The SOCIAL SCIENCE JOURNAL

Special issue on social science and health

EDITOR
Scott Alan Carson
University of Texas, Permian Basin

WITH GUEST EDITOR
Jeff Dennis
University of Texas, Permian Basin

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier’s archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/authorsrights
The effects of spatial patterns of neighborhood risk factors on adverse birth outcomes

Elizabeth Miklya Legerski a,*, Jonathan B. Thayn b

a University of North Dakota, Department of Sociology, 205 Gillette Hall, Grand Forks, ND 58202-7136, USA
b Illinois State University, Department of Geography – Geology, 200A Flemley Hall, Normal, IL 61790-4400, USA

A R T I C L E   I N F O

Article history:
Received 16 May 2013
Received in revised form 5 October 2013
Accepted 5 October 2013
Available online 2 November 2013

Keywords:
Birth weight
Neighborhood effects
Spatial analysis
Health

A B S T R A C T

Neighborhood environments play an important role in shaping health. But how do the conditions of surrounding neighborhoods affect health? Specifically, how do the spatial patterns of neighborhood characteristics shape birth outcomes? Using Census and health data from Wyandotte County, Kansas, we analyze the relationship between spatial patterns of socio-demographic risk factors and incidence of low weight births in neighboring block groups. Using spatial filtering with eigenvectors we identify significant socio-demographic patterns and use them as predictors of low-weight births in a regression model. We identify several patterns that predict significant variability in birth outcomes and find that while some factors, like unemployment, have strong internal neighborhood effects on birth weight they may not have strong external neighborhood effects. We argue spatial filtering methods may improve our understanding of persistent inequalities in health by helping to identify the differential effects of proximate social conditions and spatial interdependencies.

© 2013 Western Social Science Association. Published by Elsevier Inc. All rights reserved.

1. Introduction

Interest in understanding the relationships between neighborhoods and health outcomes grew considerably toward the end of the twentieth century and continues into the present (Ellen et al., 2001; Sampson et al., 2002). Analyses examining individual health and neighborhood environments vary widely in empirical approaches and methodological techniques, spanning disciplinary boundaries from health to the social sciences (Diez Roux, 2001). This burgeoning literature explores the relationship between neighborhood environments and a wide variety of health outcomes including mortality, mental health, rates of disease, quality of life, and birth outcomes (Ford and Browning, 2011; Kim, 2010; Masi et al., 2007; Meijer et al., 2012; Yen et al., 2006). Despite the literature’s diversity, this research consistently shows that local environments matter; individual decisions cannot fully explain the prevalence, or concentration, of poor health in certain areas (Ross and Mirowsky, 2001).

Birth weight is a particularly important indicator of health because it represents the starting gate of health position (Conley et al., 2003). Low birth weight (LBW) is related to infant mortality, childhood mortality, and developmental problems (Cramer, 1995; Li et al., 2003). The effects of LBW extend beyond a child’s early years, putting adults at greater risk for poor health and, consequently, negative socioeconomic outcomes in later life (Conley et al., 2003). LBW is also associated with negative intergenerational outcomes, with some research suggesting that a mother’s birth weight has greater implications for her future children’s health than her adult socioeconomic status (Emanuel et al., 2004).

* Special thanks to Shirley Hill and Mary Zimmerman, University of Kansas, Department of Sociology, for use of their data.
* Corresponding author at: University of North Dakota, Department of Sociology, 205 Gillette Hall, 225 Centennial Drive Stop 7136, Grand Forks, ND 58202-7136, USA. Tel.: +1 701 777 4414.
E-mail address: elizabeth.legerski@email.und.edu (E.M. Legerski).

0362-3319/$ – see front matter © 2013 Western Social Science Association. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.soscij.2013.10.006
Persistent disparities among racial/ethnic groups in many health outcomes, including birth weight, suggest that there may be important unobserved neighborhood or individual level characteristics and mechanisms that need to be accounted for to understand inequalities in health (Sastry and Hussey, 2003). In particular, greater recognition of the complexity of forces that shape health has facilitated growing interdisciplinary interest in the spatial components of neighborhoods and health (Goodchild et al., 2000). This analysis seeks to determine the effects of surrounding characteristics on an area’s LBW prevalence by examining the effects of the spatial patterns of socio-demographic risk factors within neighborhoods. We examine patterns of unemployment, single-parent households, poverty, educational attainment, owner-occupied housing and minority presence. Given the usefulness of spatial modeling techniques in many social science fields, external neighborhood effects are assessed using a method called spatial filtering with eigenvectors. To situate our analysis, we first summarize research examining the relationship between neighborhood characteristics and birth outcomes.

