Published in final edited form as:

Health Place. 2012 July ; 18(4): 824–831. doi:10.1016/j.healthplace.2012.03.010.

Area variations in health: A spatial multilevel modeling approach[☆]

Mariana Arcaya^a, Mark Brewster^c, Corwin M. Zigler^b, and S.V. Subramanian^{a,*}

Mariana Arcaya: marcaya@hsph.harvard.edu; Mark Brewster: markbrewster@post.harvard.edu; Corwin M. Zigler: cmzigler@gmail.com; S.V. Subramanian: svsubram@hsph.harvard.edu

^aDepartment of Society, Human Development and Health, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA

^bDepartment of Biostatistics, Harvard School of Public Health, USA

^cDepartment of Epidemiology, Columbia University, USA

Abstract

Both space and membership in geographically-embedded administrative units can produce variations in health, resulting in geographic clusters of good and poor health. Despite important differences between these two types of dependence, one is easily mistaken for the other, and the possibility that both are at work is commonly ignored. We fit a series of hierarchical and spatially-explicit multilevel models to a U.S. county-level life expectancy in 1999 to demonstrate approaches for data analysis and interpretation when multiple sources of area-clustering are present. We demonstrate the methods to detect, interpret, and differentiate evidence of spatial and geographic membership effects and discuss key considerations for analyzing data with spatial or/and membership dimensions. We find evidence that life expectancy is driven by both within-state geographic process, and by spatial processes. We argue that considering spatial and membership processes simultaneously yields valuable insights into the patterning of area variations in health.

Keywords

Life expectancy; Spatial analysis; Multilevel modeling; Geographic variability; U.S. counties; States

1. Introduction

Since the 1990s, researchers have paid increasing attention to the role of geographic setting in shaping health (Kearns, 1993; Macintyre and Ellaway, 2000; Jones and Moon, 1993).

This trend is apparent in two main streams of literature that reflect longstanding theoretical distinctions between the concepts of “place” and “space” (Kearns and Joseph, 1993). First,

[☆]*Support*: No direct funding was available for this study. S V Subramanian is supported by the National Institutes of Health Career Development Award (NHLBI K25 HL081275) and the Robert Wood Johnson Investigator Award in Healthy Policy Research.

© 2012 Elsevier Ltd. All rights reserved.

^{*}Corresponding author. Tel.: +1617 432 6299; fax: +1617 432 3123.

Contributions: SVS and MA conceptualized and designed the study. MA led the analysis, interpretation of the data, and wrote the manuscript. MB contributed to the design, data analysis, interpretation and writing of the manuscript. CMZ contributed to data analysis, interpretation and writing of the manuscript. SVS contributed to the interpretation of the data and writing of the manuscript, and provided overall supervision for the study. All authors reviewed and approved the final version of the manuscript submitted for publication. All the authors confirm no conflict of interest.

social epidemiologists have argued for the need to consider “place” as a source of area-level health exposures such as welfare policy (Gnanasekaran et al., 2008), social environment quality (Jia et al., 2009), and neighborhood resources (Diez Roux and Mair, 2010). This focus on the substantive role of place in health has been spurred, in part, by the emergence of multilevel models (Raudenbush and Bryk, 2002), seeking to understand geographic clustering in health at the area level (Subramanian, 2004; Subramanian et al., 2003). Traditionally, this approach conceptualizes place as a matter of group membership whereby subjects (level-1) are nested in shared, geographical units (level-2), often defined on the basis of administrative boundaries. Geographic information is used to establish a nested hierarchy that describes shared contexts among observations.

In the second stream, medical geographers and spatial epidemiologists have demonstrated that “space” matters for health. This is particularly true in the fields of infectious disease epidemiology, surveillance, and environmental health, where proximity structures contagious processes, and measures of nearness can be used to detect sources of health-damaging environmental exposures (Elliott and Wartenberg, 2004; Gatrell and Foundation, 1998; Waller and Gotway, 2004). The proliferation of user friendly spatial statistics programs and increased computing power in recent years have allowed researchers to take advantage of computationally intensive spatial econometric (Fischer and Getis, 2010) and Bayesian inference (Lawson, 2009) methods to understand how health is patterned in space and why. In contrast to a “place” perspective, which uses geographic information to form groups, the “space” perspective uniquely defines each observation according to its proximity to all other observations, ignoring the potentially meaningful commonalities among observations that are generated by shared geographic or political boundaries. The merits of the two perspectives are rarely compared to identify the better approach for a specific dataset or problem, and even less often are they integrated, though there are notable exceptions in the literature (e.g., (Chaix et al., 2005; Chaix et al., 2005)).

In addition, the roles of place and space in understanding area variations in health (Mobley et al., 2010; Lebel et al., 2010; Moore and Carpenter, 1999) may also be addressed simply from a nuisance perspective, motivated by the need to satisfy the assumptions of parametric statistical methods (Holsten, 2009; Freisthler, 2004). Though often treated separately, integrating concepts of space and place while considering both nuisance and substantive perspectives may yield valuable insights. Spatial clustering that persists after accounting for place-based membership in geographically-embedded administrative units could suggest that spatially-explicit processes should be explored. Similarly, persistent membership effects after accounting for space may imply that policy factors are important.

