

THESIS
3
2010

LIBRARY
Michigan State
University

This is to certify that the
dissertation entitled

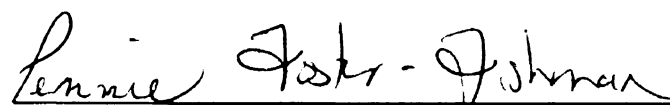
USING GEOSTATISTICAL MODELS TO STUDY NEIGHBORHOOD
EFFECTS: AN ALTERNATIVE TO USING HIERARCHICAL LINEAR
MODELS

presented by

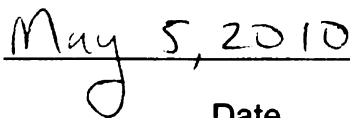
Steven James Pierce

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Psychology



Major Professor's Signature



Date

MSU is an Affirmative Action/Equal Opportunity Employer

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

**USING GEOSTATISTICAL MODELS TO STUDY NEIGHBORHOOD
EFFECTS: AN ALTERNATIVE TO USING HIERARCHICAL LINEAR
MODELS**

By

Steven James Pierce

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Psychology

2010

ABSTRACT

USING GEOSTATISTICAL MODELS TO STUDY NEIGHBORHOOD EFFECTS: AN ALTERNATIVE TO USING HIERARCHICAL LINEAR MODELS

By

Steven James Pierce

Neighborhoods are important ecological contexts that influence the development, behavior, health, and welfare of their residents. Community psychologists studying neighborhood effects usually turn to hierarchical linear modeling (HLM) to test multilevel theories that explain neighborhood effects by examining the links between neighborhood characteristics and resident outcomes.

Geostatistical modeling (GSM) can also test such theories, but it relies on a different way of conceptualizing neighborhoods than used in HLM and few social scientists have ever applied this method. This study developed an argument for why GSM may be a valuable alternative to HLM, then applied both methods to study the effects of neighborhood crime and neighborhood socioeconomic status (NSES) on residents' perceptions of neighborhood problems. Applying them to the same data allowed the study to examine the effect of varying the neighborhood boundaries used to measure crime and NSES and to explore whether the conceptual and statistical differences between HLM and GSM led to different scientific inferences about crime and NSES effects on residents' perceptions.

While HLM and GSM models detected similar amounts of neighborhood-level variance and autocorrelation in perceived neighborhood problems, GSM provided a better description of the data from this sample because crucial HLM assumptions about the independence of the residuals were violated. The specific neighborhood boundaries

used to measure crime and NSES in this study had important implications for the size and statistical significance of their effects.

For this sample, GSM showed that circular buffers centered on residents' homes provided better operational definitions of the neighborhoods than the fixed cluster boundaries required by HLM. The HLM models overestimated the size and significance of the NSES effect on perceived neighborhood problems due to inaccurate assumptions about the residuals at both levels of analysis. The GSM models did not suffer problems with their residuals and showed that while a cluster-based NSES measure did not affect residents' perceptions in these data, NSES measured in 0.2 km radius buffers around residents' homes did (but not as strongly as indicated by the HLM models).

The GSM models showed that residents' perceptions of neighborhood problems were more sensitive to crime occurring inside 1.1 km radius buffers around their homes than they were to the level of crime occurring inside the much smaller neighborhood cluster boundaries used in the HLM models. Thus, HLM underestimated how strongly crime affected residents' perceptions in this study because crime was not measured on the right spatial scale, despite following "best-practice" advice from the HLM literature to choose the smallest neighborhood units that are feasible.

The study concludes by discussing the implications of the findings for conceptualizing and operationally defining neighborhoods, measuring neighborhood-level constructs, and applying research findings to inform community intervention efforts. Future directions for research are suggested, as are some ways of dealing with the practical issues of using GSM.

Copyright by

STEVEN JAMES PIERCE

2010

DEDICATION

This work is dedicated to my parents, Rex and Lynn Pierce. My achievements rest on the foundation that they laid. Because they wrought well, I can reach the sky.

ACKNOWLEDGEMENTS

This work would not have been possible without the efforts of the William K. Kellogg Foundation and the residents of Battle Creek, Michigan. In committing to the *Yes we can!* initiative, the foundation sought to give back to its own home community. The residents responded by vigorously engaging with *Yes we can!* in many different ways, including participating in the survey that provided the outcome data for this study. I sincerely hope that the seeds planted in their collective effort to build a stronger, more vibrant community that offers all of its residents the opportunities and resources necessary to thrive will ultimately bear bountiful fruit.

On the personal side, I want to thank my wife Lori Corteville, whose unfailing support and encouragement kept me going through the rough patches in graduate school. She patiently shouldered more tasks at home so that I could devote the necessary time to my academic work. Words cannot fully express my gratitude to her, but I'm sure it will become apparent to her in other ways as we move into the next stage of our life together and implement the plans that have been waiting on my graduation.

I am thankful for the guidance offered by my advisor, Pennie Foster-Fishman, and the rest of my doctoral committee. Her questions at various points along the way led me to more clearly articulate the theoretical and conceptual basis of my work. The spatial data analysis class I took from Ashton Shortridge prompted me to apply those methods in community psychology, plus it introduced me to statistical computing tools that have proven incredibly valuable both in this study and in my work at CSTAT. Robin Miller emphasized the importance of connecting the statistical methods and results back to

substantive inferences and issues relevant to community psychologists. Finally, Rick DeShon challenged me to do a more sophisticated comparison of the two statistical methods than I had initially planned. All of these things contributed greatly to this work.

Thanks are also due to Brian Silver, who provided me with all the support and flexibility I could have asked for in my job at CSTAT while I worked on finishing this dissertation. Additional thanks are due to Erik Segur at CSTAT for providing the technology infrastructure that made it possible to actually run the analyses, and to all my other colleagues at CSTAT for working so hard to keep the consulting service running smoothly while I've been splitting my time between my job and my dissertation.

In addition to enthusiastically cheering me on, Melissa Quon Huber helped me obtain some of the required GIS data files. Without her, that would have been harder and likely would have taken longer. I deeply appreciate her support and help. My fellow students in the Ecological-Community Psychology program, who are too numerous to name individually, offered invaluable social support. Hearing that so many people found my work interesting and exciting was a great motivator. My friend Sara Repen's talent for sending encouraging cards at just the right time also deserves special recognition.

Finally, I would like to thank my former boss and mentor, Greg Cline, who nudged me back into graduate school to pursue my doctorate. As he promised, doing that opened up career options that would have been otherwise unobtainable. Given how we celebrated Greg's own dissertation defense a few years ago, I fully suspect he will return the favor if given even half a chance. Fortunately, I am prepared for such shenanigans!

TABLE OF CONTENTS

LIST OF TABLES	xiii
LIST OF FIGURES	xiv
INTRODUCTION	1
Using HLM to Study Neighborhood Effects.....	4
Neighborhoods as places in discontinuous space.	5
Problems with neighborhoods as places in discontinuous space.	8
Lack of meaningful boundaries.	9
Ignoring spatial proximity.....	11
Ignoring spatial variability in contextual conditions.	12
Poor handling of spatial scale.....	12
Summary.....	14
Using GIS to Study Neighborhood Effects.....	15
Accounting for spatial proximity with GIS.	16
Measuring neighborhood characteristics with GIS.	18
Summary.....	21
The Current Study: Comparing GSM and HLM.....	22
Focal constructs.....	23
Research questions.	24
Organization of the text.	29
LITERATURE REVIEW	31
Neighborhoods and Multilevel Research.....	31
Neighborhoods as meaningful contexts for residents.	31
Theoretical mechanisms underlying neighborhood effects.	32
Multilevel assumptions in neighborhood research.....	36
Defining Neighborhoods.....	37
Defining neighborhoods: Social versus geographical units.....	39
Neighborhoods as social units.	40
Neighborhoods as geographic units.....	40
Relating places to geographic space.	41
Neighborhoods as places in discontinuous geographic space.	42
Census geography.....	43
The MAUP.	44
Lack of meaningful boundaries.	45
Ignoring spatial proximity.....	48
Ignoring spatial variability in contextual conditions.	50
Poor handling of spatial scale issues.....	50
Summary.	53
Neighborhoods as places in continuous geographic space.....	54
Conceptual definition of neighborhood.	56
Making boundaries more meaningful.	56
Using buffers as neighborhood boundaries.....	58

