. They also report that neighborhoods with higher densities of blue-collar workers have higher rates of air pollution-associated fatality than others. Chung *et al.* [13] present evidence of inequalities conditioned on age as a control variable. The authors investigate the impact of socioeconomic status amid the rapid economic development of Hong Kong. Their findings suggest a decline in socioeconomic disparity in mortality risks across the distrcits of Hong Kong from 1976 to 2010. They also show that various health benefits brought by economic growth are greater for regions with higher socioeconomic status. The market share of health benefits is unequally distributed among groups of varying status: Individuals with higher socioeconomic status have access to greater benefits than those of lower socioeconomic status. In the present study, we use a set of socioeconomic attributes including income, unemployment, and mobility, to define and capture the some of the ramifications of social deprivation in Hong Kong.

2.2 Spatial association of mortality risks

Spatial associations between income disparity and health risks are widely understood both internationally, and for individual cities and states [14–20]. Local attributes play a powerful role in the model dynamics, given the assumption that socioeconomic factors vary geographically. Studies have shown the importance of spatial associations in identifying relations between socioeconomic deprivation and longevity. Although researches have examined income inequality, they often use a spatially localized approach in their investigations [21–25].

Geographically weighted regression (GWR) is a commonly used method designed to examine spatial associations [26, 27]. Fotheringham *et al.*argue socioeconomic features are intrinsically intra-connected over space because of the mechanisms by which communities develop. Their study makes an empirical comparison of their proposed method (GWR) to other stationary regression models to investigate the spatial distribution of long-term illness in the UK.

Others have looked into the spatial association between air pollution and mortality in Hong Kong [28], Czechia [29], Rome [30], and France [31]. Cossman *et al.* [32] examine the spatial distribution of mortality rates over 35 years, starting from 1968 to 2002 across all counties in the US. The study highlights a nonrandom pattern of clustering in mortality rates in the US, where high fatality rates are primarily driven by economic decline.

To assess geospatial associations between pollution and mortality in Hong Kong, Thach *et al.* [33] examine the spatial interactions of tertiary planning units (TPUs) [34]—similar to census-blocks in the US. The authors show a positive spatial correlation between mortality rates and seasonal thermal changes in Hong Kong. They argue that the variation between TPUs is a key factor for cause-specific fatality rates. Their results show that socioeconomically deprived regions have higher fatality rates, especially during winter.

2.3 Sociospatial factors of death

Studying the relative spatial interactions of social and economic indicators dates back to decades ago. Researchers delve into measuring nonlocal and/or interdependent interactions of inequality in life expectancy [35], health care [36], education [37, 38], and decision-making [39, 40]. Many methods have been proposed to identify and examine broader dimensions of inequality from a spatial point of view such as Moran's I and spatial auto-regression [41–43]. Yang *et al.* [44] argue that mortality rates of counties in the US are associated with social and economical aspects found in neighboring counties. Their findings suggest that fatality rates in a county are remarkably driven by social signals from bordering counties because of the spillover of socioeconomic wealth or social deprivation across neighborhoods. Another recent work by Holtz *et al.* [45] highlights the significant influence of nonlocal interactions and spillovers on regional policies regarding the global outbreak of COVID–19. Employing a network-based approach to explore the dynamics of communities and their impact on mortality risks, we present here a small-case study using a collection of datasets from Hong Kong. In our study, we use three different models to illustrate the role of spatial associations by comparing models with spatial features to a baseline model without spatial factors.

3 Materials and methods

3.1 Data sources

Census data.   We collected socioeconomic variables curated by the Census and Statistics Department of Hong Kong [46]. We have three snapshots at 5-year intervals, 2006, 2011, and 2016. For each year, the dataset includes the total population density by district, median income, median rent to income ratio, median monthly household income, unemployment

rates, unemployment rates across households, unemployment rates among minorities, the proportion of homeless people, the proportion of homeless mobile residents, the proportion of single parents, the proportion of households with children in school, the proportion of households with children (aged under 15), and the proportion of households with elderly (aged 65 or above).

Mortality data.   We use the official counts of known and registered deaths provided by the Census and Statistics Department of Hong Kong [47]. The data set contains 892,055 death records between 1995 and 2017. Every record includes a wide range of information such as age, gender, and place of residence (TPU) [34]—a geospatial reference system used to report population census statistics. Our mortality records have both *place of occurrence* and *place of residence*. Each of those spatial features is certainly important to paint a better picture of the microscopic spatial associations of mortality and other social and economic factors. For our study, however, we only use *place of residence* as our primary geographic unit and discard all years other than the census years to cross reference our death records with other socioeconomic characteristics for each district. We derive an annual crude death rate of each district by simply dividing the number of registered death records by the population size for each calendar year such that:

$$Y_{it} = \frac{N_{it}(\text{deaths})}{N_{it}(\text{census population})},$$

for district i and $t \in [2006, 2011, 2016]$.

Life insurance data.   We have obtained data from a Hong Kong based life insurance provider. According to the Hong Kong Insurance Authority [48], our provider had roughly 2.5% market share of all non-linked individual life insurance policies issued in Hong Kong in 2016. We normalize the number of polices issued at the district level by population size for each time snapshot to report the proportion of individuals insured by each district. Notably, our variable is limited to policies sold by a single company and thus affected by the sociospatial features of the company market share such as the spatial sparsity of its agents and offices, and the social characteristics of consumers who would choose our provider over other life insurance providers in the area. However, statistical and detailed data sources regarding life insurance policies are often proprietary, especially with a similar spatial resolution to the one presented in our study. Although our data on life insurance policies may not represent the full population, it provides an example of the data that an insurance company can use to build their models. Given the scarcity of such data, we use our records of life insurance policies as a useful complementary wealth indicator—among other variables such as income, rent and unemployment—which is absent from most studies.

Geospatial unit.   Initially, we planned to use TPUs as the main geospatial units to cross-reference our data sources. However, we identified a large subset of missing TPUs in death records, as records in small TPUs may reveal sensitive information about specific individuals there. To avoid any risk of identifying individuals in the data set, we use districts as our main spatial unit of analysis [49]. This choice is consistent with prior work, where most studies have either filtered out small TPUs in their analyses [13, 28], or aggregated their records at the district level [33] to overcome this challenge.

Categorization.   We organize our features into three different categories.

1. **Base**: This set has most of the socioeconomic features in our data sets such as population density, unemployment rates, the proportion of homeless people, mobile residents, and single-parent households. However, we do not include wealth-, age-, or race-related features here.

2. **Wealth**: Besides the base features described above, we include median income, median rent-to-income ratio, median monthly household income, and life insurance coverage by the district.

3. **All**: This set includes all features in our data set, including sensitive variables—from a socio-political perspective—such as the proportion of minorities and unemployment rates among minorities. This set also includes age-related features such as the proportion of young and elderly residents for each district.

3.2 Statistical methods

There are many statistical modeling paradigms to tackle this task. Each of which comes with its own costs and benefits. Researchers sometimes use Poisson models to investigate mortality risks [1, 4, 28, 33]. Others use general linear or multivariate regression models [31, 35, 50]. We can also see modifications of this family of approaches in the literature such as geographically weighted regression [26, 27]. In this study, however, we consider the simplest approach. We use a set of three Bayesian multivariate linear regression models. Our goal is to examine the addition of nonlocal information, regardless of the distributional assumptions placed on the response variable Y. Therefore, we keep our model as simple as possible to allow us to investigate the implications of two different spatial models compared to a baseline model that does not factor in any nonlocal information. We treat the design tensors $\vec{X}$ as exogenous variables and do not model their evolution across time. A "design tensor" is a rank 3 tensor given by $\vec{X} = (\vec{X}_1, \vec{X}_2, \ldots, \vec{X}_T)$ where $\vec{X}_t$ is the design matrix for time period $t \in [2006, 2011, 2016]$. Each design matrix $\vec{X}_t$ is of dimension $N \times (p + 1)$, where N is the number of observations which accounts for 18 districts in Hong Kong, and p is the number of predictors. We add an extra variable to the design matrix to account for a constant in our linear model.

3.2.1 Local model (Baseline). The dynamics of the local models are described by a system of linear equations,

$$\vec{y}_t = \vec{X}_t \vec{\beta}_t + \sigma_t \vec{u}_t, \tag{1}$$

$$\vec{\beta}_t = \vec{\beta}_{t-1} + \vec{\mu} + \vec{L}\vec{v}_t, \tag{2}$$

$$\log \sigma_t = \log \sigma_{t-1} + \mu + \ell w_t, \tag{3}$$

for $t = 1, \ldots, T$.