We present a series of statistical models that offer empirical insights into how health clusters in both space and in geographically-defined administrative units. We use a life expectancy dataset to demonstrate approaches for data analysis and interpretation when multiple sources of clustering are present. Because multilevel modeling methods are traditionally applied to datasets that feature individual-level data, it is important to note that our lowest-level unit of analysis is the county. While this precludes individual-level causal inference, which is often the primary focus of epidemiologic research, the goal of this paper is to introduce a generic method for analyzing data that are structured by both spatial and membership processes. The overview is organized as follows: we first demonstrate the methods to detect, interpret, and differentiate evidence of spatial and geographic membership effects. We then discuss key considerations for epidemiologists analyzing data with spatial or/and geographic membership dimensions.

1.1. Aims

This paper aims to clarify differences and similarities between two broad methodological approaches to location that have separately dominated the literature around health geography. We review how clustering, also referred to as dependence or autocorrelation, is detected and interpreted from the multi-level and spatial perspectives in this life expectancy dataset for illustration. The main purposes of this exercise are to: (1) highlight that multiple forms of dependence may exist in geographically-referenced, hierarchical data, and that such clustering must be considered; (2) propose one method of integrating place and space perspectives when such clustering is present; and (3) to highlight the substantive insights that might be gained by examining clustering, which is often treated as a nuisance, through these place and space lenses. Finally, the paper discusses extensions to the simple framework presented with the goal of elucidating ways to tailor this multilevel spatial approach to other datasets and research questions.

2. Methods

2.1. Data

We base our analysis on county-level life expectancy estimates for 1961–1999 compiled by Ezzati and colleagues (Ezzati et al., 2008), publicly available at PLoS Medicine. This dataset contains 2068 geographic units comprising the 3141 United States (US) counties, with smaller counties combined to form geographies with a total population of at least 10,000 males and 10,000 females in 1990 and adjacent counties merged when needed to smooth changes in county boundaries over the study time period. Ezzati and colleagues estimated mortality rates using vital statistics from the National Center for Health Statistics (NCHS) and county population by age from the US Census Bureau. They used standard life table techniques to estimate life expectancy (Ezzati et al., 2008). For brevity, we refer to these geographic units as “counties” throughout this paper. For this analysis, we used 1999 life expectancy estimates for the 2063 counties that fall within the continental US. The District of Columbia and 48 states (i.e., 49 “states”) comprise these counties.

2.2. Models

We fit four null models to the data to estimate average yearly life expectancy in the US, as well as to detect groups of counties with particularly high and low life expectancies in 1999. We omitted known predictors of life expectancy, both to maintain a focus on model characteristics rather than the drivers of mortality during this time period and, more importantly, to maintain unexplained variability in life expectancy estimates for the purposes of partitioning it into geographic membership (place) and spatial components.

2.2.1. Model 1: Basic single-level model—We first fit a single-level null model to the data:

$$LE_i = \beta + (e_i), \quad e_i \sim N(0, \sigma_e^2)$$

where LE_i is the life expectancy of county i ($i = 1, \dots, 2063$) in 1999, β is the average life expectancy across all counties, and e_i is a county-specific residual capturing the difference between the actual life expectancy in that county and the overall mean. This specification was a simple Ordinary Least Squares (OLS) model that assumed independence among observations, ignoring dependence among observations created by shared “place” and proximity in space.

2.2.2. Model 2: Two-level model accounting for hierarchy—We extended this to a two-level hierarchical model, conceptualizing our data in a multilevel framework, and fitting a random effects model where counties were nested within states and each state had its own unique average life expectancy. Our basic random effects model was specified:

$$LE_{ij} = \beta + (u_j + e_{ij}), \quad u_j \sim N(0, \sigma_u^2), \quad e_{ij} \sim N(0, \sigma_e^2)$$

where LE_{ij} is the average life expectancy of county i ($i = 1, \dots, 2063$) within the j th contiguous state ($j = 1, \dots, 49$).

β still represents average life expectancy for all counties and bracketed terms represent random effects associated with state and county. The term u_j is a state-specific residual that gives each state its own average life expectancy, and e_{ij} is a residual for each county. Assuming a normal distribution with a 0 mean, the model estimated a variance at level-1 (: between-counties, within states) and level-2 (: between-states) in life expectancy. These variances are described by two parameters, σ_u^2 and σ_e^2 , respectively, where $u_j \sim N(0, \sigma_u^2)$ and $e_{ij} \sim N(0, \sigma_e^2)$, such that state residuals are identically and independently distributed (iid) and county residuals are iid within states. This multilevel model is also often referred to as a mixed model, a random intercepts model, a hierarchical model, or a variance components model (Singer, 1998). The multilevel modeling approach employed provided the advantages of decomposing the variation in life expectancy attributable to states and also precision-weighting estimates for state-specific predictions (Goldstein, 2010).

