Methods and Challenges of Analyzing Spatial Data for Social Work Problems: The Case of Examining Child Maltreatment Geographically

Bridget Freisthler, Bridgette Lery, Paul J. Gruenewald, and Julian Chow

Increasingly, social work researchers are interested in examining how "place" and "location" contribute to social problems. Yet, often these researchers do not use the specialized spatial statistical techniques developed to handle the analytic issues faced when conducting ecological analyses. This article explains the importance of these techniques when analyzing spatial data, describes appropriate spatial statistical techniques, provides an illustration of how using inappropriate statistical techniques with spatial data can produce biased estimates of statistical tests, and discusses challenges to conducting spatial analysis. The study involved analyzing data for 941 census tracts for structural factors related to child maltreatment using traditional ordinary least squares (OLS) regression and generalized linear squares (GLS) spatial regression. When using OLS, results showed that immigrant populations/child care burden was negatively related to rates of child maltreatment but not related when using GLS, which controls for correlations between these spatial units. Relying on OLS regression techniques to create interventions to reduce child maltreatment in spatial areas could result in developing ineffective strategies that fail to reduce maltreatment.

KEY WORDS: *child maltreatment; GIS; neighborhoods; spatial analysis*

Advances in geographic information systems (GIS) technology are leading social work researchers and practitioners to think about social welfare problems more "spatially." For example, Albert and Catlin (2002) studied whether states adjusted their welfare benefits to be more like adjacent states to discourage welfare recipients from crossing state borders to receive more benefits. Essentially, their question was one of space: Are welfare benefits in states located next to each other more alike than states farther away? Social work researchers increasingly pose similar types of questions, yet the methods used in social work research to examine this greater awareness of space have not kept pace with the questions asked.

One area where the biggest awareness of space in social work practice and research can be seen is through the increased use of GIS-generated maps, which show the distributions of problems across communities or neighborhoods (Noble & Smith, 1994; Queralt & Witte, 1998; Robertson & Wier, 1998). Although the use of these maps has proliferated, analyses of spatial data from these maps continue to be conducted using traditional analytic techniques, such as regression, analysis of variance, or tabulated frequencies of events across spatial areas including neighborhoods and communities. Such techniques do not account for several of the unique problems associated with analyses of spatial data. At the very least, the neglect of these techniques when studying neighborhoods or other areas may result in specification error and inconsistent findings across studies. At the very worst, interventions to reduce social problems may be recommended to policymakers based on biased statistical tests. The purpose of this article is to explain the importance of using specialized statistical techniques for analyzing spatial data, describe these techniques, provide an example that illustrates how using inappropriate statistical techniques of spatial data can produce biased estimates of the statistical tests, and discuss challenges to conducting spatial analyses.

WHY SPATIAL DATA ANALYSIS?

Spatial data analysis refers to the examination of some process in space and its relationship to other spatial phenomena (Bailey & Gatrell, 1995) and can be used to describe relationships between points, lines, or areas. This article focuses on those techniques used to describe areal data—data collected from geographic areas such as cities, zip codes, and census tracts. In the past 10 years, there has been a proliferation of studies of neighborhood areas examining many different social problems (see Burton & Jarrett, 2000; Leventhal & Brooks-Gunn, 2000; Sampson, Morenoff, & Gannon-Rowley, 2002 for a review). Predicated on the person-in-environment philosophy that characterizes the social work profession, social welfare research is being conducted that examines how environment may affect the prevalence of social problems. Social workers have embraced the change in focus of social science research from examining solely the individual to investigating the person in his or her environment. Yet research techniques that have been developed for analysis of small areas (for example, neighborhoods) are not being applied in these studies.

Although social work problems always occur in some "place," researchers rarely account for or examine that location in detail. Place or space is a useful piece of knowledge for many reasons. For example, knowing where populations who experience different social welfare problems reside can aid practitioners in targeting services in areas where they are most needed. The concentration of problems in certain geographic locations allows social work researchers and practitioners to begin to examine how the physical environment in those areas may be contributing to problems faced by its residents. By examining place more explicitly in research and practice, social workers can begin to take a more methodical approach to combating social problems that are the result of both individual and environmental determinants of behavior.

One analytic issue that can bias results in studies of geographic places is that area units located next to each other may share similar characteristics (Bailey & Gatrell, 1995). It is often the case that measures taken from adjacent spatial units (that is, neighborhoods that share a boundary) are correlated (Bailey & Gatrell). It is also true that errors in measurement from statistical models of problem outcomes in these areas may also be correlated, violating the assumption of unit independence required for unbiased application of many traditional statistical techniques (Cliff & Ord, 1973, 1981). Statistical procedures have been developed to assess and control for spatial autocorrelation and are described in this article.

A second analytic issue that can bias results in neighborhood studies is the failure to recognize that neighborhood boundaries are permeable. For this reason, aspects of the social and physical environments that affect problems in one area (for example, availability of social services) may also affect problems elsewhere. People move in and out of neighborhood areas to go to work, school, shop, or visit with friends and family. Thus, although problems may be concentrated in certain areas of a community, this does not mean that either people living in those areas or their specific neighborhood environments are the only causes or correlates of those problems.

To attribute problems that occur in neighborhoods solely to characteristics of neighborhood residents can lead to the misallocation of resources or interventions to populations who do not need them or to places that do not contribute to the problem. For example, it has been suggested that some types of drug sales occur more frequently in neighborhood areas with high rates of disadvantage and low guardianship (Eck, 1995; Saxe et al., 2001). However, researchers have found that these areas do not experience greater incidence of drug dependence compared with wealthier areas (Saxe et al.). This suggests that residents of less disadvantaged neighborhoods may be traveling to more disadvantaged neighborhoods to purchase illicit drugs. Focusing efforts to prevent drug use and abuse primarily in disadvantaged neighborhoods will not eliminate the demand for drugs as long as residents from other areas are continuing to use these illegal substances. As a result, social workers may fail to see rates of these problems decline in those areas. Techniques have been developed that enable researchers to account for these spatial relationships between characteristics of neighborhood areas and local problems. Using these techniques, researchers can better model these problems and create better-informed interventions.