2. Neighborhood inequalities in birth weight and health

About 8% of all births in the U.S. are considered low weight (Centers for Disease Control and Prevention, 2011). While telling, national statistics often mask important inequalities among subpopulations (Robert and House, 2000). For example, Geronimus (1996) found that the odds of low weight births increase dramatically for women in the lowest socioeconomic categories, especially for black women. Maternal race and ethnicity is significantly associated with risk for LBW. But while black mothers are significantly more likely to have LBW babies than whites, Hispanic mothers are significantly less likely than black mothers to have low weight infants (Buka et al., 2003; Roberts, 1997).

Evidence suggests that neighborhood factors may be responsible for a significant proportion of the gap in birth weights between blacks and whites (Sastry and Hussey, 2003). Analyses that examine differences by racial/ethnic status in birth outcomes find community level factors may have differential effects for racial/ethnic groups. For example, some results suggest that while increasing neighborhood poverty exacerbates the problem of LBW for black and Asian mothers, the effects of local socioeconomic conditions may be less significant for white or Hispanic mothers (Pearl et al., 2001; Reagan and Salsberry, 2005). Similarly, Masi et al. (2007) find that the effects of racial density on birth weight are stronger for whites and Hispanics than for blacks. Thus, the processes whereby neighborhood environments affect health outcomes like birth weight may differ across groups, making certain individuals more or less susceptible to local environmental conditions than others (Ellen et al., 2001; Sastry and Hussey, 2003).

Despite variations across groups, individual access to economic resources generally acts as an important buffer to predispositions in LBW (Conley et al., 2003). Although individual factors may be important, they cannot fully explain the incidence of poor birth outcomes (Ross and Mirowsky, 2001). Aggregate community socioeconomic status measures may also have their own independent effect on health outcomes (Robert, 1999). Even when controlling for individual level risk factors, the relationship between concentrated poverty and health persists; living in communities with high levels of poverty puts individuals at greater risk of poor birth outcomes (Buka et al., 2003; Pearl et al., 2001; Roberts, 1997; Sampson, 2003).

Neighborhoods have multiple dimensions, including structural, compositional, social, and cultural characteristics that may have important effects on health (Aber et al., 1997). Neighborhoods may influence health outcomes by shaping access to local institutions and resources, creating stress in the physical and social environment, and by shaping neighborhood networks and norms (Ellen et al., 2001). Neighborhood environments create important contexts that may expose individuals to harmful conditions like violent crime, thus playing a mediating relationship between community socioeconomic level and individual health (Masi et al., 2007; Roberts and House, 2000; Ross and Mirowsky, 2001). While economic hardship may account for some of the variation in birth weight between neighborhoods, when socio-demographic characteristics like income are accounted for, a large proportion of the difference in average birth weight between racial/ethnic groups still remains unexplained (Buka et al., 2003; Cramer, 1995; Roberts, 1997).

Because spatial concentrations of poverty tend to cluster with other forms of disadvantage it is possible that socioeconomic conditions are only one of many different mechanisms that affect health (Gephart, 1997). For example, Sampson et al. (2002) find that the spatial distribution of white and black neighborhoods is responsible for part of the disparity in health outcomes by race. Whereas living in predominantly black neighborhoods may lower the risk of LBW for black women, being in racially isolated neighborhoods may actually increase their risk for adverse pregnancy outcomes (Bell et al., 2006; Roberts, 1997). Pickett et al. (2005) find that the health benefits of positive income incongruity – living in an area where the people around you have higher socioeconomic status than yourself – may be reversed due to the effects of racism. In addition, a growing body of research shows that because neighborhoods are spatially interdependent, the social environment beyond the immediate neighborhood also has important effects on health outcomes like birth weight (Morenoff, 2003). For example, Inagami et al. (2007) find that routine exposure to healthier non-residential neighborhoods in the course of daily activities may improve health. Findings such as these illustrate the importance of considering the proximate and spatial dimensions of neighborhood socio-demographic risk factors for understanding health.

Studies that examine neighborhood effects on socioeconomic variables typically fall into two broad
methodological categories – those that examine internal area effects and those that examine external area effects (Anselin, 2003; Morenoff, 2003).1 Internal area effects are those caused by the phenomena and characteristics of the observation area. For example, birth outcomes may be affected not only by the income of the child’s parents, but also by the median income of the area in which the family lives. These studies typically use Census block groups, or similar geopolitical boundaries, to delineate areas and frequently ignore the effects that proximal areas have on one another.