The state-specific average life expectancy from Model 2 ($\beta + u_j$), which we will refer to as β_j , is a weighted average of the state-specific averages that would be estimated if each state's mean life expectancy were calculated separately (which we will refer to as β_j^*), and the overall multilevel mean estimated in Model 2 (β), as follows:

$$\beta_j = w_j \beta_j^* + (1 - w_j) \beta$$

where the overall multilevel average from Model 2 (β) is the weighted average of all state-specific means where state means are calculated independently of one another:

$$\beta = \left(\sum w_j \beta_j^* \right) / \sum w_j$$

The weights, w_j , indicate the precision of the state-specific means, and are calculated by taking the ratio of the between-state variance to the total variance. The total variance includes the within-state sampling variance attributable to observing a sample of counties within states. As a result, states with fewer counties and consequently higher sampling variances are assigned a lower weight and are therefore pulled more towards the overall mean intercept, as given by

$$w_j = \frac{\sigma_u^2}{(\sigma_e^2/n_j) + \sigma_u^2}$$

where n_j is the number of counties within each state and σ_e^2 is the total between-county variance. In more traditional applications of multilevel modeling techniques where individuals, rather than geographies, serve as the lowest-level unit of analysis, the number of

individuals contributing data to each higher level unit would determine the precision of the level-2 estimates.

Together, σ_e^2/n_j gives the random sampling variance of each parameter estimated in a separate estimation (i.e., fixed effects approach) to account for state membership.

In contrast to Model 1, this specification no longer demands complete independence among observations. Rather, it allows counties from the same states to be more similar to each other than they are to counties from other states, reflecting the fact that state-specific policies, economic conditions, and other state-level factors may influence life expectancy and create clusters of longevity or premature mortality. Like Model 1, however, this multilevel specification is still a-spatial. That is, it cannot account for spatial relationships among counties.

2.2.3. Model 3: Two-level model accounting for spatial clustering—Model 3 ignored the role of states and built directly on Model 1 by situating counties within “spatial patches” rather than within administrative boundaries. For each county ($j = 1, \dots, 2063$), a spatial patch was constructed that consisted of county j and its adjacent counties. Model 3 is therefore a spatial model that ignores hierarchy and accounts for spatial proximity. To account for proximity, information about space is needed. We constructed a weights matrix that defined the spatial relationships among all of the counties in our dataset to provide this information. This spatial weights matrix has dimensions 2063×2063 so that for every pair of counties, a value is given to identify their spatial relationship. We employed a first-order neighborhood structure using Queen-based contiguity to create the spatial weights matrix used in this example. The term “Queen-based contiguity” derives from chess where the queen can move in any direction from her starting location and implies that all shared borders constitute first order neighbors (Fischer & Getis, 2010). To illustrate this scheme, imagine a given map of nine counties $a, \dots, i$ as shown in Fig. 1a.

Using a first-order regional structure using Queen-based contiguity, this map translates into a 9×9 adjacency matrix as shown in Fig. 1b.

Pairs of adjacent counties receive a value of 1 and all other pairs are assigned 0. To look up the relationship between counties a and b , we find county a in the first row and county b in the second column. Reading across the matrix’s first row, the second cell contains 1 because a is adjacent to county b on the map. By contrast, the cell indicating the relationship between counties a and c contains a 0 (first row, third column) because the two counties are not adjacent on the map. Of course, first-order adjacency matters for some research questions while different conceptualizations are more appropriate in other settings. Because of this, there are many ways to construct a spatial weights matrix, as we review in the discussion section of this paper. For our analyses, cell values of zero indicated county pairs that did not share a border and whose estimates did not directly influence each other, while values of 1 indicated neighbor county pairs whose estimates did.

We fit a conditional autoregressive (CAR) model to the data, which is simply another extension of our original random effects model. Instead of nesting a given observation completely within one higher-level unit (the state), Model 3 allows counties to be “cross-classified,” such that their estimates are influenced by a *group* of neighbors, according to our spatial weights matrix as described above. While some counties have few neighbors, the spatial weights matrix is well populated as a whole and allows us to investigate the role of space in the dataset overall. Bayesian estimation of CAR models, which accounts for uncertainty in the data, also helps smooth estimates over space so that areas with few

neighbors are borrowing information from other regions (LeSage and Pace, 2004; Rue and Held, 2005).

In Model 3, counties were situated in spatial patches, which themselves were collections of adjacent counties:

$$LE_i = \beta + (e_{\text{neighbors}(i)} + e_i), \quad e_{\text{neighbors}(i)} \sim N(\bar{e}_{\text{neighbors}(i)}, \sigma_{(e, \text{neighbors})}^2 / n_i), \quad e_i \sim N(0, \sigma_e^2)$$

where $\text{neighbors}(i) = \sum_{j \in \text{neighbors}(i)} w_{ij} e_j / n_i$, and n_i is the number of neighbors for a given county i such that $e_{\text{neighbors}(i)}$ is a second, spatially structured, county-specific deviation from the overall mean life expectancy that is influenced by estimates from adjacent counties in the spatial patch. That is, Model 3 assumes that the distribution of $e_{\text{neighbors}(i)}$ is conditional on neighboring counties. County-level residuals e_i are assumed to be iid after dependence within spatial patches is accounted for. Fig. 2 illustrates the cross-classified structure of this model.

The difference between Model 2, the purely hierarchical multilevel model that accounts for membership in a state, and Model 3, which accounts for spatial effects, is substantive. Model 2 allows for a correlation among the life expectancies of counties in the same state, regardless of their relative geographic locations. Model 3 allowed for correlations among estimates for counties that physically touched, regardless of in which states those counties were located. Model 4 combines features of Models 2 and 3, accounting for both spatial relationships and the effect of county membership in states.