SPATIAL STATISTICS

Descriptive Spatial Techniques

As with traditional statistical analysis, techniques that describe spatial data have been developed to help researchers more fully examine their data. Descriptive spatial techniques can be used to generate hypotheses about whether data collected from different areas exhibit "complete spatial randomness" or "spatial dependence" (Bailey & Gatrell, 1995). For example, do measures of problem outcomes taken from different areas, such as zip codes or census tracts, exhibit some systematic pattern or are these problems distributed randomly across these areas?

GIS can facilitate mapping problem outcomes, providing descriptions of the distributions of outcomes that are similar to those provided by frequency tables. The unique contribution of geographic representations is that the frequencies, rates, or counts shown in maps can suggest spatial patterns that may have empirical validity once more rigorous statistical techniques are applied. Thus, descriptive spatial statistical techniques form a bridge between GIS mapping and spatial statistics in three ways. First, descriptive spatial techniques present spatial data in ways that get at the spatial structure or patterning of the data across geographic locations. Second, problems of spatial scale and resolution can be addressed directly; patterns that are apparent at one geographic scale may exhibit random variation at another scale. Third, as with any graphical medium, good data quality is essential and bad data are easy to overlook when presented in graphical, rather than tabular form. The process of engaging in exploratory analysis with spatial data (that is, searching for outliers, clusters, sparsity, correlation between spatial units) can identify peculiarities and errors in the data (Anselin, 2003).

Known as exploratory spatial data analysis (ESDA), these techniques identify and describe properties unique to spatial data, like spatial autocorrelation and spatial outliers. Figure 1a shows an example of ESDA quartile map of substantiated rates of child maltreatment for 2000 for three counties in California (Alameda, Sacramento, and Santa Clara). The map appears to show clustering of high rates in the central and northern parts of Sacramento County and in the northwest portion of Alameda County. However, this map alone provides only descriptive information about the spatial patterns. A more advanced diagnostic technique, the Moran coefficient, allows researchers to explore relationships among spatial units to quantify the extent to which a spatial association is present. This spatial pattern occurs when spatial units are correlated with one another on some dependent measure. The Moran coefficient estimates the degree of relationships between measures taken from adjacent spatial units, for example, a target unit and its adjacent neighbors on a map. The expected value for the Moran coefficient is $(-1/(N-1)$, is bounded approximately by -1 and $+1$, similar to a correlation coefficient, and has a significance level associated with it. A significant Moran coefficient indicates that the rates of problems across neighborhood areas are correlated so spatial statistics may be necessary.

ESDA can also be used to identify spatial clusters and spatial outliers. In Figure 1b statistically significant clusters are designated by high-high and low-low census tracts. This is done through the use of local indicators of spatial autocorrelation (LISA) analyses. Rather than looking at the entire area as a whole, LISA statistics assess the level of spatial autocorrelation for each individual spatial unit comparing it to neighboring areas (Anselin, 2003). Through this procedure the Moran coefficient and its significance level is calculated for each spatial unit. Spatial clusters are identified as those spatial units that have high levels of spatial autocorrelation and are surrounded by units that also have high levels of spatial autocorrelation or those units with low levels of spatial autocorrelation surrounded by units with low spatial autocorrelation. Spatial outliers are those spatial units that may have high LISA but are surrounded by spatial units with low levels of spatial autocorrelation and vice versa. The high-low and low-high census tracts in Figure 1b point to statistically significant spatial outliers. A word of caution is appropriate at this point: As with all statistical analyses, descriptive statistics are not able to ascertain information about the underlying nature of these spatial patterns. For instance, spatial outliers identified through ESDA procedures may be the result of the unique characteristics of that place. Once those characteristics are controlled in spatial multivariate analyses, these spatial outliers may disappear. If they do not, spatial smoothing procedures where the outliers are replaced by the value of surrounding areas can be used.

Connection and Distance Matrices

To estimate a Moran coefficient, it is necessary to have a map of the spatial units for which the coefficient will be calculated and a procedure by which the "adjacencies" will be identified; a method that establishes what units are neighbors to one another. This connection or distance matrix is the element

Figure 1: Quartile Map, Cluster and Outlier Maps for Incidence of Substantiated Child Maltreatment Reports in Alameda, Santa Clara, and Sacramento Counties, by Census Tract