External neighborhood effects are caused when outcomes in an area are affected by the properties of neighboring areas. “For example, if an area has a low crime rate but the areas around it all have high crime rates, then crime in the surrounding areas could still be an important source of stress for people who live in the low-crime area” (Morenoff, 2003, p. 984; Anselin, 2003). These spillover effects occur on a broader scale and may represent phenomena with spatial effects beyond that of the study area (Morenoff, 2003). Morenoff (2003) suggests that analysts who focus only on internal neighborhood effects will likely miss potentially important neighborhood effects from the wider environment. He also argues that most previous research on health-related neighborhood effects focuses on internal effects and ignores external effects.

Given the importance of examining different levels of neighborhood effects, what are the external effects of neighborhood socio-demographic risk factors on adverse birth outcomes? More specifically, how does proximity, or the conditions of a block group’s neighbors, affect birth outcomes within a given block group? Furthermore, how might spatial methods of data analysis contribute to our understanding of the impact of neighborhood characteristics on health?

3. Methods

3.1. Data

To answer these questions, we analyze socio-demographic and birth data from Wyandotte County, Kansas from 2000 to 2002. Exploratory spatial analysis techniques are employed to identify the multi-neighborhood spatial patterns of several socio-demographic risk variables. These spatial pattern variables are then combined with block group data to determine if these unique patterns have significant effects on LBW prevalence within a neighborhood. The data is aggregated at the Census block group level. Wyandotte County has a total of 161 block groups, eight of which are excluded from analysis due to low (less than 100) or no population (n = 153).

3.2. Independent variables

The independent variable data were collected from the 2000 Census. Wyandotte County has a substantial proportion of both blacks (28.3%) and Hispanics (16.0%), and is one of the poorest counties in the state. Socioeconomic disadvantage is an important predictor of poor birth outcomes and evidence suggests that poverty, unemployment, and lack of education are associated with increased LBW prevalence (Buka et al., 2003; Catalano et al., 1999; Pearl et al., 2001). Research also shows that single-parent families and housing conditions are important predictors of prenatal care and birth outcomes (Charreire and Combier, 2009).

The variable unemployed measures the percentage civilian population unemployed within each block group. Single-parent measures the percentage of single-parent households. HS degree measures the percentage of adults with a high school degree or equivalent. Own house measures the percentage of residents in owner-occupied housing. Poverty measures the percentage of the population with incomes below the poverty level in 1999. Because black women have on average higher LBW prevalence (Pearl et al., 2001; Pickett et al., 2005; Reagan and Salsberry, 2005; Roberts, 1997; Sastry and Hussey, 2003) percent black is included as a measure of the proportion of black (non-Hispanic) residents within each block group.

3.3. Dependent variable

The dependent variable in our analysis is LBW prevalence, the percentage of low weight births within a block group. Low weight births are defined as infants born weighing less than 2,500 g. Birth data were obtained through the Kansas Department of Health and Environment and collected from certificates of live birth in Wyandotte County from 2000 to 2002. During that time period 8,342 births were recorded in the county. Of them, 8% were considered LBW—higher than both the state rate of 6.5% and the national rate of 7.6% in 2000 (State of Kansas, 2003, p. 1).

To avoid confusion, the following definitions are used throughout this paper. The study area is the full extent of the region under investigation (Wyandotte County, Kansas). An area is a subregion within the study area (Census block groups). A neighborhood is a collection of areas that are near to or otherwise spatially related to a central area (an area and its neighboring areas). Thus, a neighborhood is a collection of spatially related areas within the study area. Within the literature on this topic, the phrase “neighborhood effects” is used in a general sense to describe the effects of proximal observations on one another, usually without reference to a specific scale or degree of relatedness and frequently without differentiating between internal and external effects.

3.4. Spatial filtering

The strength of the relationship between areas and their neighborhoods can be measured using Moran’s I (Moran, 1950), which measures the correlation of a variable with itself across space. Moran’s I (MI) is a modification of the Pearson correlation coefficient (Rogerson, 2001). While MI
is not restricted to the range –1 to 1, it generally falls within that window. It can sometimes be as high as 1.3 and it rarely drops below about –0.6. A MI near 1 indicates that most areas are very similar to their neighbors; a MI near –1 indicates that most areas are very different from their neighbors, and a MI of zero indicates that the spatial arrangement of observations’ values is random. MI values can be converted to z-scores and the concomitant p-values can be found (Bivand et al., 2008). Fig. 1 shows examples of a spatial pattern with strong positive spatial autocorrelation, a random spatial pattern, and a pattern with strong negative spatial autocorrelation.

Griffith (2000, 2003) and Tiefelsdorf and Griffith (2007) have developed a process called spatial filtering with eigenvectors that can be used to separate a variable into its spatial component and its non-spatial component. The spatial component is the smooth, strongly spatially autocorrelated pattern within the variable. This component could be considered the external neighborhood effects of the pattern or the portion of the variable that is shared between neighboring areas. Typically, the spatial component has a high and statistically significant MI. The non-spatial component is the randomly distributed component of the variable or the portion of the variable that is not shared between neighboring areas. This component can be considered the internal neighborhood effects of the pattern and it typically has an MI near zero that is statistically insignificant.