Addressing spatial proximity.....	60
Addressing spatial variation in contextual conditions.	60
Addressing issues of spatial scale.....	61
Summary.	61
Hierarchical Linear Modeling Methods for Testing Contextual Effects	62
Why HLM is useful.....	64
The HLM statistical model.	66
Assumptions in HLM.	69
Autocorrelation in HLM.	70
Controlling for composition.....	72
Neighborhoods as level 2 units in HLM.....	73
Considering space in HLM.	75
Geostatistical Modeling Methods for Testing Contextual Effects	79
Why GSM is useful.	81
The GSM statistical model.	83
Assumptions in GSM.	84
Autocorrelation in GSM.	86
Controlling for composition.....	91
Neighborhoods as level 2 units in GSM.....	92
Considering space in GSM.	94
Comparing HLM and GSM Approaches	95
Comparing models of infectious disease in Haiti.	97
Comparing models of health care utilization in France.	98
Comparing models of substance abuse disorders in Sweden.	99
Summary.....	101
Background on the Substantive Constructs.....	102
Selection of constructs.....	102
Theoretical mechanisms.	104
Perceived neighborhood problems.	105
Crime.	109
Neighborhood SES.....	114
Individual-level predictors.....	119
Linking Gaps in the Literature to Hypotheses for the Present Study	121
Taking full advantage of spatial information.....	121
Detecting autocorrelation.....	121
Modeling autocorrelation.....	124
Testing assumptions by examining residuals.....	125
Testing contextual effects.	127
Examining spatial scale.	129
Alternative possibilities.	131
Summary of the Study	133
Limitations	135
METHOD	137
Study Context.....	138
Data Sources.....	140
Survey sample.	140

Crime data.....	143
Property data.....	144
Procedures.....	145
Survey consent.....	145
Geocoding.....	146
Surveys.....	146
Crime data.....	147
Property data.....	147
Neighborhood-Level Contextual Measures.....	147
Crime.....	148
Neighborhood SES.....	150
Individual-Level Measures.....	150
Neighborhood problems.....	151
Age.....	151
Gender.....	151
Race.....	151
Marital status.....	152
Education.....	152
Employment status.....	152
Income.....	152
Home ownership.....	152
Presence of children in the home.....	152
Analysis.....	153
Imputation of missing data.....	153
Bayesian modeling.....	154
Model building sequence.....	159
Testing Hypothesis 1.....	161
Testing Hypothesis 2.....	161
Testing Hypothesis 3.....	162
Testing Hypothesis 4.....	163
Testing Hypothesis 5.....	164
Testing Hypothesis 6.....	164
Testing Hypothesis 7.....	164
Testing Hypothesis 8.....	166
Criteria for evaluating and comparing models.....	167
Software.....	169
RESULTS.....	170
Exploratory and Descriptive Analyses.....	170
HLM and GSM Analyses.....	174
Research Question 1.....	174
Hypothesis 1.....	175
Hypothesis 2.....	183
Research Question 2.....	184
Hypothesis 3.....	184
Hypothesis 4.....	186
Hypothesis 5.....	188

Hypothesis 6.....	188
Research Question 3.....	191
Optimal buffer size for crime.....	191
Optimal buffer size for NSES.....	193
Hypothesis 7.....	193
Adding crime alone.....	195
Adding NSES alone.....	202
Adding both crime and NSES.....	204
Comparing CAR HLM to standard HLM.....	206
Comparing CAR HLM to cluster-based GSM.....	207
Comparing CAR HLM to buffer-based GSM.....	208
Research Question 4.....	209
DISCUSSION.....	211
Detecting Neighborhood-Level Variability in Perceived Neighborhood Problems	
.....	214
Previous research.....	215
Current findings.....	217
Spatial Scale of Autocorrelation for Perceived Neighborhood Problems.....	219
Spatial scale in HLM.....	219
Spatial scale in GSM.....	221
Modeling Autocorrelation: Spatial Versus Hierarchical Structure.....	223
Comparing model fit.....	224
Diagnostic analyses of HLM residuals.....	225
Diagnostic analyses of GSM residuals.....	228
Summary.....	230
Testing Crime and NSES Effects.....	231
Conceptualizing and defining neighborhoods.....	232
Comparing HLM and GSM.....	234
Current findings.....	235
Comparing model fit.....	235
Spatial scale of crime and NSES.....	236
Contextual effect of NSES.....	236
Contextual effect of crime.....	239
Implications for Defining Neighborhoods.....	240
Implications for Community Interventions.....	242
Crime effect.....	243
NSES effect.....	244
Summary.....	245
Feasibility of Applying GSM in Community Psychology Research.....	246
Limitations.....	247
Directions for Future Research.....	249
Use simulation studies to compare HLM and GSM.....	250
Explore alternative methods for defining buffers.....	250
Conclusion.....	251
APPENDIX.....	253

REFERENCES 257

12

13

12

12

12

三

25

124

25

23

7

LIST OF TABLES

Table 1: Conceptual comparison of HLM and GSM.....	96
Table 2: Demographic characteristics of survey participants (N = 1049)	142
Table 3: Overview of primary model building sequence.....	160
Table 4: Criteria for evaluating and comparing models	167
Table 5: Parameter estimates and model fit statistics for HLM Models 1-6.	176
Table 6: Parameter estimates and model fit statistics for GSM Models 1-5 and 56-58.....	179
Table 7: Parameter estimates and model fit statistics for empty GSM models fit to individual-level residuals from HLM Models 1 and 2.....	186
Table 8: Parameter estimates and model fit statistics for empty HLM models fit to individual-level residuals from GSM Models 1 and 2.....	188
Table 9: Parameter estimates and model fit statistics for empty HLM models fit to neighborhood-level residuals from GSM Models 1 and 2.....	190
Table 10: List of research questions and hypotheses.....	213
Table 11: Amount of change required on each predictor to reduce mean perceived problems by half a standard deviation (0.74 Points).....	243
Table 12: Parameter estimates and model fit statistics for GSM Models 15-17 and 31-33.	254

LIST OF FIGURES

- Figure 1: Hierarchical nesting of blocks and block groups within three census tracts in Battle Creek, Michigan. The two larger tracts are subdivided into multiple block groups, but the smallest tract contains only one block group; all block groups are further subdivided into blocks. Source: Map produced by the author from GIS files prepared by the U.S. Census Bureau (2007a, 2007b, 2007c). 7**
- Figure 2: Illustration of using buffers as neighborhood boundaries. Contextual characteristics could be measured within circular buffers of different sizes (e.g., with radii of 100 m and 200 m) centered on points A and B, which are shown in relation to census tract and block boundaries in Battle Creek, MI. For point A, both buffers overlap portions of multiple tracts. For point B, only the larger buffer overlaps portions of more than one tract. The 200 m buffers for these two points partially overlap. Source: Map produced by the author from GIS files prepared by the U.S. Census Bureau (U.S. Census Bureau, 2007a, 2007c). 59**
- Figure 3: Exponential, spherical, and Gaussian variogram models displayed in three different metrics. Each panel illustrates the autocorrelation between observations as a function of the distance between them. The top two rows show these models in correlation (top) and covariance (middle) metrics, which are measures of similarity; the bottom row shows them in a semivariance metric, which is a measure of dissimilarity. All three models have the same partial sill ($\sigma^2 = 1$), range ($\phi = 1$), and nugget ($\tau^2 = 1$) parameters, but differ in shape. 88**
- Figure 4: Map of neighborhood cluster boundaries (N = 52) and ESCA boundaries. These clusters were used as Level 2 units during the survey sampling and to represent ecologically meaningful neighborhoods for grouping residents in the HLM analyses. Source: Map prepared by the author..... 139**
- Figure 5: Distribution of the imputed neighborhood problems scores (N = 1049). The vertical line is located at the mean (M = 3.77, SD = 1.48); the circles at the bottom show individual data points (vertically jittered to reduce overlap). 171**
- Figure 6: Boxplots of neighborhood problems scores for each cluster, in ascending order by cluster-level mean score. The dots show the clusters' medians. 171**
- Figure 7: Map of the imputed neighborhood problems scores (N = 1,049). 172**
- Figure 8: Distributions of pairwise distances between cluster centroids (N = 52) and between survey locations (N = 1,049). The vertical lines mark the medians, which are almost identical (Mdn = 2.91 and Mdn = 2.95, respectively). 174**