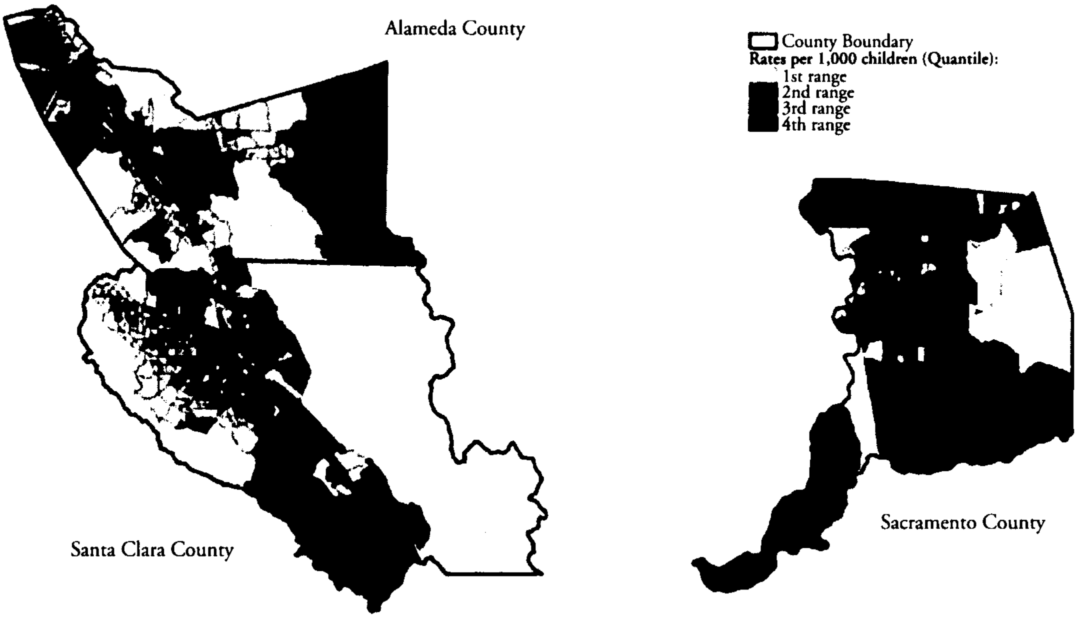


Figure 1a: Quartiles of Substantiated Rates per 1,000 Children

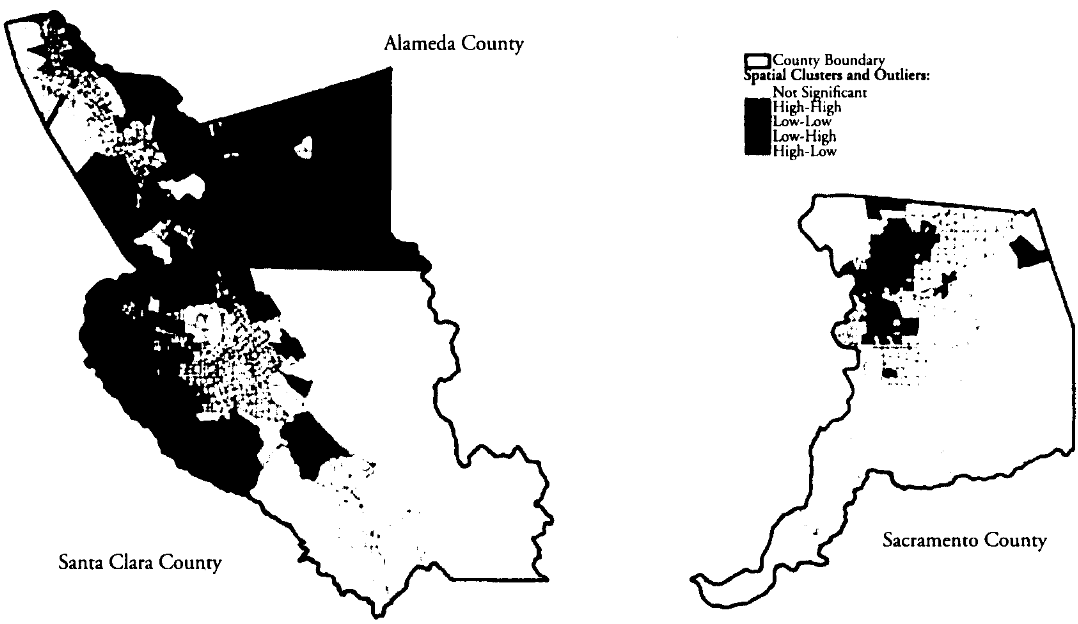


Figure 1b: Spatial Cluster (High-High, Low-Low) and Outlier (Low-High, High-Low) Maps

that puts the "space" into spatial analysis. For example, these matrices allow us to identify whether one zip code is "next" to another. Connection matrices are squares containing the same number of columns and rows equal to the total number of area units. The most simple connection matrix identifies whether neighborhood areas share a common boundary and is often used in the absence of better information about relationships among spatial units. For neighborhoods that share a common boundary, a 1 is placed in the cell of the connection matrix that refers to those neighborhoods. If two neighborhoods do not share a common boundary, a 0 is place in the matrix. By convention, zeros are placed along the diagonal, indicating that a neighborhood is not adjacent to itself. Note that cases for which only the point touches another neighborhood are often not considered adjacent and a 0 is placed in the connection matrix. (Using a metaphor from chess, these are often referred to as "bishop's cases," distinct from "rook's cases" in which extensive borders are shared between places.)

For illustration, Figure 2 shows a connection matrix that assigns "neighbor" status to adjacent areas using rook's criteria, and does not consider neighbors of neighbors (second order connections). Figure 2a is an example of a map with nine spatial areas, A through I. The matrix in Figure 2b summarizes the spatial relationships of these areas: A is a neighbor of B and D but not E, as indicated by 1s in the matrix. The matrix will change in size and content according to what spatial scale and what connection criteria is chosen. For instance, Alameda, Sacramento, and Santa Clara counties (from Figure 1) have 941 census tracts and the connection matrix is composed of 885,481 entries (941 x 941 units). When units are aggregated or disaggregated, the number of neighbors each unit shares may change.

The second type of matrix, a distance matrix, uses interval measures to identify connections between neighborhood areas. Rather than denoting connections between areas with a 0 or 1, a distance matrix may use the length of a shared boundary between neighborhood areas. This type of matrix

Figure 2: Development of Connection Matrix Using Rook's Criteria and Depiction of Spatial Autocorrelation

Figure 2a: Map of Adjacencies

A	B	C
D	E	F
G	H	I

Figure 2b: 9 x 9 Connection Matrix

	A	B	C	D	E	F	G	H	I
A	0	1	0	1	0	0	0	0	0
B		0	1	0	1	0	0	0	0
C			0	0	0	1	0	0	0
D				0	1	0	1	0	0
E					0	1	0	1	0
F						0	0	0	1
G							0	1	0
H								0	1
I									0

Figure 2c: Positive Spatial Autocorrelation

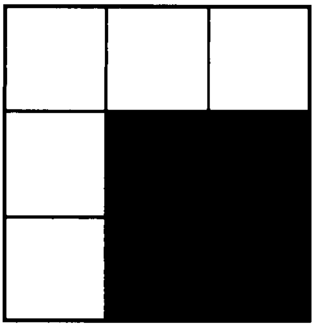
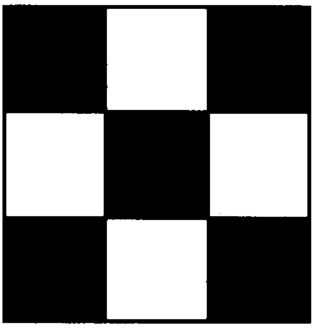