2.2.4. Model 4: Three-level model accounting for hierarchy, and space—Model 4 simultaneously accounted for both state membership and spatial adjacency. We extended Model 2 to Model 4, also a CAR model, by adding a second cross-classified level where spatial patches became level-2 and states were specified as level-3. This data structure is illustrated by Fig. 3. Thus, both spatial dependence of adjacent counties and state membership-induced dependence were modeled simultaneously.

Model 4 situated counties within spatial patches, and at the same time allowed counties to be nested within states:

$$LE_{ij} = \beta + (u_j + e_{\text{neighbors}(i)} + e_{ij}), \quad u_j \sim N(0, \sigma_u^2), \quad e_{\text{neighbors}(i)} \sim N(\bar{e}_{\text{neighbors}(i)}, \sigma_{(e, \text{neighbors})}^2 / n_i), \quad e_{ij} \sim N(0, \sigma_e^2)$$

where $\text{neighbors}(i) = \sum_{j \in \text{neighbors}(i)} w_{ij} e_j / n_i$, and n_i is the number of neighbors for a given county i . Model 4 therefore assumes that state and county-level residuals within states and spatial patches, u_j and e_{ij} , respectively, are iid, while the distribution of $e_{\text{neighbors}(i)}$ is conditional on neighboring counties.

To test the distributional assumptions of each model, we checked whether residuals that were assumed to be iid in fact exhibited evidence of spatial dependence.

Though Models 1 and 2 were fit using iterative generalized least squares estimation, we also fit all four models using Monte Carlo Markov Chain (MCMC) methods to produce a comparable measure of model fit across all models and to accurately quantify spatial dependence in the error terms from Models 2 to 4. Gibbs samplers estimated the fixed and random components of the model separately within the chain, and priors were set to be non-informative. MCMC chains were run for 70,000 iterations, discarding the first 10,000 as burn in and using every 10th sample for posterior inference. We monitored chain

convergence for all models to ensure the parameter spaces were fully explored. We used MLWiN and WinBUGS (Rasbash et al., 2005; Spiegelhalter et al., 2000), software programs especially designed to handle multilevel and spatial data structures, to fit these models.

2.3. Quantifying spatial dependence

Retaining our Queen-based contiguity first-order regional structure, we checked county-level residuals from the four models for evidence of spatial autocorrelation. Testing residuals for spatial autocorrelation involves investigating whether the residual at location i is correlated to other residuals at nearby locations j beyond what would be expected by chance. In other words, it requires examining whether units of observations that are spatially similar also have similar values.

One formal test for spatial clustering of any variable involves calculating a Moran's I statistic (Fischer and Getis, 2010). The Moran's I statistic for each geographic area i is given

$$I_i = \frac{y_i - \bar{y}}{(1/n) \sum_{i=1}^n (y_i - \bar{y})^2} \times \sum_{j=1}^n W_{ij} (y_i - y_j),$$

where W_{ij} comes from the spatial weights matrix. Thus, Moran's I for the residual of a given county i equals the value of that residual, y_i , minus the average county-level residual, $\bar{y}$, divided by the average sum of squares of this difference across all counties,

$(1/n) \sum_{i=1}^n (y_i - \bar{y})^2$. This ratio is then multiplied by the weighted sum of the differences

between county i 's and all other counties' residuals, $\sum_{j=1}^n W_{ij} (y_i - y_j)$, where the weights are derived from the spatial weights matrix. Values of Moran's I value are calculated for each county and then added together for all I_i ($i = 1, \dots, n$) to create a sum that is proportional to a global statistic I indicating overall spatial clustering in the dataset, where the proportionality factor is

$$\gamma = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n W_{ij} \sum_{i=1}^n (y_i - \bar{y})^2$$

and

$$E(I_i) = -\frac{1}{n} \sum_{j=1}^n W_{ij}$$

Possible values of the global Moran's I range from -1 to 1 , with -1 indicating perfect dispersion, values near 0 indicating spatial randomness, and 1 indicating perfect clustering. Under the null hypothesis of no spatial autocorrelation, regression residuals have a slightly negative expected value that approaches zero in large samples (Tiefelsdorf, 1998; Anselin and Rey, 1991; Anselin, 1995). Dispersion would mean that *dissimilar* counties were located together, while clustering would manifest as spatial groupings of *highly similar* counties. By analogy, the black and white squares on a chess board are perfectly dispersed; perfectly clustered squares would form a single white and a single black region. No pattern would be evident under spatial randomness.

In addition to carrying substantive implications, dependence indicates a technical problem for researchers: geographically clustered residuals are not independent and therefore, violate assumptions of many parametric statistical methods. A form of the Moran's I statistic for detecting clustering of regression residuals is given by,

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n W_{ij}} \times \frac{\varepsilon^T W \varepsilon}{\varepsilon^T \varepsilon}$$

where ε is the vector of residuals and ε^T is the matrix transpose.

We calculated in R (R Team & others, 2010) Moran's I values for the county (e_i , e_{ij}) and state-level residuals (u_j) assumed to be iid that were obtained from each of the four models above. Although we report posterior mean Moran's I values and their associated 95% credible intervals for Models 2–4, it is useful to note that Moran's I is Z distributed under the null hypothesis of no spatial autocorrelation and so can be compared to critical z values of the normal distribution for hypothesis testing and inference in frequentist settings, as in the case of our OLS Model 1.