Eq 1 is an ordinary linear model for the response vector $\vec{y}_t$ as a function of the design matrix $\vec{X}_t$ and coefficients $\vec{\beta}_t$. We presently define the quantities that compose Eqs 2 and 3. We set

$$\vec{u}_t, \vec{v}_t \sim \text{MultivariateNormal}(\vec{0}, \vec{I}) \tag{4}$$

in Eqs 1 and 2, while $w_t \sim \text{Normal}(0, 1)$. Our identity matrix $\vec{I}$ is informed by the number of predictors in our model, and has a dimension of $(p + 1) \times (p + 1)$. Hence the model likelihood is

$$\begin{aligned} p(\vec{y}|\vec{\beta}, \sigma) &= \prod_{t=1}^{T} \prod_{n=1}^{N} p(y_{tn}|\vec{X}_{tn}\vec{\beta}_t, \sigma_t) \\ &= \prod_{t=1}^{T} \prod_{n=1}^{N} \text{Normal}(\vec{X}_{tn}\vec{\beta}_t, \sigma_t^2). \end{aligned} \tag{5}$$
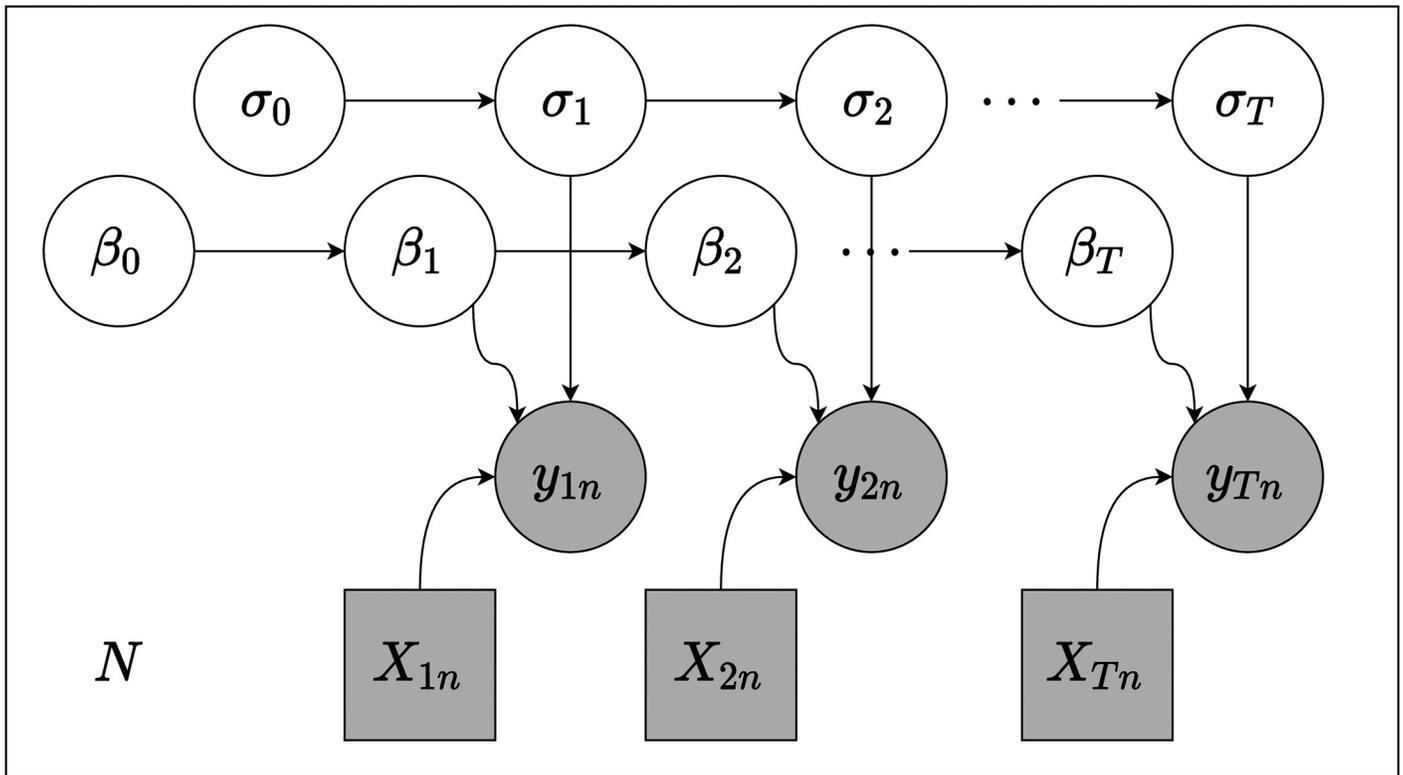
Fig 1. A graphical model representing the likelihood function given in Eq 5. Latent log σ_t and β_t evolve as biased random walks, while y_{tn} and $\vec{X}_{tn}$ are treated as observable random variables and exogenous parameters respectively. The entire temporal model is plated across districts $N = 18$.

A graphical model corresponding to Eq 5 is displayed in Fig 1. We also *a priori* believe that $\vec{\beta}_t$ does not remain constant throughout the time under study, though we are unsure of exactly how it changes over this time. Thus, we assume a prior on $\vec{\beta}_t$ that evolves as a biased random walk with drift given by $\vec{\mu}$ and correlation matrix $\vec{\Sigma}$ with Cholesky decomposition $\vec{L}$.

Likewise, we suppose that log σ_t evolves according to a univariate random walk with drift given by μ and standard deviation ℓ. We make this assumption for the same reason: We do not believe it is likely that σ_t remains constant over the entire time period of study. The random walk priors for $\vec{\beta}_t$ and σ_t are each centered about zero because we impose a zero mean prior on $\vec{\mu}$ and μ. We initialize these random walks with zero-centered multivariate normal initial conditions,

$$\vec{\beta}_0 \sim \text{MultivariateNormal}(\vec{0}, \vec{\Sigma}), \tag{6}$$

and

$$\log \sigma_0 \sim \text{Normal}(0, \ell^2). \tag{7}$$

The distribution of $\vec{\beta}_{1:T} \equiv (\vec{\beta}_1, ..., \vec{\beta}_T)$ is thus given by

$$p(\vec{\beta}_{1:T}|\vec{\mu}, \vec{\Sigma}) = \prod_{t=1}^{T} p(\vec{\beta}_t|\vec{\beta}_{t-1}, \vec{\mu}, \vec{\Sigma}) \tag{8}$$

$$= \prod_{t=1}^{T} \text{MultivariateNormal}(\vec{\beta}_{t-1} + \vec{\mu}, \vec{\Sigma}), \tag{9}$$

with an analogous (univariate) distribution holding for $\log \sigma$. We set

$$\vec{\mu} \quad \sim \text{MultivariateNormal}(\vec{0}, \vec{I}), \tag{10}$$

$$\mu \quad \sim \text{Normal}(0, 1), \tag{11}$$

so that the prior distribution over the paths of the regression coefficients and log standard deviation are centered about zero—the null hypothesis—for all t.

We place a uniform prior (LKJ(1)) over the correlation component of $\vec{\Sigma}$. The mean of this prior is at the identity matrix. The vector of standard deviations of $\vec{\Sigma}$, $\vec{s}$, is hypothesized to follow an isotropic multivariate log normal distribution as $\vec{s} \sim \text{LogNormal}(0, 1)$. We also place a univariate LogNormal(0, 1) prior on ℓ, the standard deviation of the increments of $\log \sigma_t$. We make this choice because we do not possess prior information about the appropriate noise scale of $\vec{\beta}_t$ or $\log \sigma_t$ and the log normal distribution is a weakly-informative prior that does not encode much prior information about their noise scales.

We did not perform exact inference of this model but rather fit parameters of a surrogate variational posterior distribution. We use here variational inference to approximate the posterior probability of the design tensor and ultimately run Bayesian inference over these features [51, 52]. Although traditional methods such as Markov chain Monte Carlo sampling (MCMC) can offer guarantees of accurate sampling from the target density, it does not necessarily outperform variational inference in terms of accuracy [53–55]. Evaluating the costs and benefits for accurate estimation of posteriors is indeed an open area of research, however, variational inference offers a much faster and effective method to approximate probability densities through optimization, even for small datasets [56, 57]. Using variational inference also allows us to have an agile development cycle and flexible models, as access to more economic and social data features will continue to evolve, and change over time. The effectiveness and versatility of variational inference has left a remarkable positive impact across disciplines [58–62].

Denoting the vector of all latent random variables by

$$\vec{z} \equiv (\mu^{(\sigma)}, \ell, \vec{\mu}^{(\vec{\beta})}, \Sigma^{(\vec{\beta})}, \log \sigma_{0:T}, \vec{\beta}_{0:T}), \tag{12}$$

we fit the parameters ψ of an approximate posterior distribution $q_\psi(\vec{z})$ to maximize the variational lower bound, defined as the expectation under $q_\psi(\vec{z})$ of the difference between the log joint probability and $\log q_\psi(\vec{z})$ [52]. We chose a low-rank multivariate normal guide with rank equal to approximately $\sqrt{\dim \vec{z}}$. This low rank approximation allows for modeling of correlations in the posterior distribution of $\vec{z}$ with a lower number of parameters than, for example, a full-rank multivariate normal guide distribution. All bounded latent random variables are reparameterized to lie in an unconstrained space so that we could approximate them with the multivariate normal guide.