Figure 2d: Negative Spatial Autocorrelation



can also be used to represent the distance between one neighborhood area and all other neighborhood areas by recording the distance from the geographic center, or the *centroid*, of the neighborhood to other neighborhoods and recording that in the cells of the distance matrix. A distance matrix may be used when theory suggests that distance across areas may better explain rates of problems in local areas. For example, neighborhoods that share a longer boundary may experience more spillover or movement across that boundary, resulting in a greater effect of people living in that neighborhood on rates of problems there than in places where the shared boundary is a shorter length. The decision to use one type of matrix or the other depends on the theory guiding the overall research question and how this theory describes the possible relationship or "connectedness" between spatial units.

Multivariate Spatial Statistical Techniques

As mentioned earlier, researchers have found that measures from adjacent spatial units (that is, neighborhoods that share a boundary) are often correlated and that errors in measurement from statistical models may also be correlated between spatial units (Bailey & Gatrell, 1995). The biases that such correlations introduce into spatial models are important to diagnose and control. One way to do this is through a GLS model that assesses and controls for spatial autocorrelation. The majority of neighborhood or small area studies do not take into account this spatial relationship between neighborhood areas. Instead, they treat each neighborhood as an individual unit, violating the assumption of unit independence.

Spatial autocorrelated error can be either negative or positive (Cliff & Ord, 1973). In Figure 2c positive spatial autocorrelation is depicted, with the dark areas representing areas with high rates of child maltreatment and white areas representing areas with low rates of maltreatment. Adjacent spatial units have the same or similar value on neighborhood measures. When positive spatial autocorrelation exists, the probability of committing Type I errors is increased. In other words, positive spatial autocorrelation inflates nominal significant relationships. On the other hand, negative spatial autocorrelation (as shown in Figure 2d) increases Type II errors—failing to find significant relationships when they do in fact exist. This figure has a checkerboard appearance, indicating that spatial units with high levels of some variable (dark areas) are surrounded by neighborhoods with low levels of the same variable (light areas).

Statistical procedures have been developed that assess the level of spatial autocorrelation in residuals from spatial models and control for the degree of spatial autocorrelation found. OLS regression models are the typical analysis procedures used in neighborhood studies of child abuse and neglect. These models generally regress some measure of problem rates (for example, number of child abuse reports per 1,000 children) over a set of independent measures (often obtained from census data), each measure taken from some collection of geographic areas such as census tracts. In this setting, the use of OLS regression procedures in the presence of significant spatial autocorrelated error biases statistical tests of model coefficients and can result in substantive Type I or Type II errors (Griffith, 1988). Examining the residuals from a regression model, researchers can assess the occurrence of significant spatial autocorrelation in an OLS regression model. A significant Moran I statistic measured on the residuals will identify spatial autocorrelation. A nonsignificant Moran I statistic means that the assumption of unit independence for that OLS regression model is appropriate (Griffith).

Once the level of spatial autocorrelation has been assessed and found to be present in the analysis, a GLS regression procedure allows researchers to control for spatial autocorrelation in the model (Griffith, 1988). One approach to dealing with spatial autocorrelation is to assume that spatial dependence is found to be a nuisance, related only to correlated error and otherwise unrelated to the independent and dependent measures (Bailey & Gatrell, 1995). In this model, spatial autocorrelation is controlled for by using the connection matrix discussed earlier as part of the error term. This is called the nuisance parameter (also called spatial error or simultaneously autoregressive) model.

An alternative approach to the nuisance parameter model is the fixed autoregressive response model. The fixed autoregressive response (also called spatial lag or conditionally autoregressive) model assumes that spatial dependence is a function of the value of the dependent measure itself (Bailey & Gatrell, 1995). Like models of disease contagion, autoregressive response models assume that levels of the outcome measure within each unit are causally related to the observed outcome measure in spatially related units. This model is appropriate when it is assumed that the spatial patterns observed are not due to error, but rather due to some spatial process, such as diffusion, that is contributing to the social problem. Thus, this model may be used to study how police enforcement changes patterns of drug markets in local areas (that is, diffuses drug markets from local areas to places in adjacent areas). For the example used here, it is assumed that child maltreatment across neighborhood areas is not represented by a diffusion process but, rather, separate spatial processes conjointly affect risks for this problem outcome within areas, with errors correlated between areas. Correlated measurement error is assumed independent of the outcome, but likely due to unmeasured exogenous spatial determinants of the outcome.

Spatial Lags

Although either the errors in measurement or values of the dependent variables measured across spatial units may be autocorrelated, suggesting the use of one or another statistical model, an essential and often overlooked feature of much spatial analysis is its ability to model relationships **between** neighborhoods. Having controlled for potential sources of spatial autocorrelation, the researcher is free to examine relationships among independent variables and outcomes **across** spatial units. Because neighborhood boundaries are permeable and people are able to move between neighborhood areas fairly easily, neighborhood characteristics of one area (for example, poverty) may be associated with rates of problems in adjacent areas (for example, maltreatment). For this purpose, the same connection matrix that is used to diagnose spatial autocorrelation can be used to calculate statistics from areas that are connected to each target area. In the simplest case, this entails taking the average of some measure from spatially connected units (for example, levels of poverty in adjacent areas) and including it as a regressor in a spatial regression model of a problem outcome measured within each target unit. If the coefficient relating the measure from surrounding units is correlated with the outcome within target units, a spatial lag effect has been identified; some statistical relationship exists between characteristics of populations living outside each target area and outcomes measured within each target area.