Figure 9: Estimated variance components and levels of autocorrelation from HLM and GSM Models 1 and 2. The plot shows the posterior means (symbols) plus the central 95% credible intervals (whiskers) around those estimates. Model 1 for each method was an empty model that included no predictors, while Model 2 included all individual-level predictors..... 183

Figure 10: Exponential variograms and correlation functions for GSM Models 1 and 2. The maximum value for the autocorrelation is the PSR for each model (.288 and .283, respectively). Vertical lines mark the practical ranges of spatial autocorrelation for the two models (2,962 m and 3,058 m, respectively), which occur at the distances where the neighborhood-level covariances have decreased to 5% of their initial sizes, leaving little residual autocorrelation..... 184

Figure 11: Exponential variograms and correlation functions for empty GSM models fit to the individual-level residuals for HLM Models 1 and 2. The maximum value for the autocorrelation is the PSR for each model (.441 and .425, respectively). Vertical lines mark the practical ranges of spatial autocorrelation for the two models (4.4 m and 4.5 m, respectively), which occur at the distances where the neighborhood-level covariances have decreased to 5% of their initial sizes, leaving little residual autocorrelation. 187

Figure 12: Estimated variance components and levels of autocorrelation from empty HLM and GSM models fit to the individual-level residuals from GSM and HLM Models 1 and 2. The plot shows the posterior means (symbols) plus the central 95% credible intervals (whiskers) around those estimates. Predictors included refers here to the predictors included in the model that generated the residuals being analyzed: Model 1 for each method was an empty model that included no predictors, while Model 2 included all individual-level predictors. ICC = intra-class correlation; L1 = level 1 (individual); PSR = partial sill ratio.... 189

Figure 13: Parameter estimates and model fit criteria for GSM Models 6-30, shown as a function of buffer radius. The dashed, vertical line marks the optimal buffer radius (1.1 km) for measuring crime. Crime = coefficient for the buffer-based crime measure; DIC = deviance information criterion; L2 PCV = level 2 proportional change in variance relative to GSM Model 1 (level-specific R^2); PSR = partial sill ratio; R^2 = overall proportion of variance explained; Range = practical range (in km) of variogram; t = t-statistic for the crime coefficient. These models also included all individual-level predictors. 192

Figure 14: Parameter estimates and model fit criteria for GSM Models 31-55, shown as a function of buffer radius. The dashed, vertical line marks the optimal buffer radius (0.2 km) for measuring NSES. NSES = coefficient for the buffer-based measure of neighborhood socioeconomic status; DIC = deviance information criterion; L2 PCV = level 2 proportional change in variance relative to GSM Model 1 (level-specific R^2); PSR = partial sill ratio; R^2 = overall proportion of variance explained; Range = practical range (in km) of variogram; t = t-statistic for the NSES coefficient. These models also included all individual-level predictors. 194

Figure 15: Model fit criteria for HLM Models 3-6 and GSM Models 3-5 and GSM Models 56-58. Both = both crime and NSES; CAR = conditional autoregressive model at level 2 with both crime and NSES as predictors; DIC = deviance information criterion; L2 PCV = level 2 proportional change in variance relative to GSM Model 1 (level-specific R^2); R^2 = overall proportion of variance explained. These models also included all individual-level predictors.... 196

Figure 16: Estimated variance components and levels of autocorrelation from HLM Models 3-6, and GSM Models 3-5 and 56-58. The plot shows the posterior means (symbols) plus the central 95% credible intervals (whiskers) around those estimates. Both = both crime and NSES; CAR = conditional autoregressive model at level 2 with both crime and NSES as predictors. These models also included all individual-level predictors..... 197

Figure 17: Exponential variograms and correlation functions for GSM Models 3-5 (top) and GSM Models 56-58 (bottom). The maximum value for the autocorrelation is the PSR for each model (for Models 3-5, PSR = .246, .247, and .217 respectively; for Models 56-58, PSR = .151, .208, and .138, respectively). Vertical lines mark the practical ranges of spatial autocorrelation for the models, which occur at the distances where the neighborhood-level covariances have decreased to 5% of their initial sizes, leaving little residual autocorrelation. For Models 3-5, range = 2,647 m, 2,451 m, and 1,964 m respectively; for Models 56-58, range = 765 m, 1840 m, and 505 m, respectively. 198

Figure 18: Estimated crime coefficients and t-statistics for HLM Models 3, 5, and 6 and GSM Models 3, 5, 56, and 58. The plot shows the posterior means (symbols) plus the central 95% credible intervals (whiskers) around the estimated coefficients. Credible intervals intersected by the dashed reference line at zero indicate non-significant effects. Both = both crime and NSES were included; CAR = conditional autoregressive model at level 2 with both crime and NSES as predictors. These models included all individual-level predictors.... 199

Figure 19: Estimated NSES coefficients and t-statistics for HLM Models 4, 5, and 6 and GSM Models 4, 5, 57, and 58. The plot shows the posterior means (symbols) plus the central 95% credible intervals (whiskers) around the estimated coefficients. Credible intervals intersected by the dashed reference line at zero indicate non-significant effects. Both = both crime and NSES were included; CAR = conditional autoregressive model at level 2 with both crime and NSES as predictors. These models included all individual-level predictors.... 200

Figure 20: Scatterplot showing buffer area (km^2) as a function of buffer radius (km), with annotations showing the optimal buffer sizes for crime and NSES. The dashed horizontal reference lines show the minimum (0.03 km^2) and maximum (0.47 km^2) areas among the 52 clusters. The mean cluster area was 0.08 km^2 . Arrows show the optimal buffer sizes for crime (1.1 km radius, 3.80 km^2), and NSES (0.2 km radius, 0.13 km^2)..... 210

USING GEOSTATISTICAL MODELS TO STUDY NEIGHBORHOOD EFFECTS: AN ALTERNATIVE TO USING HIERARCHICAL LINEAR MODELS

INTRODUCTION

Neighborhoods are potent ecological contexts that influence the development, behavior, health, and welfare of residents in a variety of ways (Gephart, 1997; Leventhal & Brooks-Gunn, 2000; Sampson, Morenoff, & Gannon-Rowley, 2002; Shinn & Toohey, 2003). As a result, many social scientists have developed multilevel theories linking neighborhood contextual conditions to outcomes for residents. Two recent papers illustrate such theories. Roosa, Jones, Tein, and Cree (2003) proposed a theoretical model where neighborhood characteristics influence residents' perceptions and experiences; those, in cascading fashion, then affect the ways that families and children react to neighborhood conditions, which in turn mediate influence on outcomes for children. Drawing on environmental stress and social disorganization theories, Kruger, Reischl, and Gee (2007) found support for their hypothesis that neighborhood physical deterioration would exert indirect effects on depression and stress through its impact on residents' social behaviors and their perceptions of social conditions in their neighborhoods. Both papers describe multilevel theories because they propose that neighborhood-level constructs represent contextual conditions that have consequences for individual-level outcomes; they postulate the existence of cross-level effects (Shinn & Rapkin, 2000) originating from the neighborhood level that, with appropriate research design and analytical methods, can be empirically tested.

Arguing that multilevel theories demand analytical methods that better capture, represent, and answer questions about context, Luke (2005) recommended that

community psychologists testing these theories increase their use of hierarchical linear modeling (HLM) and geographic information systems (GIS). HLM is an extension of regression that is designed to handle grouped data in which the intercept and/or slope coefficients are allowed to vary from group to group (Gelman & Hill, 2007; Raudenbush & Bryk, 2002; Shinn & Rapkin, 2000). Researchers can apply HLM in neighborhood research by grouping residents according to the neighborhoods in which they live.

GIS generally refers to software used for capturing, storing, and managing spatially-referenced data and then displaying those data with maps (Haining, 2003), but it also includes methods for spatial statistical analysis (Bailey & Gatrell, 1995; Haining, 2003; Luke, 2005). Because neighborhoods can be viewed as geographic places (Burton, Price-Spratlen, & Spencer, 1997; Coulton, Cook, & Irwin, 2004; Diez Roux, 2001; Gephart, 1997; Guest & Lee, 1984; Lee, 2001; Lee & Campbell, 1997; Leventhal & Brooks-Gunn, 2000; Sampson, et al., 2002) GIS statistical methods may also be applicable to neighborhood research.