3.2.2 Nonlocal models (SP and WSP). We use the road networks in Hong Kong to build a spatial network of the 18 districts [63, 64]. Each node in the network represents a single district. Nodes are linked if they share a direct road or bridge. In Fig 2A, we show a map of Hong Kong's districts. We display an undirected network of districts in Fig 2B. By contrast, we show a fully connected version of the network in Fig 2C. Edges are weighted by their spatial distance measured by the length of the shortest path d_{ij} to reach from district i to district j

$$w_{ij} = \exp\{-(d_{ij} - 1)\}, \tag{14}$$

and weights decays exponentially as the length of the shortest path increases between any two districts.
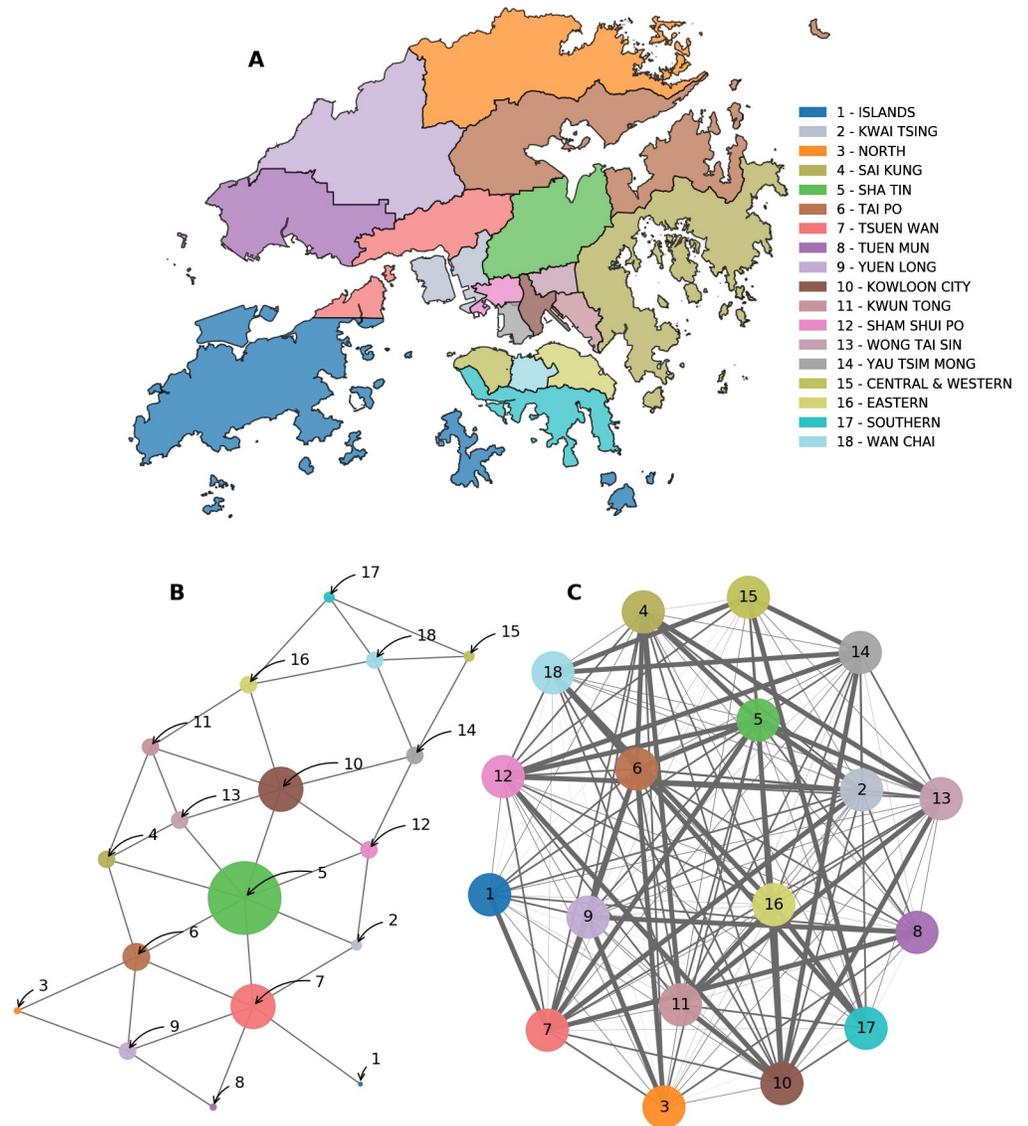
Fig 2. Spatial networks of Hong Kong's districts. (**A**) We cross-reference the roads and bridges connecting these districts to build a spatial network [63, 64]. (**B**) We demonstrate the first undirected network layout of Hong Kong's districts. Districts (nodes) are linked if they border each other or share a direct road/bridge in a binary fashion. (**C**) We show a fully connected version of the network. For the fully connected network, edges to neighboring districts are weighted exponentially. Different weighting schemes can be applied here, however, for our application we use the spatial distance measured by the length of the shortest path connecting any two districts on the network.

https://doi.org/10.1371/journal.pone.0247795.g002

Similarly, we treat the adjacency matrix $\vec{A}$ as an exogenous variable. We fit two nonlocal models that leverage the design matrix associated with each district's neighbors; a spatial model (SP) that uses the binary adjacency matrix and a weighted spatial model (WSP) that uses the weighted adjacency matrix. The equations describing the time evolution of this data generating process are

$$\vec{y}_t \quad = \vec{X}_t \vec{\beta}_t + \vec{f}(\vec{X}_t \otimes \vec{B})\vec{\gamma}_t + \sigma_t \vec{u}_t, \tag{14}$$

$$\vec{\beta}_t \quad = \vec{\beta}_{t-1} + \vec{\mu} + \vec{L}\vec{v}_t, \tag{15}$$

$$\vec{\gamma}_t \quad = \vec{\gamma}_{t-1} + \vec{m} + \vec{\Lambda}\vec{q}_t, \tag{16}$$

$$\log \sigma_t \quad = \log \sigma_{t-1} + \mu + \ell w_t, \tag{17}$$

for $t = 1, \ldots, T$. The rank three tensor $\vec{X}_t \otimes \vec{B}$ is the outer product of $\vec{B} \equiv \vec{A} - \vec{I}$ with the design matrix $\vec{X}_t$. The function $\vec{f}$ is a reduction function that lowers the rank of the tensor by one by collapsing the first dimension. Here we take $\vec{f}$ to be the mean across the first dimension. In other words, $\vec{f}(\vec{X}_t \otimes \vec{B})$ is a design matrix where $\vec{f}(\vec{X}_t \otimes \vec{B})_{ij}$ is the average of the values of predictor j over all the neighbors of district i in the network.

The prior distributions for $\vec{\gamma}_t$, $\vec{m}$, $\vec{\Lambda}$, and $\vec{q}_t$ are identical to those for $\vec{\beta}_t$, $\vec{\mu}$, $\vec{L}$, and $\vec{u}_t$ except their dimensionality is lowered from $p+1$ to p since we do not include an intercept term in $\vec{f}(\vec{X}_t \otimes \vec{B})$.

We use Pyro [65], a probabilistic programming language that operates on top of Pytorch [66], a dynamic graph differentiable programming library, to implement our models. Our source code along with our documentation is publicly available online on our Gitlab repository at: https://gitlab.com/compstorylab/asis.

4 Results and discussion

4.1 Observational and analytical findings

In Fig 3, we display the spatial distribution of socioeconomic characteristics for 2006, 2011, and 2016. Each heatmap is normalized by the mean and standard deviation for each year, such that darker shades of red show areas above the mean for each of these variables, while shades of grey show areas below the mean. We show normalized population density in Fig 3A through 3C. We see dense clusters both at the center of the country and on the northwestern side.

We are primarily interested in the geospatial trend across different variables/predictors for each year, respectively. For example, our heatmaps in Fig 3D–3F show that the southern islands have higher mortality rates than the average rates of Hong Kong. The southern islands had higher rates of new life insurance policies in 2006 (Fig 3G), followed by consistently lower rates than average when compared to the rest of the districts in Hong Kong for each year, respectively (see Fig 3H and 3I).

The northwestern territories have higher rates of unemployment compared to the southeastern side of Hong Kong, as we see in Fig 3J–3L. In Fig 3M–3O, we observe that the east and center districts have higher normalized median income when adjusted for inflation. We display additional statistics regarding households in S2 Fig in S1 File.