Spatial filtering begins by deriving a series of random spatial patterns unique to the spatial arrangement of observations under investigation. The random patterns span the possible range of spatial autocorrelation as measured by MI (Griffith, 2000). These patterns are not associated with any particular variable, but they encompass the possible range of patterns that can occur within the specific spatial arrangement of observations. Creating this series of patterns is somewhat like creating a random variable that is normally distributed.

The second step in spatial filtering is to select from the series of random spatial patterns those patterns that underlie, or are related to, the variable being filtered. The random spatial patterns are generated using eigenfunctions (Griffith, 2000; Thayn and Simanis, 2012), so they are uncorrelated and orthogonal to one another. A stepwise regression (Griffith, 2000; Tiefelsdorf and Griffith, 2007), with the random spatial patterns as the independent variables and the variable being filtered as the dependent variable, is used to select the subset of spatial patterns (Getis and Griffith, 2002; Griffith, 2003; Getis, 2010). The spatial, or external effects, component of the variable is the linear combination of the selected random spatial patterns returned by stepwise regression. The non-spatial, or internal effects, component of the filtered variable is the error term. The spatial and non-spatial components can be added together to find the value of the original, unfiltered variable. The spatial and non-spatial components sum to the original, unfiltered variable. To illustrate, Fig. 2 shows the poverty variable filtered into its spatial and non-spatial components.

There are several advantages to spatially filtering the variables into their spatial and non-spatial components before being submitted to regression models. First, it allows subsequent regression models to give more or less emphasis to the external spatial components relative to the internal, non-spatial components. Generally, more emphasis (larger coefficients in a normalized model) is given to the spatial, external components and the predictive ability of the model increases (Thayn and Simanis, 2012). The increase in R² values is very frequently statistically significant (Thayn and Simanis, 2012).

Second, spatial filtering ensures that each spatial pattern receives its appropriate weight in the regression. Suppose two variables share a common spatial pattern (this is very common). A traditional regression model assigns half of the coefficient to each of the variables so it is difficult to determine the effect of the spatial patterns relative to one another and to the non-spatial components of the variables. When the independent variables have been spatially filtered, each selected random spatial pattern is submitted to the regression only once even if multiple variables share that pattern. Since the random spatial patterns are orthogonal to one another, ensuring that there is no redundancy between patterns, spatial filtering allows the relative importance of each spatial pattern to be easily assessed. The importance of each neighborhood pattern can be determined by examining its coefficient (if the model is normalized) or its p-value.

There is another application of spatial filtering. Regression models with spatially autocorrelated error terms are spatially misspecified (Anselin, 1988), which tends to make regression models too liberal, meaning the null hypothesis is rejected more often than it should be (Clifford et al., 1988). A spatially autocorrelated error term also suggests that at least one important neighborhood effect has been

### Fig. 1. Examples of Moran’s I and spatial autocorrelation.

\[
MI = 0.970, \ p < 0.00 \\
MI = -0.002, \ p = 0.42 \\
MI = -0.546, \ p = 1.00
\]
left out of the model. The missing neighborhood effects are not accounted for in the current set of independent variables. Spatial filtering cannot identify the missing variables, but it does show the spatial pattern of their combined external neighborhood effects (Moreno, 2003; Thayn and Simanis, 2012). The spatial pattern of the missing external spatial effects is a dummy variable that represents at least one missing variable.

Analysts may then use the dummy spatial variable as an aid to identify, and then include, the missing variables. Unfortunately, finding the missing variables is usually prohibitively difficult. Even if the variables that encompass the missing external effects cannot be found, including the dummy spatial variable removes the spatial misspecification from the model (Thayn and Simanis, 2012). Including the dummy spatial variable typically increases the predictive ability of the model (Dormann, 2007) but, since its real-world counterpart is missing, the model is difficult to interpret.

3.5. Defining neighborhoods

A key issue in studying external neighborhood effects is the conceptual and operational definition of a neighborhood (Anselin, 2003; Moreno, 2003). The operational definition of neighborhoods is a critical first step in calculating Moran’s l and in performing spatial filtering. Which areas belong to a specific area’s neighborhood? All subsequent analysis depends on this definition as disparate results may occur with changes in the way neighborhoods are created. Sampson et al. (2002) looked at 40 peer-reviewed studies on the effects of neighborhood social processes on health outcomes and found very little consistency in the way that neighborhoods were operationalized.