Both HLM and GIS methods can link contextual characteristics of neighborhoods to outcomes among residents in ways that are consistent with multilevel theories and both also offer potential solutions to the statistical problem posed by the lack of independence among observations¹ in the samples needed to test theories about neighborhood effects (Haining, 2003; Raudenbush & Bryk, 2002). When the values of the same variable from different observations in a dataset are not independent from each other, they are said to be autocorrelated. The presence of autocorrelation simply means there is some structured relationship between the values of that variable associated with different observations.

¹ Violating the independence assumption in a regression model inflates the Type I error rate for statistical inferences about the regression coefficients because the standard errors are underestimated.

From a statistical perspective, the idea that neighborhoods influence residents implies that they somehow induce whatever autocorrelation exists in resident outcomes. Therefore, it is necessary to find an appropriate way to describe that autocorrelation and to build statistical models that can explain it. The two types of autocorrelation relevant to this study—hierarchical autocorrelation (associated with HLM) and spatial autocorrelation (associated with GIS methods)—are alternative ways of describing and modeling the structures that could be observed in a dataset if neighborhood effects are present. So, both of the methods advocated by Luke (2005) can be employed in neighborhood research, but they have not been applied with equal frequency in our discipline and have rarely been compared. HLM has been the tool of choice for quantitative studies of neighborhood effects recently, leaving GIS methods underutilized.

This study identifies some limitations associated with using HLM to study neighborhood effects, then develops an argument for why a specific GIS method called geostatistical modeling (GSM) (Chaix, et al., 2006; Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian, Lynch, & Chauvin, 2005) may be a valuable alternative to HLM. GSM is an extension of regression designed for analyzing relationships between variables when the spatial locations of the observations are known. In HLM, geographic space is conceptualized as a discontinuous phenomenon that is divided into discrete neighborhood units with fixed boundaries, while in GSM geographic space is conceptualized as a continuous phenomenon in which neighborhoods are more loosely defined as the areas immediately surrounding particular locations. This study compared the results of applying both HLM and GSM methods to a single set of data to examine whether the conceptual differences between HLM and GSM led to differences in their

statistical performance and in the scientific inferences they allow us to make about the phenomena under study that warrant further use of GSM.

Using HLM to Study Neighborhood Effects

Until now, most community psychologists have relied on HLM for answering questions about contextual effects, perhaps because its terminology maps directly onto the levels of analysis in our theories (Luke, 2005) and it provides the flexibility to test a wide range of multilevel hypotheses (Merlo, 2003; Merlo, et al., 2006; Merlo, Chaix, Yang, Lynch, & Råstam, 2005a, 2005b; Merlo, Yang, Chaix, Lynch, & Råstam, 2005). The consensus, both within community psychology and in other social sciences, is that HLM represents a substantial improvement over using ordinary least squares (OLS) regression models to study contextual effects (Bingenheimer & Raudenbush, 2004; C. Duncan, Jones, & Moon, 1998; Hofmann, Griffin, & Gavin, 2000; Roosa, et al., 2003). One of the reasons for that viewpoint is that HLM explicitly models autocorrelation in the data rather than assuming independence among the observations, which allows HLM to control Type I error rates better than OLS regression.

In HLM studies of neighborhood effects, one of the key premises is that the outcomes for different residents who belong to the same neighborhood are autocorrelated because something about living in that specific neighborhood actually induces similarity in the residents' outcomes. For example, living in a high-crime neighborhood might cause higher levels of fear among residents than living in a low-crime neighborhood. Because the residents are hierarchically nested within neighborhoods in HLM, we can be more specific and say that this method assumes a hierarchical autocorrelation structure: Part of each person's score on the outcome variable is assumed to be a shared residual

component that reflects the influence living in that specific neighborhood has on its residents.

The popularity of HLM among community psychologists for quantitative studies of neighborhood effects is quite evident in the literature. Recent papers in the *American Journal of Community Psychology* have used HLM to study neighborhood effects on children's behavior problems and cognitive development (Beyers, Bates, Pettit, & Dodge, 2003; Caughy, Nettles, & O'Campo, 2008; Caughy & O'Campo, 2006), residents' perceptions of collective efficacy (T. E. Duncan, Duncan, Okut, Strycker, & Hix-Small, 2003), and use of illicit drugs among low-income women (Sunder, Grady, & Wu, 2007). These and other researchers have used HLM to pursue questions about how much and in what ways neighborhoods matter by treating them as ecological settings that occupy geographic places with fixed, non-overlapping boundaries and possess contextual characteristics reflecting local conditions inside those boundaries. For example, Sampson and Raudenbush (2004) aggregated police, census, and observational data to create contextual measures of crime, poverty, disorder, and other neighborhood conditions for the block groups they used as neighborhoods in their HLM analyses.

Neighborhoods as places in discontinuous space. In focusing on neighborhoods as geographic places, HLM studies ask how much resident outcomes vary from place to place and what neighborhood characteristics predict that spatial variation in outcomes. To answer those questions, most HLM studies operationalize neighborhoods with administratively defined geographic units such as census tracts or block groups (Leventhal & Brooks-Gunn, 2000; Roosa et al., 2003; Sampson et al., 2002). The U.S. Census Bureau divides the nation into a hierarchically organized set of small geographic

units (see Figure 1) to facilitate the collection and tabulation of census data (U.S. Census Bureau, 1994, 2002). Blocks are the smallest units in that boundary system, while block groups are somewhat larger because each one is composed of multiple blocks and census tracts are still larger units (each tract often encompasses multiple block groups). Tracts are the units most routinely used for reporting census data.

Studies that use block groups or census tracts to operationalize neighborhoods inherit a boundary system in which space is divided into units that occupy mutually exclusive geographic areas (U.S. Census Bureau, 1994, 2002). That makes it easy to unambiguously group residents into neighborhoods based on the locations of their homes, which is crucial to ensuring that each resident is associated with only one neighborhood.² That hierarchical nesting structure allows HLM to treat outcomes among residents of the same neighborhood as correlated with each other but uncorrelated with the outcomes of all other residents. Thus, this application of HLM relies on a discontinuous conceptualization of geographic space that is fragmented into discrete, place-based geographic units that do not overlap in order to group residents. Taking that view of space facilitates detecting spatial variation in outcomes by modeling it as autocorrelation that is a function of whether or not residents live in the same place.

² If there are multiple levels of geographic units (e.g., block groups at level 2 and census tracts at level 3), then the goal is to ensure that each resident is associated with only one geographic unit at each level.