4.2 Evaluation and comparison of the models

Although we have similar and simple building blocks for our models, they do scale differently in terms of their computational costs. The total number of parameters in our Baseline model is equal to $pN+1$, where p is the number of predictors we use for each district. Besides the set of predictors for each target district, our spatial model SP uses spatial features from the nearest neighbours (ego-network) to that district. Thus, the total number of parameters used in the SP model is $\propto pN\bar{C}$ where $\bar{C}$ is the average clustering coefficient of the network. Our WSP model leverages features from all neighbours weighted by their distance to the target district. It has the largest number of parameters, which is proportional to pN^2. The relative difference in the
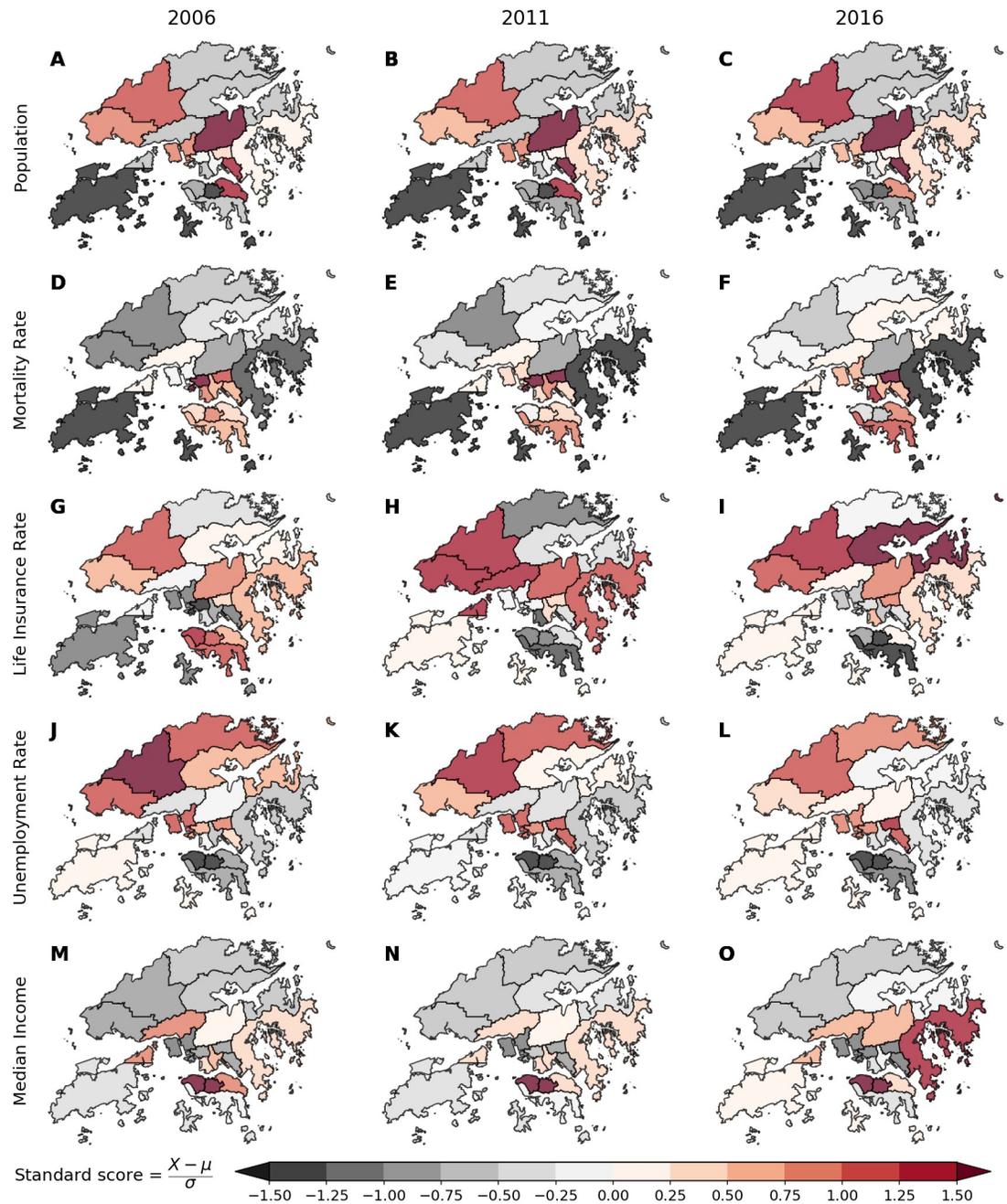
Fig 3. Temporal dynamics of spatial socioeconomic characteristics of Hong Kong. We show the spatial distribution of five features in our datasets for three different census years. Here, heatmaps are normalized by the mean and standard deviation. Darker shades of red show areas above the mean for each of these variables while shades of grey show areas below the mean. (**A–C**) We display the spatial growth of population over time. (**D–F**) We demonstrate the variation of mortality rates, and life insurance converge (**G–I**). We see some segregation of unemployment rates in (**J–L**), and median income in (**M–O**).

https://doi.org/10.1371/journal.pone.0247795.g003

number of parameters among these models urges us to further investigate the benefits of expanding the models with spatial features.

To evaluate our models, we consider two metrics: mean absolute error (MAE) to estimate our margin of error and mean signed deviation (MSD) to examine systematic bias. In Table 1,

Table 1. Model evaluation. We report the mean absolute error for each model across all districts averaged over a 1000 trials. The cells colored in blue show the model with the lowest margin of error for each feature category, whereas grey cells demonstrate ties among two models. The model highlighted in red indicates the model with the lowest margin of error.

Mean absolute error $e^t = \frac{1}{N}\sum_{i=1}^{18}|\hat{Y}_i^t - Y_i^t|$	Base (7 predictors)			Wealth (11 predictors)			All (16 predictors)		
	Baseline	SP	WSP	Baseline	SP	WSP	Baseline	SP	WSP
2006	1.19	1.25	1.17	1.25	1.22	1.25	1.32	1.40	1.46
2011	0.99	1.03	0.94	1.02	1.00	1.11	1.13	1.24	1.37
2016	0.90	0.92	0.86	251.09	251.09	1.26	1.20	1.34	1.51

we report the mean absolute error defined such that:

$$e^t = \frac{1}{N}\sum_{i=1}^{N=18}|\hat{Y}_i^t - Y_i^t|,$$

for each calendar year $t \in [2006, 2011, 2016]$ in our dataset across all districts. We highlight cells in blue to show the model with the lowest margin of error, and red to indicate the best model for all years. We also color cells in grey to demonstrate a tie between two models for a year.

For our default set of features (Base), we note our WSP model outperformed the rest of the models in most districts. However, as we add more predictors to the models, we observe a pattern whereby models with fewer parameters perform better. Our results show the SP model has the lowest MAE across districts when we use the Wealth category, which has 11 predictors including some wealth-related features as described in Sec. 3.1. The Baseline model—which does not account for spatial information—has a lower MAE when we feed all predictors to the models. This is an expected behavior because our larger and richer spatial models get overwhelmed with too many parameters and very limited data points. We show a detailed breakdown for each model and each district for the calendar year 2016 in Table 2.

We also compute a probability density function (PDF) of signed deviation $(\hat{Y}_i^t - Y_i^t)$ for each model, which is possible since our models are fully Bayesian and generate a distribution of possible outcomes. If the models accurately associate features with observed mortality rate, the distributions would be centered on zero. Conversely, if the models display systematic over- or under-estimation of mortality rate, the distributions will diverge away from zero, whereby negative numbers show underestimation and positive numbers indicate overestimation.

In Fig 4, we display the empirical distributions of signed deviation to examine the relative likelihood of systematic bias for models trained on the default set of features in 2016. We assess significance of model coefficients using centered $Q\%$ credible intervals (CI). A centered $Q\%$ credible interval of a probability density function $p(x)$ is an interval (a, b) defined such that $\frac{1}{2}\left(1 - \frac{Q}{100}\right) = \int_{-\infty}^{a} dx\, p(x) = \int_{b}^{\infty} dx\, p(x)$. We measure the significance of systematic errors in each model by computing the 80% CI, whereby systematic overestimation is highlighted in orange (CI > 0), and systematic underestimation is colored in blue (CI < 0).

We note that our spatial models, especially the weighted spatial model, are effective at reducing systematic over- and under-estimations. For example, the spatial models reduce the margin of error in panels B, H, L, and R of Fig 4. By contrast, all three models either overshoot or undershoot mortality rates drastically in a few districts (see panels D, E, M, and Q in Fig 4).

Our models also provide evidence to suggest that there are significant relationships between socioeconomic variables, such as household unemployment, percentage of single parents, and mortality rate. Many of these relations are significant in each of the census periods under study (2006, 2011, and 2016) while other relations are significant for some census periods but

Table 2. Model evaluation by district. We report the mean absolute error for each model in 2016 averaged over a 1000 trials. The cells colored in blue show the model with the lowest margin of error for each district and feature category, whereas grey cells demonstrate ties among two models for a district. The model highlighted in red indicates the model with the lowest margin of error across all districts, respectively.