The simplest, and a commonly used, definition of neighborhoods is based on contiguity – the neighbors of an area are those that share some portion of its boundary. Another common definition calls areas neighbors if their center points are closer to one another than a specified threshold, so the areal extent of the neighborhood is held constant but the number of areas in the neighborhood varies. A third common definition includes in each area’s neighborhood a specified number of closest areas, keeping constant the number of areas in the neighborhoods but allowing the areal extent to vary.

While these definitions are simple to apply, they have some significant problems. First, they are an oversimplification of a complex spatial pattern. For example, two areas may share a border but have little effect on one another. Similarly, large, sparsely populated areas that are distant from one another may have little effect on each another, and thus should not be included in the same neighborhood despite being within a specified number of closest areas.

Second, these neighbors are based only on the spatial arrangement of areas, and ignore patterns within the values under observation. Consequently, Spielman and Yoo (2009) argue that neighborhoods need to be defined more empirically and Sampson et al. (2002) suggest that the neighborhood definition should be based on a more ecological definition, such as how far children are allowed to wander.

Third, these definitions of neighborhoods rest on the assumptions of spatial isotropy and stationarity. Spatial isotropy is the assumption that the degree of similarity between neighbors is consistent in all directions. Under this assumption, an area is just as related to its northern neighbor as it is to its eastern neighbor. Stationarity is the assumption that the degree of similarity between neighbors is consistent for all neighborhoods. So, if an area on the west of the study area is strongly related to its neighbors, then an area on the east of the study area is just as strongly related to its neighbors. These assumptions are violated in nearly all real data. The assumptions of spatial isotropy and stationarity are relaxed somewhat when using spatial filtering since each randomly generated spatial pattern can encompass different scales of external neighborhood patterns (Diniz-Filho and Bini, 2005; Dormann et al., 2007; Thayn and Simanis, 2012).

To further relax the assumptions of stationarity and spatial isotropy we developed an operational definition of neighborhoods that allows the area of each neighborhood, and the number of neighbors in each neighborhood, to vary for each area empirically. When the size of each area’s neighborhood is allowed to vary, the assumption of stationarity is eliminated and the assumption of spatial isotropy is relaxed. Thus, in this analysis the number of areas in each area’s neighborhood was allowed to increase incrementally until the significance of local spatial autocorrelation was maximized. In this way, areas within more homogeneous regions of the study area were allowed to have more neighbors and areas within heterogeneous regions were allowed to have fewer neighbors. Local spatial autocorrelation (the spatial autocorrelation of an individual neighborhood rather than the aggregate spatial autocorrelation of all neighborhoods across the study area) was measured using the local Moran statistic, proposed by Anselin (1995). The optimal neighborhood was deemed to be the one with the smallest p-value for the area’s local
Morgan statistic. Since the focus of this study is on the prevalence of LBW, the neighborhoods were determined by maximizing the significance of the local spatial autocorrelation of the LBW variable. This definition of neighborhoods was used in all subsequent analyses.

Under this new definition of neighborhoods the mean number of neighbors per area was 4.8 with a minimum of 1 and a maximum of 10. For comparison, under the contiguity definition, the mean was 4.6, with a minimum of 2 and a maximum of 9. The number of neighbors under the contiguity definition were normally distributed, according to a Shapiro–Wilks test (W = 0.917, p < 0.001). The number of neighbors under the new definition was bimodal – many areas had only one neighbor, and a few had as many as ten neighbors. Table 1 shows the MI values for each variable under the new variable–size definition of neighborhoods proposed herein, and the contiguity definition for comparison.

There are several disadvantages associated with this new definition of neighborhoods. First, the number of neighbors for areas whose observations are near the variable mean increases until every area in the study area is included. For example, the LBW value of Census block group 438032 is 7.843 and the mean of LBW is 7.848. Since local spatial autocorrelation is a measure of the relationship between each area and the mean of its neighbors, the significance of local spatial autocorrelation is strengthened with the addition of every possible neighbor. This problem is addressed by limiting the number of neighbors to 10. This threshold was determined based on trial and error.

Second, under the new method neighbors are not always symmetrical, meaning that while area A could be a neighbor of area B, B is not necessarily a neighbor of A. But symmetry is required to generate the series of random spatial patterns used to spatially filter the variables. This problem is addressed by forcing neighbors to be symmetrical. In all other analyses, such as when calculating MI, the asymmetrical neighborhoods are used.