Finally, we compared all four models using the Deviance Information Criterion (DIC), a Bayesian measure of both model fit and complexity that was designed for comparing complex hierarchical models (Spiegelhalter et al., 2002). Lower values of the DIC indicate better model fit after penalties for complexity have been applied.

3. Results

Life expectancy across U.S. counties was about 76.3 years in 1999. Estimates of mean life expectancy did not differ substantially across the four models. Geographic clusters of high and low life expectancy counties also emerged in our analysis. We sought to quantify the degree to which clustering is attributable to hierarchical processes, independent of spatial processes, and the degree to which clustering is attributable to spatial processes, independent of hierarchical processes. It is to this end that each model contributed unique information discernible through an examination of its residuals. Mean values of the MCMC chains for the variance components of each model are reported below, with standard deviations of the MCMC chains given in brackets.

Rather than exhibiting an independent and identical distribution as required, residuals from OLS Model 1 were highly clustered with a global Moran's I value of .638 ($p < .05$) (Table 1). A map of the local Moran's I values for each county (Fig. 4) shows that counties with lower than average life expectancies are concentrated in the Southeast while those with higher than average life expectancies are concentrated in the Midwest and Northeast.

The random components of Model 2 indicated clustering of county measures within states, with level-2, the state level, accounting for roughly 50% of the observed variability in life expectancy estimates ($\sigma_u^2 = 2.15(.489)$, $\sigma_e^2 = 2.04(.065)$). However, spatial analysis of the residuals revealed geographic dependence at both the county (Moran's $I = .289$ (95% CI: .281–.298)) and state-levels (Moran's $I = .641$ (95% CI: .561–.707)), meaning that iid assumptions were also violated in Model 2. As shown in Fig. 5, geographic clusters of counties with particularly high and low life expectancies as compared to their own state-specific means existed within states. Further, geographic clusters of low and high life expectancy states were detected across the country (Fig. 6).

Under Model 3, which situated counties within spatial patches and ignored states, the majority of the total variance was spatially structured and captured by the county-level CAR effect ($\sigma^2_{(e,neighbors)}=4.026(.404)$, $\sigma^2_e=.659(.079)$). The county-level residuals from Model 3 were spatially independent (Moran's $I=.008$ (95% CI: $-.021-.038$ (Fig. 7)). We note that different neighborhood structures would produce different indicators of spatial autocorrelation at the county level, meaning that these findings are scale-dependent. We do not report sensitivity analyses conducted using different spatial weights matrices because our primary focus is to simply demonstrate that multiple forms of dependence can exist, and be accounted for, simultaneously in a dataset.

When we simultaneously controlled membership- and spatially-induced clustering in Model 4, we saw that states (level-3) and spatial structure at the county level, or the CAR effect, (level-2) each explained portions of the total variance in life expectancy. However, the spatially structured component of this variance was much larger than that structured by state membership ($\sigma^2_u=.190(.092)$, $\sigma^2_{(e,neighbors)}=3.830(.391)$, $\sigma^2_e=.664(.077)$). As Figs. 8 and 9 show, Model 4 produced spatially independent residuals at the county level (Moran's $I=.006$ (95% CI: $-.022-.036$)) and at the state level (Moran's $I=-.028$ (95% CI: $-.180-.152$)).

When comparing models for fit and complexity (Table 2), we found that Model 1, the single-level model performed worst overall; the DIC dropped substantially when we nested counties in states (Model 2). We observed additional gains when we fit spatially explicit Models 3 and 4. Model 4, which accounted for both state membership and spatial dependence, performed best overall, though Model 3 was also far superior to Models 1 and 2.

4. Discussion

Residuals from our null, single level model were highly spatially clustered, resulting in substantive and technical implications. Our analysis elucidated the levels at which potentially relevant processes were occurring without requiring detailed knowledge of these processes. Technically, spatially clustered residuals are not iid thus yielding biased standard errors and forcing us to account in some way for spatial patterning. In addition, clustered residuals in this and other settings lead to three broad classes of substantive interpretation. Below we review these three classes of interpretation and key concepts for working with various types of nested data. We also present model specification options that can be used to tailor this spatial multilevel modeling approach to other datasets and research questions.

4.1. Exclusive membership processes

Sometimes, spatial clustering of residuals is driven solely by the fact that membership in geographically-embedded units impacts the outcome of interest. Regarding life expectancy, omitted state-level policy variables could have manifested as spatial autocorrelation because county membership in states is geographically dependent even in the absence of truly spatially-explicit processes. For example, health variations could be driven by state-specific social welfare policies that affect life expectancy. Examples of such policies include income requirements for health care, housing, or food subsidies that are set by states. Research shows that residents are less likely to forego medical care in states with generous public health insurance plans (Gnanasekaran et al., 2008), which in turn could impact longevity and create dependence in life expectancy at the state level. However, this state-induced clustering would also be detected by measures of spatial autocorrelation and mistaken as such even if no relevant spatial processes existed simply because counties in the same state tend to be close together. On the other hand, if membership processes alone create dependence in the data, residuals from the higher-level administrative unit in a properly

specified multilevel model should be independent. Our failure to obtain iid state-level residuals from Model 2 suggests that state membership alone does not drive the spatial distribution of life expectancy.