Mean absolute error	Base (7 predictors)			Wealth (11 predictors)			All (16 predictors)		
$e_i^{2016} = \lvert \hat{Y}_i^{2016} - Y_i^{2016} \rvert$	Baseline	SP	WSP	Baseline	SP	WSP	Baseline	SP	WSP
ISLANDS	1.39	1.03	2.09	1.71	2.37	2.07	2.09	2.72	3.17
KWAI TSING	0.82	0.65	0.61	0.78	0.65	1.01	250.95	250.95	1.18
NORTH	0.59	0.84	0.54	1.42	1.04	1.39	1.93	1.72	1.90
SAI KUNG	251.41	251.41	1.78	1.52	1.99	1.63	1.59	1.82	1.89
SHA TIN	1.26	1.43	1.15	251.40	251.40	1.62	1.56	1.72	1.81
TAI PO	250.62	0.88	250.62	0.85	0.96	1.11	0.94	1.33	1.40
TSUEN WAN	0.36	0.57	0.49	250.55	250.55	0.67	0.63	0.78	0.77
TUEN MUN	0.65	0.71	0.60	1.47	1.16	1.72	1.49	1.18	1.45
YUEN LONG	250.63	0.81	250.63	0.54	0.68	0.80	0.69	1.00	0.99
KOWLOON CITY	0.41	0.67	0.54	0.51	0.65	0.74	0.68	0.94	0.98
KWUN TONG	0.61	0.77	0.83	0.96	0.84	1.09	1.15	1.47	1.46
SHAM SHUI PO	0.55	0.46	0.38	0.60	0.52	0.77	250.73	250.73	250.73
WONG TAI SIN	1.79	1.80	1.70	2.29	1.95	2.47	1.72	1.25	1.45
YAU TSIM MONG	0.79	0.51	0.71	1.37	1.19	1.22	0.81	0.96	0.92
CENTRAL & WESTERN	250.75	0.82	250.75	251.15	251.15	1.61	1.91	2.20	2.95
EASTERN	0.65	0.56	0.77	0.78	0.75	0.94	0.70	0.96	0.97
SOUTHERN	1.42	1.80	1.23	1.06	0.92	1.32	1.34	1.41	1.85
WAN CHAI	250.77	0.95	250.77	0.60	0.79	0.79	0.69	1.03	1.12

not others. We display the distributions of βs in S6 Fig in S1 File—the parameter for the baseline model. For each panel, we show the kernel density estimation of β as a function of each variable in the design tensor. We highlight distributions that are significantly above 0 using an orange color, while distributions significantly below 0 are colored in blue as measured by the 80% CI. Similar demonstrations of spatial and weighted spatial models can be found in S7 and S8 Figs in S1 File respectively. Besides the distributions of βs, we also show the kernel density estimation of γs—the hyperparameter used for the spatial competent in each model in S9 Fig in S1 File.

All three models are fairly accurate nonetheless. Access to a wider range of predictors and longitudinal data will help reduce our margin of error in estimating mortality rates. However, the spatial models allow us to capture nonlocal and interdependent interactions among the social and economic features across districts that would not be possible otherwise using the baseline model.

4.3 Evidence of sociospatial spillovers

The districts of Sai Kung, Sha Tin, Wong Tai Sin, and Southern are poorly fit by our models in 2016. In Fig 5, we inspect the characteristics of each district and its neighbors. The first three rows show how all three models overestimate mortality rates for the Sai Kung and Sha Tin districts (Fig 5A and 5B), while underestimating the other two districts (Fig 5C and 5D). For each district, we plot the standardized score of some socioeconomic features (black markers). We also display the corresponding average value of the same features from the neighboring districts derived from our network and denoted with orange markers to examine the effect of neighborhoods on these areas.
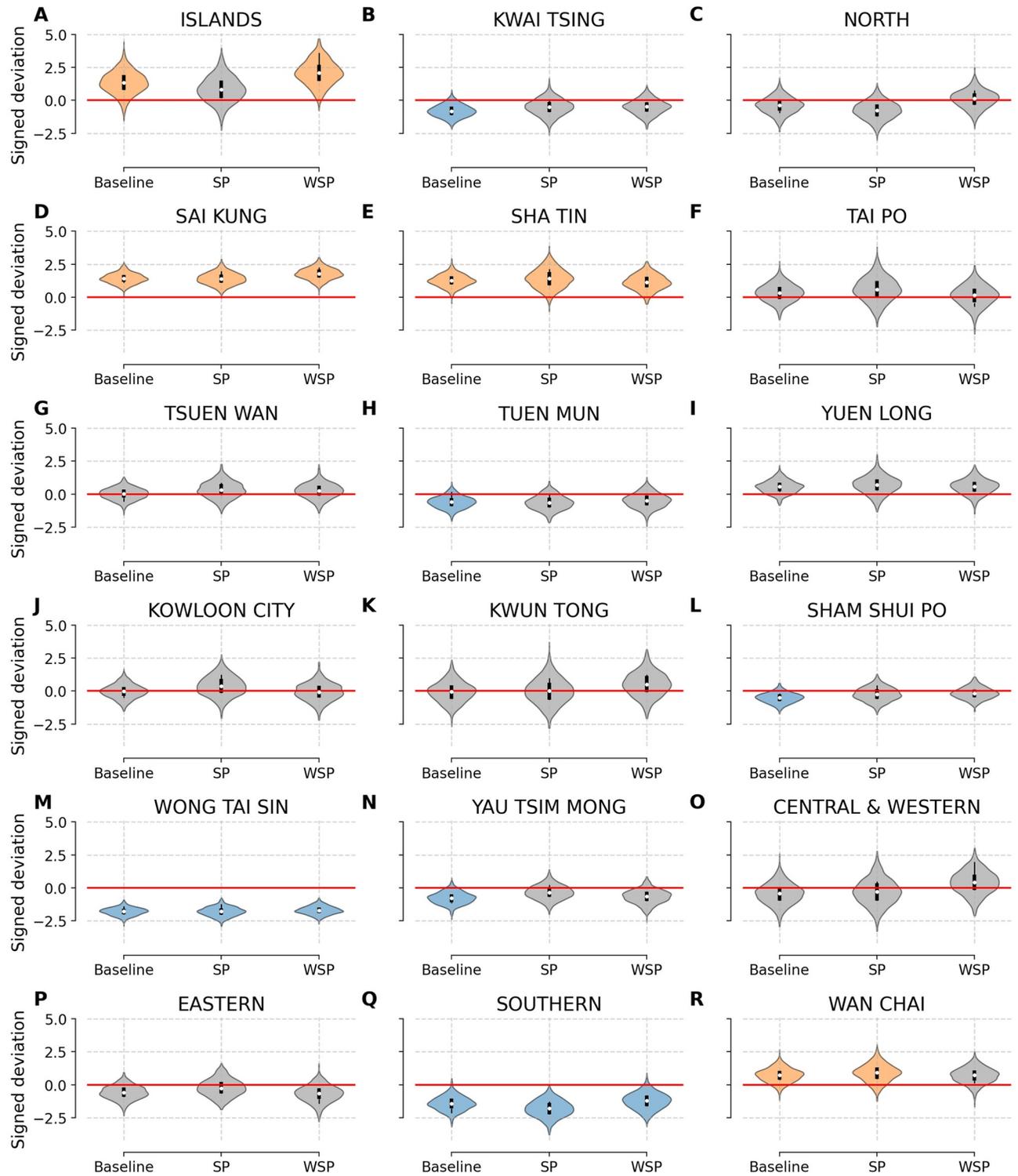
Fig 4. Relative likelihood of systematic bias for models trained on the default set of features in 2016. We examine the distribution of probable outcomes of signed deviation by computing the difference between our predictions $\hat{Y}_i^{2016}$ and the ground truth mortality rates Y_i^{2016} for each district. A perfect model would have a narrow distribution centred on 0 (solid red line going across). Positive values show overestimation, whereas negative values show an underestimation of mortality rates for each district. We color models with significance systematic overestimation in orange, while use blue to highlight models with significance systematic underestimation as measured by the 80% CI.
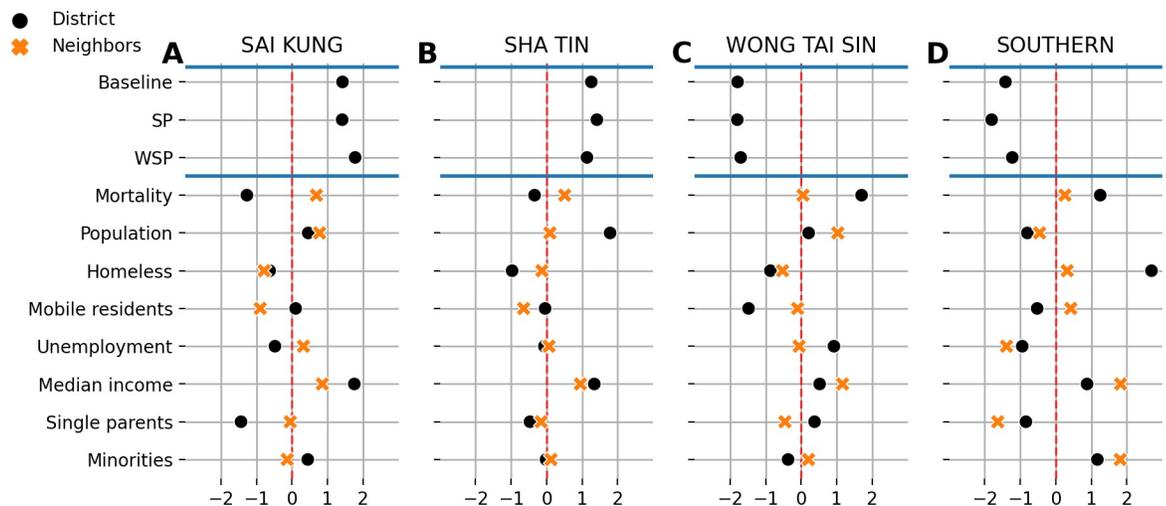
Fig 5. Impact of sociospatial factors on mortality risks. The first three rows show the mean signed deviation for four districts that are poorly fit by our models. (**A, B**) We show districts with systematic overestimation of mortality rates, while (**C, D**) show districts where mortality rates are systematically underestimated. For each district, we show the normalized value of some features of interest (black markers) along with the average value of the same features in the neighboring districts (orange markers). The red dashed line shows the average value for each of these normalized features centered at zero. We can see evidence of sociospatial factors of longevity in all four districts. Particularly, we note a spillover of wealth measured by median income. Districts in (**A**) and (**B**) maintain lower mortality rates while surrounded by districts with average mortality rates. Districts in (**C**) and (**D**) have a socioeconomic pull, driving the entire neighborhood to have higher mortality rates.