Table 1
Moran’s I values under both the contiguity and the variable–size definition of neighbors.†

<table>
<thead>
<tr>
<th>Variable-size definition</th>
<th>Contiguity definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poverty</td>
<td>0.467</td>
</tr>
<tr>
<td>Unemployed</td>
<td>0.146</td>
</tr>
<tr>
<td>Single-parent</td>
<td>0.410</td>
</tr>
<tr>
<td>HS degree</td>
<td>0.377</td>
</tr>
<tr>
<td>Own house</td>
<td>0.447</td>
</tr>
<tr>
<td>Percent black</td>
<td>0.769</td>
</tr>
<tr>
<td>LBW</td>
<td>0.522</td>
</tr>
</tbody>
</table>

† All MI values were significant at α = 0.01.

The second is a spatially filtered (SF) model where the independent variables were broken into their spatial and non-spatial components before model formulation. This explicitly includes external neighborhood effects in the model (meaning, the effects of neighboring areas are included). This model had a spatially autocorrelated error term, suggesting that not all important external neighborhood effects were included in the model. The third model removes the spatial autocorrelation from the error term and thereby provides the spatial pattern for the missing external neighborhood effects (SF+).

4. Results

4.1. Spatial autocorrelation

Table 1 provides important evidence regarding variability in the strength of external neighborhood effects across the socio-demographic risk factors analyzed. For example, the spatial autocorrelation of unemployment is very low (MI = 0.146, p < 0.001), indicating that this variable has the fewest and weakest external neighborhood effects of those examined. Alternatively, the spatial autocorrelation of percent black is very high (MI = 0.769, p < 0.001), indicating that this variable has the strongest external neighborhood effects. The external effects of the other variables examined varied somewhat between these extremes. As mentioned earlier, the spatial patterns of these external effects may be redundant, meaning their spatial patterns may be shared by more than one variable.

4.2. Model 1, OLS regression

The coefficients and p-values of the OLS model are presented in Table 2. This model is able to predict 18.2% of the variation in low-weight births (p < 0.001). The internal neighborhood effects of these data have a weak but statistically significant ability to predict LBW. Notice that the only independent variables with significant p-values are unemployed and percent black. In addition, LBW is strongly spatially autocorrelated in this model (MI = 0.571, p < 0.001), as is the error term (MI = 0.425, p < 0.001).

4.3. Model 2, spatial filtering

The first spatially filtered (SF) model, which includes external neighborhood effects, is dramatically stronger than the OLS model (R² = 0.557, p < 0.001). The coefficients and p-values of this model are shown in Table 2. In this model each independent variable was filtered into its spatial and non-spatial components. The selected random patterns that collectively define the neighborhood patterns of the independent variables, and the variables which they helped filter, are listed in Table 3. These patterns are also presented in Fig. 3.

After the spatial component has been removed from the independent variables, only the unemployed variable is still a significant predictor of LBW. Remember that this is the least spatially autocorrelated independent variable. The other variables contain such strong external neighborhood effects that, once those effects have been filtered...
from the variables, they are no longer significant predictors of low-weight births. Random patterns 13, 15, 19, and 24 are the most significant predictors of LBW (analysis not shown), indicating that these patterns may represent the dominant external neighborhood effects of these data. Table 3 shows that these spatial patterns help define neighborhood patterns in unemployment, single-parent households, owner-occupied housing, poverty, and the black population.

Fifty percent of the random spatial patterns that filtered HS degree are unique to that variable. Forty percent of the patterns used to filter single-parent are unique. As Table 3 shows, these variables contain external neighborhood effects that are not a part of any other variable used in this study. Poverty and percent black also contain unique external neighborhood effects. Since the external effects of these variables appear to be unique, perhaps they deserve further analysis and examination. At the very least, researchers should be sure to include these variables in future work since their external effects are not shared by other variables in this study. Table 3 also shows that some variables are more spatially redundant than others. For example, random pattern 17 was the most commonly selected pattern – it encompassed part of the external neighborhood effects of five of the six independent variables (excluding unemployed). As this is the predominant pattern of external effects, future research should focus on this pattern, at least in these data. Random patterns 1 and 19 each encompass part of the external neighborhood effects of three of the independent variables, indicating that these may be important secondary patterns of external neighborhood effects.

The error term of the SF model is moderately but significantly spatially autocorrelated (Mi = 0.128, \( p < 0.001 \)), indicating that there are important external neighborhood effects that contribute to modeling birth outcomes that are not included in this model. In the next model, these unobserved variables are controlled.