The inclusion of spatial lag effects within spatial statistical model often implies specific spatial dynamics among populations. For example, if population characteristics measured outside of target areas affect rates of problems within target areas, some population mixing must be taking place. A contact process must exist that relates characteristics measured in one place to outcomes measured in another. In some cases, these "contacts" may be very mundane and simply represent the movement of at-risk individuals from place to place. For example drinking drivers often crash their vehicles in places where they do not live. So population characteristics of one place (for example, sizable populations of young men) will be related to rates of car crashes elsewhere (for example, near bars and restaurants). In other cases, these "contacts" may be more complex, involving interactions between at-risk individuals living in different places (for example, young men), their use of different environments (for example, drinking at bars), and the unfortunate outcomes that sometimes arise as these populations mix (for example, violence). Thus, modeling spatial lag effects may be very suggestive of population interactions that support problem outcomes.

AN ILLUSTRATION OF A SPATIAL ANALYSIS: CHILD MALTREATMENT IN CENSUS TRACTS

Increasing numbers of neighborhood studies on child maltreatment have been conducted over the past 20 years (Coulton, Korbin, Su, & Chow, 1995; Garbarino & Sherman, 1980). Yet the majority of these did not use spatial statistical techniques to analyze the data. To show how spatial autocorrelation can bias results of studies of spatial areas, we report the results from OLS and GLS nuisance parameter regression models examining the relationship between neighborhood social organization and rates of child maltreatment. The current study uses the approach developed by Coulton and colleagues of creating factor scores to examine the effect of structural factors on child maltreatment rates in census tracts.

Method

This study used a cross-sectional ecological design to examine the relationship between social organization and rates of child maltreatment in census tracts in Alameda, Sacramento, and Santa Clara counties, California. A purposive sample was used because these counties showed variation between census tracts on overall rates of child maltreatment, ethnic and racial composition, poverty levels, and population size. For this study, census tracts (321 in Alameda County, 279 in Sacramento County, and

341 in Santa Clara County) were used to denote spatial areas. Adjacent spatial areas were defined as those census tracts that share a boundary. A connection matrix (941 x 941) was developed and used in the spatial analyses using the procedures described earlier and shown in Figure 2.

Dependent Variable. The dependent variable for the current study was rates of substantiated reports of child abuse and neglect per 1,000 children for the year 2000. The census tracts have on average 10.6 substantiated reports per 1,000 children (Table 1). These records were obtained from the Center for Social Services Research at the University of California, Berkeley, which is contracted by the California Department of Social Services to archive the official referral and placement child welfare data for the state (Needell et al., 2003). Although some children may be referred to child protective services more than once during any given year, for the purposes of this study, the most severe substantiated report was used for each child in the study. Therefore, the rates of abuse refer to the number of children abused, not the number of maltreatment events. Overall, 96% of all addresses were successfully assigned a geographic location, with geocoding rate of 96.4% of addresses in Santa Clara County, 96.2% in Sacramento, and 95.7% in Alameda.

Independent Variables. Independent variables on census tract characteristics were collected from the 2000 census data on measures of percentage unemployment, percentage poverty, percentage female-headed families, percentage African American, percentage Hispanic, percentage foreign born, percentage vacant housing units, percentage moved in the past five years, percentage recent (past year) movement, percentage moved in the past 10 years, percentage of population younger than 18, percentage of population age 65 and older, the ratio of adult men to adult women, and the ratio of children to adults. The selection of these variables was informed by Coulton and colleagues' (1995) and Sampson and colleagues' (1997) indices of social disorganization. These measures were reduced using principal components analysis of the correlation

Table 1: Descriptive Statistics for Dependent and Independent Variables for Census Tracts, by Factor Loading

Dependent variable		
No. of substantiated reports of maltreatment/1,000 children	10.6	18.3
Concentrated disadvantage		
% female-headed families with children	10.4	7.9
% population living in poverty	11.0	10.0
% unemployed residents	3.5	2.7
% welfare	4.8	5.4
% African American residents	10.7	14.6
Housing stress		
% population who moved between 1995 and 2000	50.2	12.8
% population moved in last year	21.2	10.3
% population moved since 1990	67.3	13.5
% vacant housing units	3.5	4.6
Immigrant populations/Child care burden		
Ratio children (0–12 years)/adults (≥ 21 years)	0.3	0.1
Ratio adult males (≥ 21 years)/adult females (≥ 21 years)	1.2	4.4
% children < 18	24.6	7.1
% adults ≥ 65 years	10.6	5.6
% Hispanic residents	19.0	15.5
% residents foreign born	25.4	14.4
Lagged variable		
% poverty	10.9	7.6

matrix using varimax factor rotations. This procedure resulted in three factors that explained 62.8% of the variance among these measures. These factors represent measures of (1) concentrated disadvantage (reflecting measures of poverty, female-headed families, unemployment, welfare use, and African American residents), (2) housing stress (reflecting measures of residential turnover and vacant housing units), and (3) immigrant populations/child care burden (reflecting measures of percentage foreign born, Hispanic, 18 and younger, 65 and older, and male-to-female ratio and child-to-adult ratio). Finally, a spatial lag for percentage poverty was included to examine how it may affect local rates of maltreatment.

Data Analysis. Data were first analyzed using an OLS regression. A second nuisance parameter GLS regression that assessed and controlled for spatial autocorrelation was then conducted to determine how the presence of spatial autocorrelation might change results. For this study, the presence of positive spatial autocorrelation indicated that census tracts with high rates of child abuse and neglect are adjacent to other high maltreatment census tracts. In terms of child maltreatment rates in census tracts, areas with low rates of maltreatment surround census tracts with high rates of maltreatment. To control for heteroskedasticity found in small area analyses, each model was weighted by the square root of the child population for that area (Greene, 2000).