https://doi.org/10.1371/journal.pone.0247795.g005

The first two districts (Fig 5A and 5B) have lower mortality rates than average mortality of Hong Kong. They are significantly overestimated by our models, while mortality rates in their neighboring districts hover around average mortality rate of Hong Kong. The other two districts (Fig 5C and 5D) have higher mortality rates and surrounded by districts with lower mortality rates. The qualitative difference among these districts provides additional evidence of sociospatial and economic spillovers in mortality risks. We can see spillovers of wealth in Fig 5A and 5B with a higher median income in the district being associated with a higher median income in its neighborhood. Though districts in Fig 5C and 5D are located in a wealthy neighborhood, they still have a lower median income. The Wong Tai Sin district has a lower number of mobile residents than average (see Fig 5C). In Fig 5D, we see an extraordinary higher number of homeless people compared to the neighboring population. A higher percentage of minorities can also drive our systematic errors in this district, which hints at disparities in mortality risks in the area. We would need further analyses with finer geospatial resolution to explain this behavior.

We also analyze our signed deviation distributions through a local ego network approach. An ego network of a node is the network comprising that node and its nearest neighbors. Here each node is a district and its neighbors are that district's neighboring districts. The joint distribution of mortality at time t and district i and the model mortality rate prediction conditioned on location at node i can be concisely represented in the form of an ego network for each district. We display an example of this representation in Fig 6. We color nodes by standardized mortality rate and edges by standardized signed deviation error. Though this is an exploratory method that deserves greater expansion and attention in future work, we note qualitatively that the local view of predicted mortality versus true mortality varies substantially as a function of district. For example, the district of Wan Chai is connected to four other districts, three of which have substantially higher mortality than average and in particular higher mortality than Wan Chai itself, which has lower mortality than average. The model predictions
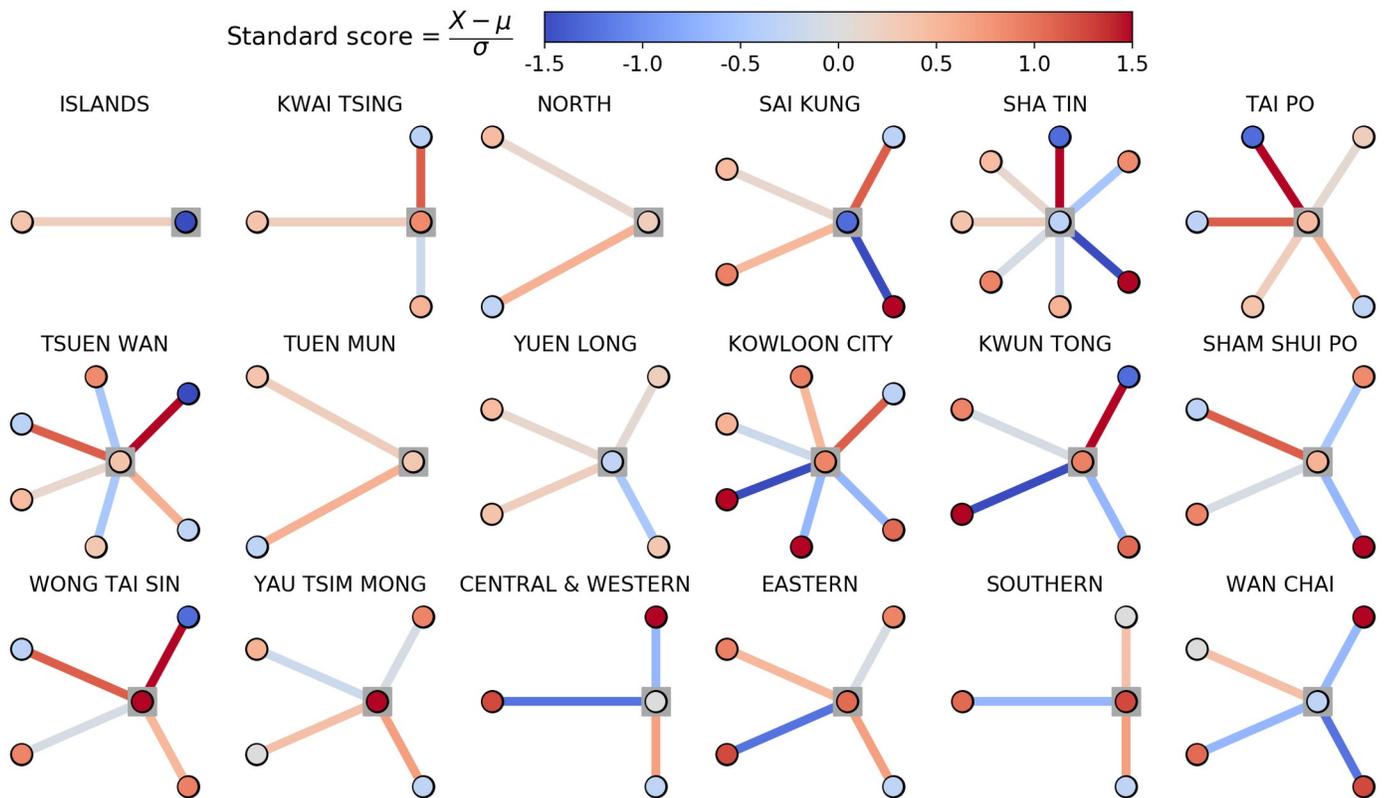
Fig 6. Ego networks of each district demonstrating sociospatial factors of mortality for the 2016 weighted spatial model. We display ego networks of each district in Hong Kong and its nearest neighbors in the road and bridge network. The central node (highlighted with a grey box) of each network corresponds to the labelled district. Neighbors are not arranged around the ego district geographically. Node color corresponds to normalized mortality rate and edge color corresponds to signed prediction error for the 2016 WSP model. These ego networks encode a qualitative measure of the sociospatial factors in mortality modeling. We display the equivalent networks for the Baseline and SP models in S10 and S11 Figs in S1 File respectively.

https://doi.org/10.1371/journal.pone.0247795.g006

for these neighbors are lower than their true mortality. An observer in Wan Chai who has access only to the model predictions and not the true mortality data would rationally assume that mortality in these districts is much lower than it truly is and could subsequently make further inferences or decisions based on faulty-but-rational assumption. Notably, socioeconomic diversity or heterogeneity of a neighborhood can change the local perception of mortality models across neighboring communities. We observe that high divergence from the neighborhood is associated with higher rates of uncertainty in the models. We envision future work incorporating this sort of information to fine tune mortality models.

5 Conclusion

Data-driven models are powerful tools often used to inform and reshape cultural, political, and financial policies around the globe. However, data scarcity and data sparsity pose an enormous challenge for some domains such as mortality modeling, especially for small territories. In this work, we studied the implications of that on the development of mortality models in Hong Kong with restricted access to publicly available data sources. We carried out a set of experiments to identify and explore how nonlocal and sociospatial interactions can systemically influence the outcome of a mortality model.

Our results support our hypothesis that spatial associations of wealth or social deprivation among neighborhoods have a direct and sometimes substantial impact on mortality risks. Our

examination reveals that localized models—which do not account for sociospatial factors—can systematically over- or under-estimate mortality rate—while spatial models reduce the error of predicting mortality rate. In our investigation, we show how our models scale differently regarding their complexity and statistical inference. We illustrate how the local perception of predicted mortality varies qualitatively and substantially as a function of the spatial unit. Future work can also improve upon our exploratory method to study spatial interdependency of social and economic factors of longevity, and identify sociospatial spillovers across neighbourhoods and communities.

We acknowledge our findings are limited for a few reasons. We only have access to census data for three individual years spanning a decade and a half. A better explanation of the nonlocalized effect of neighborhoods could be achieved by testing our hypotheses on additional years, along with more socioeconomic features to enrich our design tensor.