### 4.4. Model 3, spatial pattern for missing effects

Model 3 (SF+) is similar to the previous spatially filtered model except that a selection of the random spatial patterns have been used to create a spatial pattern that, when used as an independent variable, removes the spatial autocorrelation from the model error term. This pattern

---

**Table 2**

Regression results predicting low-weight births for OLS and spatially filtered models (n = 153).^a^  

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>Model 1 (OLS) standard model</th>
<th>Model 2 (SF) spatially filtered variables</th>
<th>Model 3 (SF+) spatially filtered and spatial misspecification variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poverty</td>
<td>−0.021^c^</td>
<td>0.006</td>
<td>0.015</td>
</tr>
<tr>
<td>Unemployed</td>
<td>0.063^†^</td>
<td>0.042^*</td>
<td>0.212^*</td>
</tr>
<tr>
<td>Single-parent</td>
<td>0.024</td>
<td>−0.018</td>
<td>−0.035</td>
</tr>
<tr>
<td>HS degree</td>
<td>−0.003^*</td>
<td>0.017</td>
<td>0.030</td>
</tr>
<tr>
<td>Own house</td>
<td>0.001</td>
<td>−0.009</td>
<td>0.000</td>
</tr>
<tr>
<td>Percent black</td>
<td>0.037^*</td>
<td>−0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>Pattern 1</td>
<td>−0.088***</td>
<td>−0.132***</td>
<td></td>
</tr>
<tr>
<td>Pattern 9</td>
<td>−0.055^*</td>
<td>−0.101^*</td>
<td></td>
</tr>
<tr>
<td>Pattern 10</td>
<td>−0.044^*</td>
<td>0.082^+</td>
<td></td>
</tr>
<tr>
<td>Pattern 12</td>
<td>0.059</td>
<td>−0.172^*</td>
<td></td>
</tr>
<tr>
<td>Pattern 13</td>
<td>−0.118***</td>
<td>0.165^+</td>
<td></td>
</tr>
<tr>
<td>Pattern 15</td>
<td>0.089***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pattern 17</td>
<td>0.060^*</td>
<td>0.114^*</td>
<td></td>
</tr>
<tr>
<td>Pattern 19</td>
<td>−0.071***</td>
<td>−0.125^*</td>
<td></td>
</tr>
<tr>
<td>Pattern 22</td>
<td>0.066</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pattern 23</td>
<td>−0.040</td>
<td>−0.075^*</td>
<td></td>
</tr>
<tr>
<td>Pattern 24</td>
<td>−0.086^*</td>
<td>−0.107^*</td>
<td></td>
</tr>
<tr>
<td>Pattern 25</td>
<td>−0.051^*</td>
<td>−0.065^+</td>
<td></td>
</tr>
<tr>
<td>Pattern 29</td>
<td>−0.086^*</td>
<td>−0.099^*</td>
<td></td>
</tr>
<tr>
<td>Missing pattern</td>
<td></td>
<td>0.139^+</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.051^***</td>
<td>0.275^^**</td>
<td>−0.021</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.182^**</td>
<td>0.557^***</td>
<td>0.739^^**</td>
</tr>
</tbody>
</table>

^a^ Standardized coefficients shown.  
^†^ \( p < 0.05 \).  
^*^ \( p < 0.01 \).  
^^^ \( p < 0.001 \).

**Table 3**

Unique spatial patterns that define the neighborhood patterns of the independent variables.\(^b^\)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Spatial patterns used to filter the variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poverty</td>
<td>1, 12^a^, 17, 24^a^, 29^a^</td>
</tr>
<tr>
<td>Unemployed</td>
<td>19^a^, 22</td>
</tr>
<tr>
<td>Single-parent</td>
<td>1, 9^a^, 10, 13^a^, 17, 19^a^, 22^b^, 23, 24^b^</td>
</tr>
<tr>
<td>HS degree</td>
<td>17, 25^a^</td>
</tr>
<tr>
<td>Own house</td>
<td>17, 19^a^</td>
</tr>
<tr>
<td>Percent black</td>
<td>1, 10, 13^a^, 15^b^, 17, 23</td>
</tr>
</tbody>
</table>

\(^b^\) Each unique pattern identified was randomly numbered. Fig. 3 shows each pattern and its associated MI value.  
\(^a^\) Indicates that the associated pattern is unique to that variable.  
\(^b^\) Is used to identify the most significant spatial patterns identified (analysis not shown, see Footnote 2).

---

\(^2^\) The selected random patterns were regressed against LBW in a model that examined only external neighborhood effects (\( R^2 = 0.53 \)). These are the four patterns that were significant at \( p < 0.001 \).
represents the missing external neighborhood effects that contribute to LBW. As Table 2 shows, including this pattern in the model increases the strength of the model significantly ($R^2 = 0.739$, $p < 0.001$) over both the OLS and SF models. The error term of the SF+ model is no longer spatially autocorrelated ($MI = -0.117$, $p = 0.594$), so this assumption of a linear model is now met. The pattern associated with the missing external neighborhood effects is shown in Fig. 4 and is strongly and significantly spatially autocorrelated ($MI = 0.456$, $p < 0.001$). The predicted low-weight births for each model are shown in Fig. 5.