4.1.1. Membership-induced dependence—Dependence must be considered any time multiple observations are available within the same higher level unit to produce a nested dataset, and may be handled using a multilevel model. Examples include datasets of children within classrooms or homes within towns. Longitudinal datasets are special cases of two-level data structures (time nested within subjects), and require additional considerations (Singer & Willett, 2003).

Whereas single-level models contain only one error term, e_i , whose estimate is calculated as $y_i - \hat{y}_i$, or equivalently $y_i - \hat{\beta}$, a two-level model introduces error terms at both level 1 (observations) and level 2 (higher level units). The level-2 residual, u_j , is equivalent to the average of e_{ij} for values for all $i = 1, \dots, n$ for subject j , times a multiplier that accounts for between-unit variability in the outcome and the number of observations we have for each unit.

When between-unit variability is small and there are few observations for unit j , the residual u_j is ‘shrunk’ by this multiplier towards the overall mean. Under opposite conditions, u_j is pulled very little towards the overall mean. Once u_j is calculated in the two level model, e_{ij} is then calculated as $e_{ij} = y_{ij} - \hat{\beta} - \hat{u}_j$. The total variance in our dataset is therefore portioned into between-unit and within-unit variability, σ_u^2 and σ_e^2 , respectively, where $u_j \sim N(0, \sigma_u^2)$ and $e_{ij} \sim N(0, \sigma_e^2)$. An estimate of σ_u^2 whose credible interval does not include zero generally indicates that observations within units are correlated and that a two-level model specification is required.

Alternatively, fixed effects for higher level units may be included as indicator variables in the fixed part of a single-level model, and robust standard error and clustering techniques such as Huber-White ‘sandwich’ estimator can effectively control for the dependence created by within-unit membership effects. However, this approach consumes many more degrees of freedom than does a random effects approach, and prevents us from generalizing about the overall degree of within- and between-unit variability in our data, which may itself be of substantive interest. It also precludes the analysis of higher level predictor variables. While we intentionally omitted known state-level predictors of life expectancy from this analysis in order to generically demonstrate a method, including higher level covariates is straightforward and extremely useful in most analyses. In fact, it can help identify what characteristics of the higher level units explain observed patterning according to membership in these units. However, a fixed effects approach to modeling membership-induced dependence thwarts such analyses. In this case, putting states in the fixed versus random portion of a model would make it impossible to estimate the effect of state-level income on life expectancy, for example.

4.2. Exclusive spatial processes

It could be that spatial processes influence the outcome of interest without any relevant membership effects. For example, life expectancy could be driven by important, but omitted, spatially-patterned predictor variables that transcend administrative boundaries, such as air quality, for example. Sunlight exposure, environmental contamination, and proximity to job centers are other examples of health exposures that are clustered in space. Contagion is another purely spatial explanation for observed clustering (Barreto et al., 2008). If researchers only consider hierarchy as a possible source dependence in their data, important spatial processes such as these might easily be confused for membership-induced impacts

simply because political and administrative boundaries comprise areas that are adjacent in space. If spatial processes alone drove the clustering of Model 1 residuals, accounting for states should not have improved our models. However, the consideration of states in the purely spatial Model 3 substantially improved the model fit according to the DIC, and variance component estimates from Model 4 showed that states accounted for a significant, though small, component of the total variance in life expectancy. These results suggest that factors beyond spatial proximity also structured our data.

4.2.1. Spatial dependence—Testing for and understanding spatial clustering requires hypotheses of how relevant processes function over space. Contagion and/or influence by spatially-patterned variables may be structured by distance, adjacency, shared access to transportation networks, or other factors, and each of these constructs can be translated into a spatial weights matrix. Researchers should choose a spatial weights matrix based on expert knowledge of the topic under study. Below are commonly employed specifications.

Distance: Distance should be considered for processes that literally operate through space. For example, if studying a disease spread by insect vectors that can only travel up to 800 m from their nests, researchers might create a spatial weights matrix where all pairs of points within 800 m of each other are neighbors and all others are not.

Spatial relationships could alternatively be described by the distance between the two points where cell values would simply be the number of meters between point pairs. To provide further flexibility, these distances could be calculated either in Euclidian terms, along a road network, or by another scheme (Apparicio et al., 2008). Geographic information systems will also allow users to calculate “cost distances” between points that can account for the slope of a walking route, the cost of tolls on a road, or even total trip time so that distances could be conceptualized in terms of minutes commuting or gasoline costs.

Adjacency: Where shared borders or regionalism matter (e.g., immigration or trade), adjacency may be a more useful concept. Adjacency can be conceptualized flexibly (Fischer and Getis, 2010), where the presence, amount, or even spatial orientation of a shared border is used to construct a spatial weights matrix. Further, second-order and higher neighbors (i.e., a neighbor’s neighbor) can be ignored or included.

Nearest neighbors: A hybrid approach between the distance and adjacency approaches is to select k nearest neighbors, with proximity based on centroid distances, from a pool of adjacent areas for each subject.