We used variational inference to estimate the posteriors analytically because of its effectiveness and versatility. Although our decision of using variational Bayes is substantive, future studies can further explore and examine the costs and benefits for accurate estimation of posteriors using variational inference compared with other classical Bayesian methods. Varying the model parameters temporally ensures a modular design, as access to richer and longitudinal sociotechnical data features continues to evolve. However, our evaluation suggests that using fixed parameters over time can reduce the number of tuneable parameters in the models substantially to overcome the challenge of high bias-variance tradeoff of the spatially rich and larger models.

Our geospatial resolution is unfortunately not high enough to identify some dynamics of connected communities. Our method, however, can be implemented similarly regardless of the spatial unit used for the experiment. Our spatial network is mainly based on the road network of Hong Kong, which could be extended to account for roads/bridges and public transport across any desired spatial unit.

Finally, we have only explored a distance-based weighting scheme for the connections across districts in the network. Population density could be included to enrich the socioeconomic effect of neighboring regions on a node within the network (for example, theory of intervening opportunities [67]). Other attributes such as geographic information associated with community health services could help us assess their value and reallocate these centers to more optimized locations.

Supporting information

S1 File.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: Thayer Alshaabi, David R. Dewhurst, Peter S. Dodds, Christopher M. Danforth.

Data curation: Thayer Alshaabi.

Formal analysis: Thayer Alshaabi, David R. Dewhurst.

Funding acquisition: Peter S. Dodds, Christopher M. Danforth.

Investigation: Thayer Alshaabi, James P. Bagrow, Peter S. Dodds, Christopher M. Danforth.

Methodology: Thayer Alshaabi, David R. Dewhurst.

Project administration: Peter S. Dodds, Christopher M. Danforth.

Software: Thayer Alshaabi, David R. Dewhurst.

Supervision: James P. Bagrow, Peter S. Dodds, Christopher M. Danforth.

Validation: Thayer Alshaabi, David R. Dewhurst.

Visualization: Thayer Alshaabi.

Writing – original draft: Thayer Alshaabi, David R. Dewhurst.

Writing – review & editing: Thayer Alshaabi, David R. Dewhurst, James P. Bagrow, Peter S. Dodds, Christopher M. Danforth.

References

1. Wong IO, Schooling C, Cowling BJ, Leung GM. Breast cancer incidence and mortality in a transitioning Chinese population: Current and future trends. British Journal of Cancer. 2015; 112(1):167–170. https://doi.org/10.1038/bjc.2014.532 PMID: 25290086

2. Wu P, Presanis AM, Bond HS, Lau EH, Fang VJ, Cowling BJ. A joint analysis of influenza-associated hospitalizations and mortality in Hong Kong, 1998–2013. Scientific Reports. 2017; 7(1):929. https://doi.org/10.1038/s41598-017-01021-x PMID: 28428558

3. Wong CM, Ma S, Hedley AJ, Lam TH. Effect of air pollution on daily mortality in Hong Kong. Environmental Health Perspectives. 2001; 109(4):335–340. https://doi.org/10.1289/ehp.01109335 PMID: 11335180

4. Qiu H, Tian L, Ho Kf, Pun VC, Wang X, Ignatius T. Air pollution and mortality: Effect modification by personal characteristics and specific cause of death in a case-only study. Environmental Pollution. 2015; 199:192–197. https://doi.org/10.1016/j.envpol.2015.02.002 PMID: 25679980

5. Lam TH, Ho SY, Hedley AJ, Mak KH, Leung GM. Leisure time physical activity and mortality in Hong Kong: Case-control study of all adult deaths in 1998. Annals of Epidemiology. 2004; 14(6):391–398. https://doi.org/10.1016/j.annepidem.2003.09.005 PMID: 15246327

6. Hayward MD, Pienta AM, McLaughlin DK. Inequality in men's mortality: The socioeconomic status gradient and geographic context. Journal of Health and Social Behavior. 1997; p. 313–330. https://doi.org/10.2307/2955428

7. Deaton A, Lubotsky D. Mortality, inequality and race in American cities and states. Social Science & Medicine. 2003; 56(6):1139–1153. https://doi.org/10.1016/S0277-9536(02)00115-6 PMID: 12600354

8. Rey G, Jougla E, Fouillet A, Hémon D. Ecological association between a deprivation index and mortality in France over the period 1997–2001: Variations with spatial scale, degree of urbanicity, age, gender and cause of death. BMC Public Health. 2009; 9(1):33. https://doi.org/10.1186/1471-2458-9-33 PMID: 19161613

9. Messer LC, Laraia BA, Kaufman JS, Eyster J, Holzman C, Culhane J, et al. The development of a standardized neighborhood deprivation index. Journal of Urban Health. 2006; 83(6):1041–1062. https://doi.org/10.1007/s11524-006-9094-x PMID: 17031568

10. Puterman E, Weiss J, Hives BA, Gemmill A, Karasek D, Mendes WB, et al. Predicting mortality from 57 economic, behavioral, social, and psychological factors. Proceedings of the National Academy of Sciences. 2020.

11. Santos-Lozada AR, Howard JT, Verdery AM. How differential privacy will affect our understanding of health disparities in the United States. Proceedings of the National Academy of Sciences. 2020; 117 (24):13405–13412. https://doi.org/10.1073/pnas.2003714117 PMID: 32467167

12. Ou CQ, Hedley AJ, Chung RY, Thach TQ, Chau YK, Chan KP, et al. Socioeconomic disparities in air pollution-associated mortality. Environmental Research. 2008; 107(2):237–244. https://doi.org/10.1016/j.envres.2008.02.002 PMID: 18396271

13. Chung RY, Lai FT, Chung GK, Yip BH, Wong SY, Yeoh EK. Socioeconomic disparity in mortality risks widened across generations during rapid economic development in Hong Kong: An age-period-cohort analysis from 1976 to 2010. Annals of Epidemiology. 2018; 28(11):743–752. https://doi.org/10.1016/j.annepidem.2018.06.004 PMID: 30392585

14. Buckingham K, Freeman PR. Sociodemographic and morbidity indicators of need in relation to the use of community health services: Observational study. BMJ. 1997; 315(7114):994–996. https://doi.org/10.1136/bmj.315.7114.994 PMID: 9365299

15. Kim D, Kawachi I, Vander Hoorn S, Ezzati M. Is inequality at the heart of it? Cross-country associations of income inequality with cardiovascular diseases and risk factors. Social Science & Medicine. 2008; 66(8):1719–1732. https://doi.org/10.1016/j.socscimed.2007.12.030 PMID: 18280021

16. Bjornstrom EE. An examination of the relationship between neighborhood income inequality, social resources, and obesity in Los Angeles county. American Journal of Health Promotion. 2011; 26(2):109–115. https://doi.org/10.4278/ajhp.100326-QUAN-93 PMID: 22040392

17. Chen Z, Crawford CAG. The role of geographic scale in testing the income inequality hypothesis as an explanation of health disparities. Social Science & Medicine. 2012; 75(6):1022–1031. https://doi.org/10.1016/j.socscimed.2012.04.032 PMID: 22694992

18. Fan JX, Wen M, Kowaleski-Jones L. Tract-and county-level income inequality and individual risk of obesity in the United States. Social Science Research. 2016; 55:75–82. https://doi.org/10.1016/j.ssresearch.2015.09.008 PMID: 26680289

19. Kim D, Wang F, Arcan C. Peer Reviewed: Geographic Association Between Income Inequality and Obesity Among Adults in New York State. Preventing Chronic Disease. 2018; 15. https://doi.org/10.5888/pcd15.180217

20. Yang TC, Matthews SA, Sun F, Armendariz M. Modeling the Importance of Within-and Between-County Effects in an Ecological Study of the Association Between Social Capital and Mental Distress. Preventing Chronic Disease. 2019; 16:E75–E75. https://doi.org/10.5888/pcd16.180491 PMID: 31198163

21. Kawachi I, Kennedy BP. The relationship of income inequality to mortality: Does the choice of indicator matter? Social Science & Medicine. 1997; 45(7):1121–1127. https://doi.org/10.1016/S0277-9536(97)00044-0 PMID: 9257403

22. Ross NA, Wolfson MC, Dunn JR, Berthelot JM, Kaplan GA, Lynch JW. Relation between income inequality and mortality in Canada and in the United States: Cross sectional assessment using census data and vital statistics. BMJ. 2000; 320(7239):898–902. https://doi.org/10.1136/bmj.320.7239.898 PMID: 10741994

23. Major JM, Doubeni CA, Freedman ND, Park Y, Lian M, Hollenbeck AR, et al. Neighborhood socioeconomic deprivation and mortality: NIH-AARP diet and health study. PLOS ONE. 2010; 5(11). https://doi.org/10.1371/journal.pone.0015538 PMID: 21124858

24. Reardon SF, Bischoff K. Income inequality and income segregation. American Journal of Sociology. 2011; 116(4):1092–1153. https://doi.org/10.1086/657114 PMID: 21648248

25. Yang TC, Jensen L. Exploring the inequality-mortality relationship in the US with Bayesian spatial modeling. Population Research and Policy Review. 2015; 34(3):437–460. https://doi.org/10.1007/s11113-014-9350-9 PMID: 26166920

26. Fotheringham AS, Charlton ME, Brunsdon C. Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. Environment and Planning A. 1998; 30(11):1905–1927. https://doi.org/10.1068/a301905

27. Fotheringham AS, Brunsdon C, Charlton M. Geographically weighted regression: The analysis of spatially varying relationships. John Wiley & Sons; 2003.