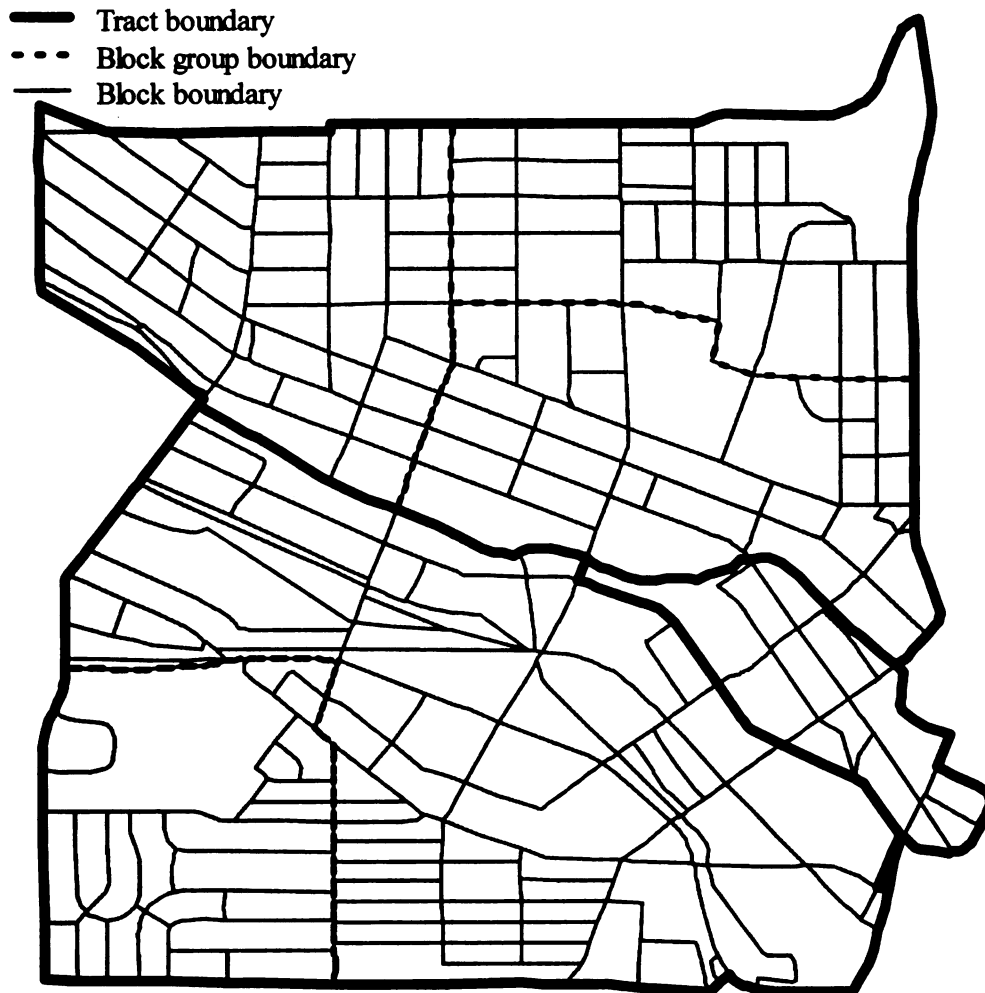


Figure 1: Hierarchical nesting of blocks and block groups within three census tracts in Battle Creek, Michigan. The two larger tracts are subdivided into multiple block groups, but the smallest tract contains only one block group; all block groups are further subdivided into blocks. Source: Map produced by the author from GIS files prepared by the U.S. Census Bureau (2007a, 2007b, 2007c).

The conceptualization of geographic space in neighborhood research also informs decisions about measuring neighborhood characteristics. Viewing a neighborhood as a discrete geographic unit that cannot overlap with other neighborhood units demands that

the boundary separating it from other neighborhoods be meaningful: people, objects, and events inside the boundary are part of that setting, but anything outside the boundary is part of a different neighborhood setting. So, the same boundary that groups residents into a neighborhood should also be used to define the geographic area within which all neighborhood-level characteristics should be measured, regardless of whether one is aggregating survey data to measure collective efficacy (Sampson, Raudenbush, & Earls, 1997), police data to estimate crime rates (Sampson & Raudenbush, 2004), census data to measure neighborhood poverty and racial composition (Franzini, Caughy, Nettles, & O'Campo, 2008), or measuring some other characteristic of the neighborhood setting. Using different boundaries for measuring a neighborhood characteristic than for grouping residents (at a particular level of analysis) would be undesirable because it would contaminate the measurements with data from outside the neighborhood setting.

Problems with neighborhoods as places in discontinuous space. The practice of treating neighborhoods as places within discontinuous space leads to the *modifiable areal unit problem (MAUP)*, which breaks down into two more specific issues: the *boundary problem*, and the *scale problem* (Downey, 2006). The MAUP refers to the problems caused by the fact that there are many different ways in which a region can be subdivided into smaller areal units. The boundary problem manifests when one considers what might happen to the study results if the researcher changes where the boundaries between units are placed while holding the number of units constant; the scale problem manifests when one considers the potential impact of changing the number of the units (and hence their size or geographic scale and also their boundaries). In either manifestation of the MAUP, choosing different geographic definitions of the

neighborhood units would reallocate portions of the population from one unit to another (thereby changing how people are grouped) and change the geographic area over which neighborhood constructs are measured for each area (Bailey & Gatrell, 1995; Coulton et al., 2004). As a result, the variances of the neighborhood constructs would change, as would their covariances and correlations with other variables (Downey, 2006).

Lack of meaningful boundaries. Despite its extensive use in neighborhood research, HLM makes four assumptions that may be inconsistent with the underlying phenomena in neighborhood research (Mowbray, et al., 2007), some of which are closely related to aspects of the MAUP. First, HLM studies assume that the boundaries of a neighborhood unit accurately group residents with the people they see as their neighbors and adequately capture the geographic area of the neighborhood setting that really matters in terms of influencing resident outcomes. There is some research evidence to suggest that this is not always the case.

If neighborhoods are discrete, bounded entities, one would expect high agreement on where their boundaries lie, but there tends to be low agreement on the precise boundaries for particular places and residents' notions of where the boundaries lie rarely match those of census defined units (Coulton et al., 2004; Coulton, Korbin, Chan, & Su, 2001; Montello et al., 2003). Simply put, residents do not agree on where to draw neighborhood boundaries. That challenges the validity of using census-based units to define neighborhoods and suggests that we should pay more attention to how residents define neighborhoods. How they define their own neighborhoods may be quite important because it may affect whether events like crimes exert any contextual influence. For example, crimes occurring inside what a resident considers his or her own neighborhood

may be more salient and more likely to influence the resident's perception of neighborhood safety than crimes occurring outside that neighborhood, regardless of whether they occurred in the resident's census tract. The lack of agreement among residents about neighborhood boundaries makes it very difficult to argue that the kind of fixed boundaries needed for use with HLM are equally suitable for capturing the geographic areas relevant to different residents.

Furthermore, if neighborhoods are discrete, bounded entities, one would also expect that their boundaries would be good at capturing the patterns of spatial variation in all contextual measures. However, even if census-based neighborhood boundaries are well suited for their intended purpose of measuring neighborhood demographic characteristics, they are poorly suited to measuring other contextual characteristics that were not considered when defining those boundaries, such as crime (McCord & Ratcliffe, 2007), because the spatial distributions of these other characteristics do not necessarily match the spatial distributions of the demographic characteristics that informed the selection of the census boundaries. This undermines the assumption that all neighborhood characteristics should be measured in the same boundaries, further calling into question whether the representation of neighborhoods in HLM is adequate.

Without good reasons to believe that a particular discrete boundary system represents a meaningful and widely agreed upon method for dividing a study region into neighborhoods, the boundary problem associated with the MAUP takes on particular significance. In essence, different boundaries lead to different answers about what impacts (if any) neighborhoods actually have and there is little basis for singling out any particular set of boundaries and declaring it the one that provides the best answers.

Ignoring spatial proximity. Second, HLM focuses on residents' own neighborhoods, neglecting the fact that neighborhoods are embedded in a broader spatial context. The spatial arrangement of residents and neighborhoods is important because physical proximity plays a role in many aspects of daily life. For example, people living close to each other but separated by the (typically arbitrary) boundary between two census tracts may have more similar environments than people living on opposite ends of a large tract (Downey, 2006) and they may even consider themselves to be part of the same neighborhood (Coulton, Korbin, Chan, & Su, 2001). Routine activities such as traveling to work, visiting friends and relatives, shopping, and attending religious services can bring people into contact with other nearby block groups and residents (Sastry, Pebley, & Zonta, 2002). So, the relevant neighborhood setting for a resident may not really be confined to the boundaries of units like block groups or census tracts.

By ignoring spatial proximity between neighborhoods and treating them as if they are independent of and disconnected from one another, HLM assumes that all spatial correlation in resident outcomes is within-neighborhood correlation, but other spatial patterns may be apparent in the actual data, such as correlation that declines as a function of the distance between residents' homes without regard to neighborhood boundaries (Bass & Lambert, 2004; Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian, et al., 2005; Mowbray, et al., 2007). In short, most HLM analyses are purely place-based: they attend to the importance of the (narrowly defined) places residents inhabit, but ignore where those places are located in the wider geographical space and the potential importance of other nearby places that may affect residents.

Ignoring spatial variability in contextual conditions. Third, HLM assumes that contextual conditions are identical for all residents within a neighborhood (Roosa, et al., 2003). That amounts to asserting that residents are equally exposed to and affected by the conditions, events, resources, and social processes in the neighborhood unit to which they belong, regardless of where they live inside it. This may or may not be appropriate depending on the size of the neighborhood unit and the nature of the characteristic in question. A census tract with a high crime rate may contain a block plagued by frequent crimes, but otherwise consist of blocks where crime is rare. Surely the local environment with respect to crime is different for residents of the block experiencing the crime hotspot than for residents living elsewhere in the same tract, but a tract-level crime measure would ignore that local spatial variability. HLM cannot represent spatial variability in contextual conditions within neighborhood units without either entirely switching to smaller neighborhood units (e.g., from tracts to blocks) or using multiple levels of neighborhood units (e.g., both tracts and blocks) so that some characteristics could be measured at one level while others could be measured at a different level. Pursuing either of those options may increase the complexity of the research design and sampling procedures and will increase the sample size required for the study. This limits HLM's utility whenever contextual conditions do indeed exhibit important spatial variability within the selected neighborhood units.