Spatial Statistical System (S^3) version 4.32, a proprietary software of the Prevention Research Center, was used to conduct the analyses (Ponicki & Gruenewald, 2003). In the presence of statistically significant spatial autocorrelation, S^3 statistically controlled for this effect, giving less biased estimates of the association between the structural characteristics and rates of child maltreatment. To determine whether outliers existed, studentized t statistics for residuals were assessed (Cook & Weisburg, 1982). Analysis of residuals for the final models showed conditional normality of the dependent variable.

Results

Table 2 gives both the OLS regression model and the nuisance parameter GLS regression model that controls for spatial autocorrelation. The significance of the Moran coefficient (t = 31.51) indicates that there was significant positive spatial autocorrelation among the residuals of the OLS analysis. Figure 3 shows quintiles of the unstandardized residuals from this OLS regression. By looking at this figure, it is clear that errors are not distributed randomly across the spatial units but exhibit patterns related to positive spatial autocorrelation. Significant positive spatial autocorrelation results in higher rates of Type I errors, leading researchers to find nominally significant relationships when such relationships do not exist. For instance, factor 3 (immigrant populations/child care burden) is negatively related to rates of child maltreatment in the OLS regression, but the relationship is not significant in the spatial regression model.

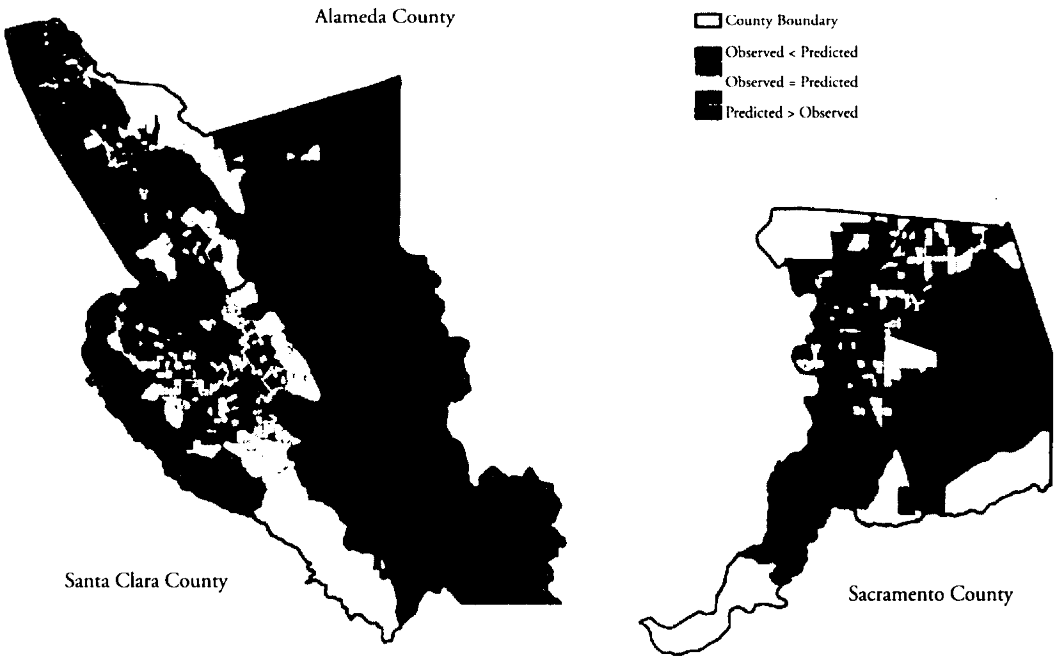
This analysis also found that concentrated disadvantage, housing stress, and percentage of poverty in adjacent areas were positively related to rates of

Table 2: OLS and GLS Spatial Regressions for Community Structural Factors, Lag of Poverty, and Child Maltreatment Rate (n = 941)

Constant	5.16	0.73***	7.85	0.94***
Factor 1 – Disadvantage	5.09	0.49***	6.56	0.43***
Factor 2 – Housing stress	1.58	0.33***	0.86	0.31***
Factor 3 – Immigrant populations/child care burden	–0.73	0.29*	–0.13	0.36
Lag of percent poverty	0.48	0.06***	0.21	0.07***
Pseudo-R^2			.689	
Spatial Autocorrelation	**Moran**	**t**	**ρ**	**t**
	.587	31.51	.663	20.34

Notes: OLS = ordinary least squares. GLS = generalized least squares.
*p < .05. **p < .01. ***p <.001.

Figure 3: Residuals from OLS Regression Examining the Relationship between Neighborhood Characteristics and Child Maltreatment



Note: OLS = ordinary least squares.

maltreatment. The correction for spatial autocorrelation in the GLS model corrected for biased statistical tests in the OLS analysis.

Discussion

The findings from OLS and GLS regression models differed significantly for one of the variables in the model. Thus, this illustration shows how spatial autocorrelation can produce measurement error, causing researchers to make inconsistent conclusions. Relying on this information to create interventions to reduce child maltreatment in neighborhood areas could result in developing ineffective strategies that fail to reduce maltreatment. For example, from the OLS analysis one might conclude that developing an intervention to reduce substantiated rates of child maltreatment should incorporate strategies reducing child care burden (for example more child care centers) in nonimmigrant communities. However, an intervention based on the GLS spatial regression does not find this factor to be significant so one would not expect to see a reduction in substantiated rates of child maltreatment using this strategy.

CONCLUSION

As social workers continue to ask questions that examine how location affects the prevalence of social problems, they need to keep pace with the strategies used to properly analyze the results. Neighborhoods, like individuals, are not isolated. They may affect and be affected by processes in nearby locations. These possibilities should be considered in many analyses of social data for the same reasons that the data should be checked for normality and homoskedasticity. Nonindependence among the study units or error terms can be a nuisance that simply biases regression coefficients and parameters and therefore must be controlled in a regression model, or it can be an interesting phenomenon in itself. Spatial dependence may represent a key explanatory process in answering the research question.