4.3. Membership and spatial processes

Both geographic membership and spatial processes may matter. This appears to have been the case in our life expectancy example. Because the DIC showed more support for model 4 than the other three Models, we know that integrating membership and space adds enough information about the structure of county-level life expectancy to warrant Model 4’s complexity. Further, level-2 and level-3 variance estimates from Model 4 showed that life expectancy was patterned according to both spatial structure and state membership. Spatial and hierarchical structuring of observations is common (Hägerstrand, 2001), indicating a need for social epidemiologists to better understand the substantive and technical implications of dependence, and to familiarize themselves with methods developed to detect these processes.

5. Conclusion

This dataset of county-level life expectancy in the U.S. exhibited multiple forms of dependence, indicating a simultaneously hierarchical and spatial structure in life expectancy. These findings suggest that life expectancy is associated with both within-state features that are distinct from spatial proximity, and by separate spatial processes. Through this example, we show how the place and space perspectives contribute to our understanding of health and may be modeled simultaneously. For example, by partitioning state membership from spatial effects, we may reduce bias in the associations we estimate between state-level policy or economic factors and health outcomes, such as life expectancy. We also have evidence that spatial processes, either in the form of spatially distributed omitted variables or contagion, matter in addition to state membership.

Multiple forms of simultaneous dependence likely structure many geo-referenced datasets, not just county-level life expectancy in the U.S. Researchers should be aware of the substantive and technical implications of clustering, using methods such as those presented above to detect, interpret, and control for these often overlooked structures.

References

- Anselin L. Local indicators of spatial association — LISA. *Geographical Analysis*. 1995; 27(2):93–115.
- Anselin L, Rey S. Properties of tests for spatial dependence in linear regression models. *Geographical Analysis*. 1991; 23(2):112–131.
- Apparicio P, et al. Comparing alternative approaches to measuring the geographical accessibility of urban health services: distance types and aggregation-error issues. *International Journal of Health Geographics*. 2008; 7(1):7. [PubMed: 18282284]
- Barreto FR, et al. Spread pattern of the first dengue epidemic in the city of Salvador, Brazil. *BMC Public Health*. 2008; 8:51. [PubMed: 18257919]
- Chaix B, Merlo J, Subramanian SV, et al. Comparison of a spatial perspective with the multilevel analytical approach in neighborhood studies: the case of mental and behavioral disorders due to psychoactive substance use in Malmö, Sweden, 2001. *American Journal of Epidemiology*. 2005; 162(2):171–182. [PubMed: 15972939]
- Chaix B, Merlo J, Chauvin P. Comparison of a spatial approach with the multilevel approach for investigating place effects on health: the example of healthcare utilisation in France. *Journal of Epidemiology and Community Health*. 2005; 59(6):517–526. [PubMed: 15911650]
- Diez Roux AV, Mair C. Neighborhoods and health. *Annals of the New York Academy of Sciences*. 2010; 1186:125–145. [PubMed: 20201871]
- Elliott P, Wartenberg D. Spatial epidemiology: current approaches and future challenges. *Environmental Health Perspectives*. 2004; 112(9):998–1006. [PubMed: 15198920]
- Ezzati M, et al. The reversal of fortunes: trends in county mortality and cross-county mortality disparities in the United States. *PLoS Medicine*. 2008; 5(4):e66. [PubMed: 18433290]
- Fischer, MM.; Getis, A. *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*. Springer; 2010.
- Freisthler B. A spatial analysis of social disorganization, alcohol access, and rates of child maltreatment in neighborhoods. *Children and Youth Services Review*. 2004; 26(9):803–819.
- Gatrell, AC.; Foundation, ES. *GIS and Health*. Houghton Mifflin Harcourt; 1998.
- Gnanasekaran SK, et al. State policy environment and delayed or forgone care among children with special health care needs. *Maternal and Child Health Journal*. 2008; 12(6):739–746. [PubMed: 17975719]
- Goldstein, H. *Multilevel Statistical Models*. John Wiley and Sons; 2010.
- Hägerstrand, T. *A look at the political geography of environmental management, Sustainable Landscapes and Lifeways: Scale and Appropriateness*. Cork University Press; 2001. p. 35-58.