Poor handling of spatial scale. Fourth, HLM is limited in its ability to answer questions about the geographical or spatial scale on which neighborhood characteristics influence outcomes. This usage of the term spatial scale refers to the size of the geographic area over which neighborhood characteristics are measured. In HLM, the

spatial scale for a contextual characteristic can be described by referring to the geographic units of analysis (e.g., blocks, census tracts) for which it is measured.³

Because of the nested nature of census geography, block level measures are at a smaller spatial scale than measures at the block group or tract levels (U.S. Census Bureau, 2002). Finding the spatial scale at which a neighborhood characteristic operates requires examining how the strength of its effect on outcomes changes as one varies the spatial scale on which it is measured. The spatial scale associated with the measurement that produces the strongest relationship with outcomes and the best model fit indices may be considered the scale at which the neighborhood characteristic operates (Chaix, Merlo, & Chauvin, 2005).

Without a priori theoretical reasons to expect that a particular spatial scale will be the right one to use, it is insufficient to test for the effects of a neighborhood characteristic at only one spatial scale because the scale chosen may be too small or too large, leading to incorrect inferences about the magnitude and/or significance of that predictor's effect. The crime example above postulated a situation where there was important spatial variability at the block level within a census tract. If residents are most sensitive to crime occurring quite close to their homes, then measuring crime at the tract level might obscure the relationship between crime and outcomes and it might be better to use a smaller spatial scale (e.g., block-level) for the crime measure. If on the other hand, residents are sensitive to crime within wider areas surrounding their homes, then a block group or tract level measure of crime might be better than a block-level measure.

³ If those units vary in size, descriptive statistics for the amount of area they occupy should be reported.

Unfortunately, HLM can only test the geographical scale on which contextual conditions like crime rates matter by associating them with pre-assigned geographic units that differ in size, such as block groups and census tracts. That limits HLM's flexibility because the available geographic units may not be the right size to best capture the effect of a specific contextual factor. In addition, few HLM studies of neighborhood effects have assessed how sensitive their conclusions about the effects of contextual factors are to varying the spatial scale of the neighborhood units. Instead, the (largely untested) assumption appears to be that all contextual characteristics operate on the spatial scale associated with the single set of neighborhood units selected by the researcher.

Summary. Neighborhood studies employing HLM methods have produced many interesting findings, but as described above, there are a number of spatial issues that they do not address well. One risk of using HLM is that it ignores the potential importance of spatial proximity; as a result, we may underestimate the amount of variability attributable to neighborhoods if the underlying pattern of autocorrelation in the data is not really hierarchically structured. Similarly, we may fail to detect, or underestimate the strength of, the effects of theoretically important contextual characteristics when we use HLM if the sizes or shapes of the neighborhood units used to measure those characteristics do not match the geographic areas really relevant to residents. Finally, HLM provides imprecise information about the spatial scale on which outcomes are autocorrelated and on which neighborhood characteristics operate because it does not directly quantify these concepts. Spatial scale can only be described in HLM approaches by describing the size of the geographic units used, but because census-based units vary in size (U.S. Census Bureau, 1994, 2002), HLM provides only a crude method of addressing questions about spatial

scale. Community psychology may benefit from examining alternative approaches to defining neighborhoods and studying neighborhood effects that can address some of the problems with HLM. GIS methods like GSM provide one such alternative.

Using GIS to Study Neighborhood Effects

Papers on GIS methods are just starting to appear in the community psychology literature (Bass & Lambert, 2004; Kruger, 2008; Kruger, Reischl, & Gee, 2007; Mowbray, et al., 2007), so there are few examples of how to apply GIS methods in our discipline. Nevertheless, GIS methods hold great promise as tools for our discipline because they allow researchers to adopt a more flexible conceptualization of geographic space and of neighborhoods as places than HLM. In GIS approaches, a place can be considered an ecological setting that is tied to a geographic location and possesses contextual characteristics that reflect local conditions in the geographic area surrounding that location. However, conceptualizing geographic space as a continuous rather than discontinuous phenomenon allows us to discard some of the constraints tied to the conceptualization of space required by HLM. Unlike in HLM, places do not need to have sharp, fixed boundaries (Montello, Goodchild, Gottsegen, & Fohl, 2003) and they may partially overlap with other places (Coulton, et al., 2004).

One of the advantages of GIS is that it gives researchers a framework that focuses on both place and space. Places certainly play an important role in GIS approaches to studying neighborhood effects: questions about how much resident outcomes vary from place to place and what characteristics of places predict that spatial variation are just as prominent in studies that use GIS methods as they are in HLM studies, if not more so. However, unlike HLM, GIS methods like GSM do not ignore the fact that neighborhoods

exist within a larger geographic space. Researchers using GSM can ask new questions about the roles of spatial proximity and spatial scale in neighborhood phenomena and can customize the neighborhood boundaries used to measure different contextual characteristics, thereby addressing some of the problems with HLM.

Accounting for spatial proximity with GIS. Researchers can go beyond purely place-based analyses by applying GIS methods that treat physical proximity and spatial relationships between places as important features of the data (Chaix, et al., 2006; Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian, et al., 2005; Downey, 2006). Rather than ignoring proximity and spatial relationships, GIS-based analyses can attend to these issues in ways that HLM does not by incorporating distance-based spatial autocorrelation directly into statistical models. Spatial autocorrelation refers to the phenomenon where the values on the same variable observed at different spatial locations (such as at the homes of different residents) are correlated, usually to different degrees based on how much distance lies between them (Bailey & Gatrell, 1995). This is a slightly different way of thinking about autocorrelation than the hierarchical autocorrelation structure used in HLM because now the relative positions of residents in space are what matters. Spatial autocorrelation in GIS approaches depends on spatial proximity between residents, not simply whether or not they live in the same arbitrary neighborhood boundary (such as a particular census tract).

For example, Bass and Lambert (2004) used a type of GIS-based spatial analysis called the variogram⁴ to show that perceptions of neighborhood disorder were more

⁴ A variogram plots the average dissimilarity (i.e., variance) in either raw variable values or residuals from a statistical model (on the vertical axis) between data points separated by a particular physical distance (on the horizontal axis; Goovaerts, 1997). Empirical variograms can usually be summarized by a mathematical

similar among adolescents who lived close together than among those who lived far apart, regardless of whether they lived in the same or different census tracts, even after controlling for levels of poverty, homicide, and juvenile arrest rates in the participants' own census tracts. Their study suggests that some processes influencing residents span the borders between census tracts and are a function of spatial proximity; it also illustrates how GIS methods can detect and model spatial patterns that HLM cannot handle because they require accounting for both place (e.g., census tract characteristics) and space (e.g., spatial proximity between observations).

GIS methods like GSM account for spatial proximity through the way they model autocorrelation as a decreasing function of the distance between observations (Banerjee, Carlin, & Gelfand, 2004). So while HLM treats two observations as similar if they belong to the same neighborhood unit regardless of the distance between them, GSM assumes that the degree of similarity between any two observations largely depends on how far apart they are located. GSM methods focus on spatial rather than hierarchical structure, so they ignore things like census boundaries when modeling autocorrelation and focus on the shape of the function relating autocorrelation to distance. This means researchers can ask several questions about the nature of that spatial autocorrelation, such as: how far does spatial autocorrelation reach for a particular outcome (i.e., what is the spatial scale at which it is evident)? How quickly does spatial autocorrelation decrease as distance between observations increases? What is the shape of the curve that describes how the

model that draws a smooth curve based on a few parameters. Spatial autocorrelation frequently manifests as a curve showing low dissimilarity at short distances and rising toward a plateau or limit that indicates that data are no longer autocorrelated after exceeding a certain distance between observations. Variograms can be converted into either covariance or correlation metrics (i.e., from a dissimilarity measure into a similarity measure), which will typically start with high values and then decrease with distance.

level of spatial autocorrelation changes with increasing distance between observations?