28. Wong CM, Ou CQ, Chan KP, Chau YK, Thach TQ, Yang L, et al. The effects of air pollution on mortality in socially deprived urban areas in Hong Kong, China. Environmental Health Perspectives. 2008; 116(9):1189–1194. https://doi.org/10.1289/ehp.10850 PMID: 18795162

29. Branis M, Linhartova M. Association between unemployment, income, education level, population size and air pollution in Czech cities: Evidence for environmental inequality? A pilot national scale analysis. Health & Place. 2012; 18(5):1110–1114. https://doi.org/10.1016/j.healthplace.2012.04.011 PMID: 22632903

30. Forastiere F, Stafoggia M, Tasco C, Picciotto S, Agabiti N, Cesaroni G, et al. Socioeconomic status, particulate air pollution, and daily mortality: Differential exposure or differential susceptibility. American Journal of Industrial Medicine. 2007; 50(3):208–216. https://doi.org/10.1002/ajim.20368 PMID: 16847936

31. Padilla CM, Kihal-Talantikite W, Vieira VM, Rossello P, Le Nir G, Zmirou-Navier D, et al. Air quality and social deprivation in four French metropolitan areas—a localized spatio-temporal environmental inequality analysis. Environmental Research. 2014; 134:315–324. https://doi.org/10.1016/j.envres.2014.07.017 PMID: 25199972

32. Cossman JS, Cossman RE, James WL, Campbell CR, Blanchard TC, Cosby AG. Persistent clusters of mortality in the United States. American Journal of Public Health. 2007; 97(12):2148–2150. https://doi.org/10.2105/AJPH.2006.093112 PMID: 17538052

33. Thach TQ, Zheng Q, Lai PC, Wong PPY, Chau PYK, Jahn HJ, et al. Assessing spatial associations between thermal stress and mortality in Hong Kong: A small-area ecological study. Science of the Total Environment. 2015; 502:666–672. https://doi.org/10.1016/j.scitotenv.2014.09.057

34. Tertiary Planning Units; 2016. https://www.bycensus2016.gov.hk/en/bc-dp-tpu.html.

35. Boing AF, Boing AC, Cordes J, Kim R, Subramanian SV. Quantifying and explaining variation in life expectancy at census tract, county, and state levels in the United States. Proceedings of the National Academy of Sciences. 2020; 117(30):17688–17694. https://doi.org/10.1073/pnas.2003719117 PMID: 32661145

36. Erreygers G, Van Ourti T. Measuring socioeconomic inequality in health, health care and health financing by means of rank-dependent indices: A recipe for good practice. Journal of Health Economics. 2011; 30(4):685–694. https://doi.org/10.1016/j.jhealeco.2011.04.004 PMID: 21683462

37. Zimmer RW, Toma EF. Peer effects in private and public schools across countries. Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management. 2000; 19(1):75–92. https://doi.org/10.1002/(SICI)1520-6688(200024)19:1%3C75::AID-PAM5%3E3.0.CO;2-W

38. Sacerdote B. Peer effects in education: How might they work, how big are they and how much do we know thus far? In: Handbook of the Economics of Education. vol. 3. Elsevier; 2011. p. 249–277.

39. Bodine-Baron E, Nowak S, Varadavas R, Sood N. Conforming and non-conforming peer effects in vaccination decisions. National Bureau of Economic Research; 2013.

40. Albert D, Chein J, Steinberg L. The teenage brain: Peer influences on adolescent decision making. Current Directions in Psychological Science. 2013; 22(2):114–120. https://doi.org/10.1177/0963721412471347

41. Moran PA. Notes on continuous stochastic phenomena. Biometrika. 1950; 37(1/2):17–23. https://doi.org/10.1093/biomet/37.1-2.17 PMID: 15420245

42. Li H, Calder CA, Cressie N. Beyond Moran's I: Testing for spatial dependence based on the spatial autoregressive model. Geographical Analysis. 2007; 39(4):357–375. https://doi.org/10.1111/j.1538-4632.2007.00708.x

43. Yang TC, Jensen L, Haran M. Social capital and human mortality: Explaining the rural paradox with county-level mortality data. Rural Sociology. 2011; 76(3):347–374. https://doi.org/10.1111/j.1549-0831.2011.00055.x PMID: 25392565

44. Yang TC, Noah AJ, Shoff C. Exploring geographic variation in US mortality rates using a spatial Durbin approach. Population, Space and Place. 2015; 21(1):18–37. https://doi.org/10.1002/psp.1809 PMID: 25642156

45. Holtz D, Zhao M, Benzell SG, Cao CY, Rahimian MA, Yang J, et al. Interdependence and the cost of uncoordinated responses to COVID-19. Proceedings of the National Academy of Sciences. 2020.

46. Population Census Data for Hong Kong; 2016. https://www.bycensus2016.gov.hk/en/index.html.

47. Micro-data set of known and registered deaths in Hong Kong; 2017. https://www.censtatd.gov.hk/service_desk/list/microdata/index.jsp.

48. Annual Long Term Business Statistics; 2016. https://www.ia.org.hk/en/infocenter/statistics/annual_long_term_business_statistics.html.

49. District and Constituency Area; 2016. https://www.bycensus2016.gov.hk/en/bc-dp.html.

50. Kiffer CR, Camargo EC, Shimakura SE, Ribeiro PJ, Bailey TC, Pignatari AC, et al. A spatial approach for the epidemiology of antibiotic use and resistance in community-based studies: The emergence of urban clusters of Escherichia coli quinolone resistance in Sao Paulo, Brasil. International Journal of Health Geographics. 2011; 10(1):1–10. https://doi.org/10.1186/1476-072X-10-17 PMID: 21356088

51. Beal MJ. Variational algorithms for approximate Bayesian inference. UCL (University College London). 2003.

52. Hoffman MD, Blei DM, Wang C, Paisley J. Stochastic variational inference. The Journal of Machine Learning Research. 2013; 14(1):1303–1347.

53. Braun M, McAuliffe J. Variational inference for large-scale models of discrete choice. Journal of the American Statistical Association. 2010; 105(489):324–335. https://doi.org/10.1198/jasa.2009.tm08030

54. Ranganath R, Gerrish S, Blei DM. Black Box Variational Inference. In: Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics; 2014.

55. Kucukelbir A, Tran D, Ranganath R, Gelman A, Blei DM. Automatic differentiation variational inference. The Journal of Machine Learning Research. 2017; 18(1):430–474.

56. Kucukelbir A, Ranganath R, Gelman A, Blei D. Automatic variational inference in Stan. Advances in Neural Information Processing Systems. 2015; 28:568–576.

57. Blei DM, Kucukelbir A, McAuliffe JD. Variational inference: A review for statisticians. Journal of the American statistical Association. 2017; 112(518):859–877. https://doi.org/10.1080/01621459.2017.1285773

58. Ghahramani Z, Beal MJ. Propagation algorithms for variational Bayesian learning. In: Advances in Neural Information Processing Systems; 2001. p. 507–513.

59. Rossi PE, Allenby GM, McCulloch R. Bayesian statistics and marketing. John Wiley & Sons; 2012.

60. Barber D. Bayesian reasoning and machine learning. Cambridge University Press; 2012.

61. Zhang C, Bütepage J, Kjellström H, Mandt S. Advances in variational inference. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2018; 41(8):2008–2026. https://doi.org/10.1109/TPAMI.2018.2889774 PMID: 30596568

62. Bottou L, Curtis FE, Nocedal J. Optimization methods for large-scale machine learning. Siam Review. 2018; 60(2):223–311. https://doi.org/10.1137/16M1080173

63. Zhou Q, Li Z. Empirical determination of geometric parameters for selective omission in a road network. International Journal of Geographical Information Science. 2016; 30(2):263–299. https://doi.org/10.1080/13658816.2015.1085538

64. Chen BY, Lam WH, Sumalee A, Li Q, Li ZC. Vulnerability analysis for large-scale and congested road networks with demand uncertainty. Transportation Research Part A: Policy and Practice. 2012; 46 (3):501–516.

65. Bingham E, Chen JP, Jankowiak M, Obermeyer F, Pradhan N, Karaletsos T, et al. Pyro: Deep Universal Probabilistic Programming. Journal of Machine Learning Research. 2018.

66. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In: Wallach H, Larochelle H, Beygelzimer A, d Alché-Buc F, Fox E, Garnett R, editors. Advances in Neural Information Processing Systems 32. Curran Associates, Inc.; 2019. p. 8024–8035. Available from: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

67. Stouffer SA. Intervening opportunities: A theory relating mobility and distance. American Sociological Review. 1940; 5(6):845–867. https://doi.org/10.2307/2084520