Even when no significant spatial autocorrelation is present, the descriptive and analytical procedures strengthen the model-building process by carefully investigating the data properties. In this instance, spatial techniques can be used as a diagnostic technique. If this diagnostic tool is applied and the results show that spatial autocorrelation is not present, confidence in study findings increases. When significant spatial autocorrelation is present, these techniques can improve the precision of models and enable social workers to develop interventions based on more scientifically sound results.

Advances in spatial statistical methodology are increasing rapidly. Although this article focuses on exploratory spatial data analysis and spatial regression procedures, procedures have been developed to conduct spatial Poisson, spatial time series cross-sectional, and spatial hierarchical models (see Banerjee, Carlin, & Gelfand, 2003; Gruenewald, Millar, Ponicki, & Brinkley, 2000). Each of these techniques is similar to spatial regression models in that they require the use of a connection or distance matrix to explicitly describe space, and they allow social workers greater flexibility in modeling the causes and consequences of social problems.

Continuing Challenges

Those interested in examining the role of location in social work problems (including the authors) continue to be confronted with several methodological and analysis issues. These include how to define small areas for analysis, limitations of current computer software, problems of inference related to the ecological fallacy, and heteroskedasticity of small areas.

Defining Small Areas. One difficulty in conducting analyses of small areas is determining the spatial scale of the problem. Researchers must often rely on administrative units (for example, zip codes, census block groups) to make use of the information collected in those areas, yet those units may not adequately represent the environment being studied. The geographic unit of study is important for several reasons. In a study examining the geographic distribution of wildland fires in California, Chou (1991) demonstrated that as the resolution of the map increased, the level of spatial autocorrelation between spatial units also increased. Simply, at higher resolutions more measures of common areas were taken (for example, more measures within common stands of trees), making data from adjacent units more similar than not. The consequence of this effect is to alter the apparent level of association between spatial units and change the outcomes of statistical models executed using differently defined units. This effect is related to another important source of specification bias in geospatial studies, the modifiable areal unit problem (MAUP). Techniques are currently being developed to deal with the MAUP, but are outside of the scope of this article. See Openshaw (1984) for an introduction.

In this case, differently defined area units (for example, census tracts rather than zip codes) gerrymander geographic data in different ways, resulting in very different statistical models of the same outcome. In the worst case, findings that may be statistically significant with one unit of analysis may not be significant with another. As an example, for the purposes of social work research, examining poverty rates for neighborhood areas that cover a few blocks may exhibit a much different spatial pattern than poverty rates for counties in a state. Counties generally cover larger distances than neighborhoods, so examining poverty at that level will not show pockets of poverty within the county where social workers may want to target services. Thus, geostatisticians and ecological researchers concerned with the characteristics of data collected from differently defined areas focus on replication of statistical models across units with different characteristic scales and resolution.

Software Limitations. Another major challenge to conducting spatial analysis is that computer software needed to run these analyses is not straightforward or instinctual to learn. Freeware computer software programs are available on the World Wide Web (for example, GeoDa [Anselin, 2003]), but there is often a steep learning curve to become proficient enough to use them. Even if researchers are capable of using these software packages, spatial analysis of data with a large number of spatial units uses large amounts of computer resources and can take significant amounts of time to run. As a result, individual researchers often do not have the computing resources to conduct large-scale spatial analyses to answer questions of interest.

Ecological Fallacy. An often-cited problem of small area studies is the mistake of attributing findings of aggregate-level data to individuals. Termed the ecological fallacy (Selvin, 1958), critics of ecological studies point to instances where individual behavior or risk differs when compared with estimates obtained using aggregate data. Although this is of concern to researchers, ecological studies can provide additional information not available at the individual level. For example, global measures such as population density or neighborhood social disorganization affect all residents living in a particular area, but the collectiveness of these measures may contribute to rates of problems (Morgenstern, 1998). Those conducting ecological studies and spatial analysis, although being careful not to overstate their findings, can point to information that is unobtainable at the individual level.

Heteroskedasticity. In addition to spatial autocorrelation, another challenge to analyses of small areas is that spatial units with small populations are given the same weight in the analysis as spatial units with large populations (see Greene, 2000). To control for heteroskedasticity, residence-based models can be weighted by the square root of the population for that area. In a cross-sectional analysis, appropriate weighting for heteroskedasticity provides an overall unbiased assessment of effects in statistical models.

Implications for Social Work Practice

As these capabilities for spatially analyzing data continue to expand, so does the usefulness for social work practice. Understanding the geographic context of social problems more fully enhances our ability to create interventions that address the individual and environmental determinants of these problems simultaneously. For example, new methods that examine social networks in a spatial context can begin to disentangle how various types of social support can differ across spatial areas, allowing practitioners to develop better informed intervention efforts aimed at strengthening or developing networks differently for different areas. These methods provide researchers and practitioners a greater opportunity to examine the relationship among accessibility, availability, and use of services across spatial areas, thus allowing us to better target service provision in areas that are most in need of and likely to use those services.

If neighborhoods matter (that is, if individual behaviors depend on the behaviors of those in surrounding areas or if features of the environment can accelerate problem outcomes), there are strong implications for policy development as well. For instance, policies designed to change the behavior of individuals, such as Welfare-to-Work or the Adoption and Safe Families Act of 1997, usually do not consider the role of location. They are usually not geographically targeted, despite the concentration of poverty and child maltreatment in particular neighborhood areas, and therefore apply across entire municipalities. Examining whether behaviors or phenomena of interest diffuse across neighborhoods in a spatially discriminate way to exhibit patterns across areas and contagion effects (spatial externalities) might provide insight on how to develop interventions in these geographically small areas. SWR

REFERENCES

Albert, V., & Catlin, S. N. (2002). Strategic interaction among the states: An in-depth look at the welfare "race to the bottom." *Social Work Research, 26,* 199–216.

Anselin, L. (2003). *GeoDa™ 0.9 user's guide.* Urbana: Spatial Analysis Laboratory, Department of Agricultural and Consumer Economics, University of Illinois, Urbana-Champaign and Center for Spatially Integrated Social Science.