None of these questions can be answered with HLM.

The models of autocorrelation implemented by HLM and GSM both serve to group data so that researchers can detect spatial variability in outcomes. By estimating neighborhood- and individual-level variance components, they provide the means for quantifying autocorrelation. Despite the fact that GSM and HLM make different assumptions about how to detect and model autocorrelation, both do so for the same reason: accounting for autocorrelation allows them to correct for violations of the independence assumption that otherwise cause regression models to perform poorly.

Measuring neighborhood characteristics with GIS. Although GSM does not rely on neighborhood boundaries for grouping residents the way HLM does, one still needs to draw neighborhood boundaries for the purpose of measuring neighborhood characteristics when using GSM to study neighborhood effects. As noted above, any given neighborhood boundary may serve quite well for capturing some contextual characteristics, but poorly for capturing others. O'Campo (2003) suggested solving this problem by using multiple operational definitions of neighborhood within the same study. Other authors agree that the geographic area over which neighborhood characteristics are most relevant to outcomes may depend on the specific characteristic in question and have also suggested that different characteristics may need to be measured within different boundaries surrounding a resident's home (Galster, 2001; Kruger, 2008).

GIS provides this flexibility. For example, Kruger (2008) measured deterioration of both residential and commercial buildings in circular buffers centered on each resident's home, but used buffers of different sizes for the two measures. He found that

deterioration in commercial buildings correlated most strongly with fear of crime when measured in a 1.00 mile radius, but deterioration in residential buildings correlated most strongly with fear of crime when measured in a 0.25 mile radius. Thus, different geographical boundaries mattered for these two contextual variables when they are used to predict fear of crime.

Kruger's (2008) study also illustrates how neighborhoods can be allowed to partially overlap in GIS analyses. Because the buffers were centered on residents' homes, each resident effectively had a unique neighborhood boundary for each contextual measure. However, the neighborhood buffers for two residents living less than 0.25 miles apart would substantially overlap; they would only be identical if the two residents lived at the same location. Buffers surrounding participants' homes have also been used to operationalize neighborhood boundaries for measuring contextual conditions in epidemiological studies that relied on GSM (Chaix, et al., 2006; Chaix, Merlo, Subramanian, et al., 2005).

Using buffers to represent neighborhood boundaries that are positioned relative to residents' homes also allows GSM models to accommodate the fact that residents often think of their homes as the center of their neighborhood (Coulton, et al., 2001). It may be particularly useful when the contextual characteristic shows spatial variability within the boundaries of the kinds of neighborhood units used in HLM. Allowing buffers to overlap preserves the intuitive notion that people living close together are exposed to similar environments, but permits people who are farther apart to have more distinct neighborhood environments.

Another advantage of GIS methods is that they enable researchers to construct new kinds of contextual measures that take location and spatial relationships into consideration and to easily vary the spatial scale on which those measures are calculated (Haining, 2003; Luke, 2005). As an example of measuring contextual variables from a spatial perspective, GIS could measure access to health care by calculating whether the distance between a person's home and the nearest health care provider is less than some criterion, such as 2 km. Varying the spatial scale of that access measure is a simple matter of changing the distance criterion (e.g., from 2 km to 4 km). This flexibility in varying the spatial scale over which contextual factors are measured enables researchers to learn more about the geographic scale on which contextual factors matter for residents by comparing alternative models that differ only in the size of the area over which a measure is calculated (Chaix, et al., 2006; Chaix, Merlo, Subramanian, et al., 2005).

HLM can examine the spatial scale on which neighborhood characteristics matter by comparing models that differ in terms of the size of the neighborhood units, but only when the data can be matched to another available type of geographic unit. In contrast, GSM can directly examine the scale on which neighborhood characteristics matter even when it is larger or smaller than available units by using buffers to represent neighborhood boundaries (Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian et al., 2005). This is useful because it allows researchers to (a) independently vary the spatial scale on which different neighborhood characteristics are measured, (b) ask questions about the spatial scale on which a particular neighborhood characteristic exerts the strongest influence on outcomes (c), precisely quantify that scale, and (d) compare the spatial scales on which different neighborhood characteristics operate.

In an HLM framework, which requires a place-only conceptualization of neighborhoods, the access to care measure described above might be reduced to recording whether or not there was a health care provider in the resident's neighborhood. In other words, the measurement of the contextual variable is confined within the perimeter of a neighborhood area that encloses a fixed, absolute portion of geographic space. This approach has limitations, especially for people living near the edges of neighborhoods, because it might treat people whose own neighborhood had no providers but who could cross the street into another neighborhood to see a nearby provider as if they had no access to care. GSM can solve that problem by using buffers that define the neighborhood boundaries for the access to care measure so that they enclose the portion of space immediately surrounding the resident's home (out to some specified distance), allowing the buffer to cross the boundaries of traditional neighborhood units like block groups or census tracts.

Summary. GIS methods like GSM can address some of the limitations inherent in the way neighborhoods must be defined in order to use HLM to study neighborhood effects. The fact that community psychologists have rarely applied methods that can address these spatial issues represents a missed opportunity. The relative neglect of GIS methods means we are unnecessarily restricting the range of questions we can answer, excluding important spatial issues from our theories, and failing to model spatial patterns that may exist in our data. If we wish to understand neighborhood phenomena but do not attend to the spatial aspects of those phenomena, we may be missing incredibly important parts of the story and we risk arriving at inaccurate conclusions with respect to the role of some contextual factors. Neighborhood research in community psychology will benefit

from taking a closer look at the spatial issues surrounding how neighborhoods are defined and testing alternative ways to represent them in our statistical models.

The Current Study: Comparing GSM and HLM

There has been little discussion in the community psychology literature about whether there is a better way to represent neighborhoods and geographic space in multilevel neighborhood research than that offered by HLM. Consistent with the call to utilize GIS techniques and adopt statistical methods that reveal and explain a wider array of patterns in our data (Luke, 2005; Mowbray et al., 2007), this study examined whether GSM provides a valuable alternative to HLM for rigorously studying neighborhood effects on residents—one that considers both place and space and that offers a wider array of options for defining neighborhood boundaries.

So far, very few studies have directly compared HLM and GSM approaches to studying neighborhood effects (Boyd, Flanders, Addiss, & Waller, 2005; Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian, et al., 2005). There are several differences between these methods that may affect which one is better suited to the task of studying neighborhood effects on residents. For example, the present study contributes to the literature by investigating the relative value of two different conceptualizations of geographic space and neighborhoods. The discontinuous view of space underlying how neighborhoods are defined in HLM studies constrains how they are represented in statistical analyses and is poorly aligned with some empirical findings about the nature of neighborhoods. In contrast, the continuous view of space underlying GIS methods like GSM may permit models to more closely match the nature of neighborhood phenomena.

Focal constructs. This study used both HLM and GSM to quantify the amount of autocorrelation in residents' perceptions of neighborhood problems and to examine whether neighborhood crime and neighborhood socioeconomic status (NSES) exert contextual influences on those perceptions. It compared the HLM and GSM results in order to answer the research questions laid out below. The study focused on perceived neighborhood problems (Foster-Fishman, Cantillon, Pierce, & Van Egeren, 2007; Foster-Fishman, Pierce, & Van Egeren, 2009) because both HLM methods (Coulton, et al., 2004) and GIS methods (Bass & Lambert, 2004; Pierce, 2006) have previously found evidence of neighborhood-level variability in this outcome, but no prior study has examined it with both methods.

Crime and NSES (Leventhal & Brooks-Gunn, 2000; Sampson, et al., 2002) are contextual predictors that appear frequently in the neighborhood effects literature, have clear theoretical links with perceived problems, and can be measured within any set of neighborhood boundaries by aggregating point-based crime incident and parcel-based housing value data. In addition, crime and NSES are examples of neighborhood characteristics that may not be adequately captured by the boundaries of census-based units typically used in HLM studies (McCord & Ratcliffe, 2007).

Both neighborhood crime and NSES are salient to residents. Residents regard the presence of crime as a problem (Sampson & Raudenbush, 2004), while neighborhood poverty is often associated with observable signs of social and physical disorder (Sampson, 2001; Sampson, et al., 2002), hence both crime and NSES may predict perceived neighborhood problems simply because they are observable indicators of the kinds of problems assessed in those perceptions. However, the stigma associated with

poverty may also prime residents of poor neighborhoods to perceive more problems than can be accounted for by observable disorder alone (Sampson & Raudenbush, 2004).