5. Discussion

These findings illustrate the usefulness of spatial forms of data analysis to understand the complexity of factors that shape neighborhood environments and health. Like Morenoff (2003), we find evidence that the “contextual effects on birth weight extend to the social environment
Beyond the immediate neighborhood’s effects, at least four unique spatial patterns are identified that produce significant external effects on LBW prevalence (Fig. 3, patterns 13, 15, 19, and 24). These spatial patterns reflect patterns in access to work (unemployment), social support (single-parent households), and wealth (home ownership and poverty), in addition to racial composition (the percent black population).

One of our findings is that some of the spatial patterns related to racial composition have strong external neighborhood effects on LBW. When these spatial patterns are included in the models the percent black population is no longer a significant predictor of LBW. Distinct from (yet complimentary to) other research that documents the effects of racial density on health outcomes (Bell et al., 2006; Masi et al., 2007; Pickett et al., 2005), this finding suggests that patterns of racial composition may also shape health. Although we are unable to determine in this analysis the specific mechanisms responsible for this relationship, it is possible that the social effects of racism and discrimination on health may partially explain these external neighborhood effects.

We also find evidence that some factors may have strong internal neighborhood effects and yet lack strong external neighborhood effects. Although the spatial pattern of unemployment has very weak external neighborhood effects, it is the only variable analyzed that produces significant internal neighborhood effects once the spatial components of all the variables analyzed are included in the models. These findings support the work of others who find that even low concentrations of unemployment in a Census block group are related to higher LBW prevalence (Pearl et al., 2001). Similar results are not found for a more general wealth measure (percentage owner occupied housing), suggesting that the stress caused by unemployment (the lack of immediate financial resources, greater conflict within families, and the potential loss of access to health insurance) has more direct consequences for the immediate neighborhood environment in regards to health, as opposed to exerting stress on surrounding neighborhood environments.

Our findings also suggest that there are other unobserved external neighborhood factors with strong spatial components that play an important role in shaping LBW prevalence. While our analysis does not allow us to identify these unobserved factors specifically, we are able to determine that their effects on poor birth outcomes may be a result of forces related to the wider spatial region rather than the direct effects of a particular Census block group. Thus, these findings reinforce the notion that while the immediate conditions of a person’s environment certainly shape their health, the effects of neighboring neighborhood environments on health outcomes cannot be ignored (Morenoff, 2003) or the spatial components of neighborhood effects (Spielman and Yoo, 2009). Future research might include the spatial patterns of violence in communities, access to healthcare and recreation facilities, or proximity to sources of pollution, such as power plants and waste disposal sites, in order to account for some of the unobserved external neighborhood factors that shape birth outcomes.

Although common practice, measuring neighborhoods using Census defined block groups is an imperfect operationalization that may not respect street patterns and social

---

**Fig. 5.** Model results.
networks. Nonetheless, Census block groups do provide a relatively small and geographically precise unit of analysis that tends to produce results consistent with analyses examining other levels of geography (Sampson et al., 2002). This analysis has attempted to improve upon common contingency based definitions of neighborhoods by using a data driven, variable-size definition of neighborhoods. Because we are examining aggregate level data, we must also be careful to avoid the ecological fallacy – making conclusions about individuals based on group-level data. While the focus of our analysis is on the relationship between aggregate level patterns and outcomes, we recognize that individual-level factors and processes also shape individual health outcomes.

Spiegelman and Yoo (2009) demonstrate that the common practice of selecting the scale of analysis by fitting the data to multiple models and selecting the model that is most statistically significant often does not capture the true areal extent of the external neighborhood effects being studied. They argue that unaggregated data should be used and that the neighborhoods should be constructed based on more natural definitions that reflect the theories underlying the analysis. While the variable-size definition of neighborhoods that we employ here helps minimize this problem (by relaxing the assumption of spatial isotropy and stationarity and by basing the neighborhoods empirically on the data), it is unlikely that it eliminates it entirely. Nonetheless, we feel that our method is currently the best method of analyzing external neighborhood effects in aggregated data. However, we agree with Spiegelman and Yoo (2009): Whenever possible, unaggregated data are preferred.

Despite the limitations of our analysis, our findings support the use of spatial filtering with eigenvectors to identify and understand how patterns of socio-demographic risk factors shape patterns of health outcomes. Geographic data is now becoming more widely available. Most socio-economic data have a spatial component and including geographic location and broader-scale neighborhoods in analyses will enhance our understanding of social processes and the environment in which they occur. Future research ought to explore the external neighborhood effects of other socio-demographic risk factors on health. Doing so not only serves to enhance research practice, but also promises to support efforts to improve public health by allowing analysts to better target support services and interventions based on the patterns of risk within neighborhoods.


References

Author’s personal copy