- Holsten JE. Obesity and the community food environment: a systematic review. *Public Health Nutrition*. 2009; 12(03):397–405. [PubMed: 18477414]
- Jia H, Moriarty DG, Kanarek N. County-level social environment determinants of health-related quality of life among US adults: a multilevel analysis. *Journal of Community Health*. 2009; 34(5): 430–439. [PubMed: 19554435]
- Jones K, Moon G. Medical geography: taking space seriously. *Progress in Human Geography*. 1993; 17(4):515–524.
- Kearns RA. Place and health: towards a reformed medical geography. *The Professional Geographer*. 1993; 45(2):139–147.
- Kearns RA, Joseph AE. Space in its place: developing the link in medical geography. *Social Science & Medicine*. 1993; 37(6):711–717. [PubMed: 8211286]
- Lawson, AB. *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. CRC Press; 2009.
- Lebel A, et al. The geography of overweight in Quebec: analyzing and visualizing spatial inequalities using second-level residuals. *Canadian Journal of Public Health. Revue Canadienne De Santé Publique*. 2010; 101(2):133–137. [PubMed: 20524378]
- LeSage, JP.; Pace, RK. *Spatial and Spatiotemporal Econometrics*. Emerald Group Publishing; 2004.
- Macintyre, S.; Ellaway, A. Ecological approaches: rediscovering the role of the physical and social environment. In: Berkman, LF.; Kawachi, I., editors. *Social Epidemiology*. Oxford University Press; USA: 2000.
- Mobley LR, et al. Community contextual predictors of endoscopic color-ectal cancer screening in the USA: spatial multilevel regression analysis. *International Journal of Health Geographics*. 2010; 9:44. [PubMed: 20815882]
- Moore DA, Carpenter TE. Spatial analytical methods and geographic information systems: use in health research and epidemiology. *Epidemiologic Reviews*. 1999; 21(2):143–161. [PubMed: 10682254]
- R Team & others. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna Austria; 2010.
- Rasbash, J., et al. *MLwiN Version 2.02*. University of Bristol; Bristol: 2005.
- Raudenbush, SW.; Bryk, AS. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage Publications; 2002.
- Rue, H.; Held, L. *Gaussian Markov Random Fields: Theory and Applications*. CRC Press; 2005.
- Singer JD. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*. 1998; 23(4):323–355.
- Singer, JD.; Willett, JB. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford University Press; US: 2003.
- Spiegelhalter, et al. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2002; 64(4):583–639.
- Spiegelhalter, D.; Thomas, A.; Best, N. Cambridge: MRC Biostatistics Unit; 2000.
- Subramanian SV. The relevance of multilevel statistical methods for identifying causal neighborhood effects. *Social Science & Medicine*. 2004; 58(10):1961–1967. [PubMed: 15020011]
- Subramanian, SV.; Jones, K.; Duncan, C. Multilevel methods for public health research. In: Kawachi, Ichir ; Berkman, Lisa F., editors. *Neighborhoods and Health*. Oxford University Press; 2003. p. 65–111.
- Tiefelsdorf M. Some practical applications of Moran's I 's exact conditional distribution. *Papers in Regional Science*. 1998; 77(2):101–129.
- Waller, LA.; Gotway, CA. *Applied Spatial Statistics for Public Health Data*. Wiley-IEEE; 2004.

a

a	b	c
d	e	f
g	h	i

b

	a	b	c	d	e	f	g	h	i
A	0	1	0	1	1	0	0	0	0
B	1	0	1	1	1	0	0	0	0
C	0	1	0	0	1	1	0	0	0
D	1	1	0	0	1	0	1	1	0
E	1	1	1	1	0	1	1	1	1
F	0	0	1	0	1	0	0	1	1
G	0	0	0	1	1	0	0	1	0
H	0	0	0	1	1	1	1	0	1
I	0	0	0	0	1	1	0	1	0

Fig. 1.

(a) Illustrative map of adjacent counties, (b) illustrative adjacency matrix.

Fig. 2.
Model 3 cross-classified structure.

Fig. 3.
Model 4 cross-classified structure.

Local Indicators of Spatial Association Cluster Map

- ☐ Not significantly clustered
- ☒ High values surrounded by high values
- ☒ Low values surrounded by low values
- ☒ Low values surrounded by high values
- ☒ High values surrounded by low values

Fig. 4.
Model 1 residuals exhibit spatial clustering (Moran's $I=.638$).

Fig. 5. Model 2 county-level residuals exhibit spatial clustering after accounting for membership in states (Moran's $I = .289$, 95% CI: .281–.298).

Local Indicators of Spatial Association Cluster Map

- ☐ Not significantly clustered
- ☒ High values surrounded by high values
- ☒ Low values surrounded by low values
- ☐ Low values surrounded by high values
- ☐ High values surrounded by low values

Fig. 6. Model 2 state-level residuals exhibit spatial clustering (Moran's $I=.641$, 95% CI: $.561-.707$).

Local Indicators of Spatial Association Cluster Map

- ☐ Not significantly clustered
- ☒ High values surrounded by high values
- ☒ Low values surrounded by low values
- ☒ Low values surrounded by high values
- ☒ High values surrounded by low values

Fig. 7.

Model 3 county-level residuals are spatially independent (Moran's $I = .008$ 95% CI: $-.021$ – $.038$).

Local Indicators of Spatial Association Cluster Map

- ☐ Not significantly clustered
- ☒ High values surrounded by high values
- ☒ Low values surrounded by low values
- ☒ Low values surrounded by high values
- ☒ High values surrounded by low values

Fig. 8.

Model 4 county-level residuals are spatially independent (Moran's $I = .006$, 95% CI: $-.022$ – $.036$).

Local Indicators of Spatial Association Cluster Map

- ☐ Not significantly clustered
- ☒ High values surrounded by high values
- ☒ Low values surrounded by low values
- ☒ Low values surrounded by high values
- ☒ High values surrounded by low values

Fig. 9.

Model 4 state-level residuals are spatially independent (Moran's $I = -.028$, 95% CI: $-.180$ – $.152$).

Table 1Moran's *I* for county- and state-level residuals.

Model	Moran's <i>I</i> for county-level residuals	Moran's <i>I</i> for state-level residuals
1: Single-level	.638 *	—
2: Counties in states	.289 **	.641 **
3: Counties in regions (spatial CAR)	.008	—
4: Counties, in regions (spatial CAR) in states	.006	-.028

*
 $p < .05$.**
95% credible interval excludes 0.

Table 2

Bayesian deviance information criterion of model complexity and fit.

	DIC
Model 1	8696.5
Model 2	7367.0
Model 3	6126.6
Model 4	6106.2