Bailey, T. C., & Gatrell, A. C. (1995). *Interactive spatial data analysis.* Essex, England: Addison Wesley Longman.

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2003). *Hierarchical modeling and analysis for spatial data.* Boca Raton, FL: Chapman & Hall/CRC Press.

Burton, L. M., & Jarrett, R. L. (2000). In the mix, yet on the margins: The place of families in urban neighborhood and child development research. *Journal of Marriage and the Family, 62,* 1114–1135.

Chou, Y. H. (1991). Map resolution and spatial autocorrelation. *Geographical Analysis, 23,* 228–246.

Cliff, A. D., & Ord, J. K. (1973). *Spatial autocorrelation, monographs in spatial environmental systems analysis.* London: Pion Limited.

Cliff, A. D., & Ord, J. K. (1981). *Spatial processes models and applications.* London: Pion Limited.

Cook, R. D., & Weisburg, S. (1982). *Residuals and influence in regression.* New York: Chapman and Hall.

Coulton, C., Korbin, J. E., Su, M., & Chow, J. (1995). Community level factors and child maltreatment rates. *Child Development, 66,* 1262–1276.

Eck, J. E. (1995). A general model of the geography of illicit retail marketplaces. In J. E. Eck & D. Weisburd (Eds.), *Crime and place: Crime prevention studies* (Vol. 4). Washington, DC: Police Executive Research Forum.

Garbarino, J., & Sherman, D. (1980). High-risk neighborhoods and high-risk families: The human ecology of child maltreatment. *Child Development, 51,* 188–198.

Greene, W. H. (2000). *Econometric analysis.* Upper Saddle River, NJ: Prentice Hall.

Griffith, D. A. (1988). *Advanced spatial statistics.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Gruenewald, P. J., Millar, A., Ponicki, W. R., & Brinkley, G. (2000). Physical and economic access to alcohol: The application of geostatistical methods to small area analysis in community settings. In R. Wilson & M. DuFour (Eds.), *Epidemiology of alcohol problems in small geographic areas* (NIAAA Research Monograph No. 36, pp. 163–212). Bethesda, MD: U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Institute on Alcohol Abuse and Alcoholism.

Leventhal, T. L., & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence upon child and adolescent outcomes. *Psychological Bulletin, 126,* 309–337.

Morgenstern, H. (1998). Ecological studies. In K. J. Rothmann & S. Greenland (Eds.), *Modern epidemiology* (2nd ed., pp. 459–480). New York: Lippincott-Raven.
Needell, B., Webster, D., Cuccaro-Alamin, S., Armijo, M., Lee, S., Brookhart, A., Lery, B., Shaw, T., Dawson, W., Piccus, W., Magruder, J., & Kim, H. (2003). *Child welfare services reports for California*. Retrieved May 11, 2003, from http://cssr.berkeley.edu/CWSCMSreports/
Noble, M., & Smith, T. (1994). "Children in Need": Using geographical information systems to inform strategic planning for social service provision. *Children & Society, 8*, 360–376.
Openshaw, S. (1984). *The modifiable areal unit problem*. Norwich, England: Geo Books.
Ponicki, W. R., & Gruenewald, P. J. (2003). *S³: Spatial Statistical System user's guide*. Version 4.32. Berkeley, CA: Prevention Research Center.
Queralt, M., & Witte, A. D. (1998). A map for you? Geographic information systems in the social services. *Social Work, 43*, 455–469.
Robertson, J. G., & Wier, K. R. (1998). Using geographical information systems to enhance community-based child welfare services. *Child Maltreatment, 3*, 224–234.
Sampson, R. J., Morenoff, J. D., & Gannon-Rowley, T. (2002). Assessing "neighborhood effects": Social processes and new directions in research. *Annual Review of Sociology, 28*, 443–478.
Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science, 277*, 918–924.
Saxe, L., Kadushin, C., Beveridge, A., Livert, D., Tighe, E., Rindskopf, D., Ford, J., & Brodsky, A. (2001). The visibility of illicit drugs: Implications for community-based drug control strategies. *American Journal of Public Health, 91*, 1987–1994.
Selvin, H. (1958). Durkheim's suicide and problems of empirical research. *American Journal of Sociology, 63*, 607–619.

Bridget Freisthler, PhD, *is assistant professor, Department of Social Welfare, UCLA School of Public Affairs, 3250 Public Policy Building, Box 951656, Los Angeles, CA 90095-1656; e-mail: freisthler@spa.ucla.edu.* ***Bridgette Lery, PhD,*** *is a researcher, Chapin Hall Center for Children, University of Chicago.* ***Paul J. Gruenewald, PhD,*** *is scientific director and senior research scientist, Prevention Research Center, Pacific Institute for Research and Evaluation, Berkeley, CA.* ***Julian Chow, PhD,*** *is associate professor, School of Social Welfare, University of California at Berkeley.* *Preparation of this manuscript was made possible in part by the National Institute on Alcohol Abuse and Alcoholism Center grant number P60-AA06282 (Gruenewald, PI). The authors wish to thank Barbara Needell, California Department of Social Services, and the Stuart Foundation for making this research possible.*

Original manuscript received November 30, 2004
Final revision received August 10, 2005
Accepted March 30, 2006

NASW PRESS POLICY ON ETHICAL BEHAVIOR

The NASW Press expects authors to adhere to ethical standards for scholarship as articulated in the NASW *Code of Ethics* and *Writing for the NASW Press: Information for Authors.* These standards include actions such as

- taking responsibility and credit only for work they have actually performed
- honestly acknowledging the work of others
- submitting only original work to journals
- fully documenting their own and others' related work.

If possible breaches of ethical standards have been identified at the review or publication process, the NASW Press may notify the author and bring the ethics issue to the attention of the appropriate professional body or other authority. Peer review confidentiality will not apply where there is evidence of plagiarism.

As reviewed and revised by
NASW National Committee on
Inquiry (NCOI), May 30, 1997

Approved by NASW Board of
Directors, September 1997