This study also controlled for several individual-level predictors of perceived neighborhood problems. Factors such as age, sex, and race are known to be associated with this outcome (Franzini, et al., 2008; Meersman, 2005; Quillian & Pager, 2001; Sampson & Raudenbush, 2004) and may also be related to where people live through social processes that produce various forms of residential segregation. That made it important to control for the composition of the neighborhood population by adding individual-level predictors to the models to obtain better estimates of the effects of neighborhood-level predictors (C. Duncan, et al., 1998; Merlo, Yang, et al., 2005).

Research questions. If GSM yields better statistical models than HLM, it would suggest that we may need to replace the simple conceptualization of neighborhoods associated with HLM with one that is more sophisticated and more compatible with what we know about the nature of neighborhoods. In HLM, neighborhoods are places with sharp, fixed boundaries that never overlap, while in GSM they are places with fuzzy, overlapping boundaries. Whether the differences in how GSM and HLM represent neighborhoods matter in practical terms depends partly on whether the two methods yield different answers about how much variance in outcomes is attributable to neighborhoods. So, the first research question for the present study is: how do GSM estimates of neighborhood-level variance and autocorrelation compare to HLM estimates?

Partitioning the variance in resident outcomes into neighborhood- and individual-level variability is vital to multilevel neighborhood research because testing whether particular neighborhood characteristics influence outcomes is only worthwhile when

there is neighborhood-level variability to be explained. The practice of defining neighborhoods in geographic terms means that neighborhood-level variability is spatial variability, so multilevel analysis methods must be able to detect spatial variability in outcomes and assess the degree to which adding predictors to a statistical model explains that variability. Spatial variability implies that neighboring residents' outcomes are more similar than they would be if people were randomly distributed across geographic space, so it shows up as lack of independence between observations. Thus, autocorrelation is simply another name for a structured form of spatial variability.

Part of the statistical rationale for using either HLM or GSM instead of OLS regression is that failure to account for autocorrelation leads to artificially small standard errors for the regression coefficients and increased Type I error rates. HLM and GSM methods differ in how they handle autocorrelation. While HLM assumes that the autocorrelation is hierarchically structured and derives from shared membership in discrete neighborhoods units, GSM assumes it is spatially structured and a function of distance between observations. The ability of HLM and GSM to detect spatial variability and control for autocorrelation depends on how well their assumptions about autocorrelation match up with the actual structure in the data. Thus, whether HLM or GSM is more appropriate may ultimately depend on the kind of data being analyzed, but the existing empirical literature provides very limited information that might guide decisions about which method to use.

In one of the few HLM studies that has varied the size of the neighborhood units, Coulton et al. (2004) examined how the intraclass correlations (ICCs) for residents' perceptions of neighborhood safety, social cohesion, informal social control, police

relations, and disorder and incivilities varied when aggregated under different neighborhood definitions. They used five different types of geographic units. The first four types of units correspond to units of steadily decreasing size: sites selected by the Making Connections initiative, project-designated sub-areas within those sites, census tracts, and block groups. The last unit, named neighborhoods, vary in size and may be larger or smaller than most of the other units, but are always smaller than the Making Connections sites.

Coulton et al. (2004) found that the ICCs were higher for smaller spatial definitions of neighborhoods and that both statistics were more sensitive to changes in neighborhood size for some constructs (perceived safety and disorder/incivilities) than for the other constructs. Still, four of the five outcomes examined (including the measure of disorder and incivilities), showed higher levels of autocorrelation when the neighborhood units were smaller. This suggests that the geographic scale of spatial variation for different constructs may differ. They speculated that perceived safety and disorder/incivilities varied more on a block to block scale than the other constructs because the questions asked about more concrete, observable phenomena, so that the geographical area over which residents might agree in their assessments would be smaller. Overall, their results suggest that the underlying structure of the data may be better modeled as spatial rather than hierarchical autocorrelation because if autocorrelation decays with increasing distance between observations, grouping observations within larger geographic units should reduce the average level of autocorrelation observed as compared to grouping them within small geographic units.

Unfortunately, Coulton and colleagues (2004) did not try analyzing their data with spatial models like GSM, which might have provided direct evidence about whether their data showed spatial autocorrelation. Other authors have argued that HLM does not fully account for autocorrelation because spatial autocorrelation can still be discerned in HLM residuals (Boyd, et al., 2005; Chaix, Merlo, & Chauvin, 2005), but no studies have yet reported whether hierarchical autocorrelation can still be discerned in GSM residuals. Thus, we do not yet know which technique more fully accounts for autocorrelation because previous comparisons between HLM and GSM models are incomplete. Accordingly, the second research question for the present study is: which method (HLM or GSM) is more effective at modeling the autocorrelation actually observed in data from neighborhood residents? Autocorrelation is a prerequisite for finding neighborhood effects because neither occurs unless at least some variance is attributable to neighborhoods. Detecting autocorrelation and testing neighborhood effects should work best when we fit statistical models consistent with the underlying structure in the data.

The difference between HLM and GSM goes beyond how they each group observations to model autocorrelation. The two different approaches to defining neighborhoods have important implications for the measurement of neighborhood characteristics. GSM provides more flexibility to customize the boundaries used for each contextual variable than HLM and can use boundaries that are set relative to the resident's location. While that flexibility is conceptually appealing, the most important test of whether GSM offers a superior method for representing neighborhoods depends on whether the two methods yield different answers about how strong the effects of neighborhood-level predictors are and how well the resulting statistical models fit the

observed data. Therefore, the third research question is: how do GSM estimates of contextual effects and model fit compare to HLM estimates?

Finally, HLM and GSM also differ in how they handle questions about the spatial scale on which neighborhood characteristics influence outcomes. Attending to issues of spatial scale will highlight new aspects of neighborhood effects that need to be explained, thereby opening up new avenues for theory development. HLM studies frequently use only a single type of geographic unit to represent neighborhoods and thus do not attempt to study whether different contextual predictors operate at different spatial scales. They simply assume that studying all aspects of the neighborhood phenomena at a single spatial scale is appropriate. There are presumably reasons why researchers using GIS methods have found that the strength of the relationship between neighborhood-level predictors and a given outcome depends on the size of the area over which the predictor is measured and on the specific predictor being studied (Kruger, 2008; Meersman, 2005), but there is currently little theory to explain those findings. With appropriate theory, we could predict and explain which outcomes might vary on small versus large geographic scales and what geographic scale to use when measuring specific neighborhood characteristics (Messer, 2007).

Unfortunately, studies have rarely tried varying the definition of neighborhoods, so we have few empirical findings to guide hypotheses about the geographic scale at which we should expect to see the effect of a particular neighborhood characteristic on a given outcome (Diez Roux, Mujahid, Morenoff, & Raghunathan, 2007). Applying GSM in the present study has the potential to add to the emerging body of empirical evidence about the spatial scale of neighborhood effects that will then permit us to develop theories

that explain when and why a predictor should be operating on a particular scale. So, the fourth research question for the present study is: in a dataset originally collected with use of HLM methods in mind, how do the geographical scales on which different contextual factors operate (as estimated with GSM) compare to each other and to the size of the neighborhood units used in HLM?

Organization of the text. To lay the conceptual foundations for this study, the first section of the literature review focuses on neighborhoods and multilevel research, highlighting why neighborhoods are important contexts and discussing key multilevel assumptions about residents and neighborhoods as units of analysis. The second section focuses on the conceptualization of neighborhoods as places within geographic space. The third section focuses on the HLM approach to testing contextual effects, explaining its strengths and limitations for studies of neighborhood effects. It describes why HLM has become a popular tool in neighborhood research and the assumptions, advantages, and disadvantages associated with it. The fourth section focuses on the origin and conceptual underpinnings of the GSM approach to testing contextual effects, explaining how it relates to the conceptualization of neighborhoods and measurement of neighborhood context, how it addresses some of the limitations in HLM, the kinds of questions it can answer, and its recent applications in the social sciences. The fifth section reviews the handful of previous studies that have compared HLM and GSM approaches. The sixth section provides background material relevant to the specific substantive example that was used to compare HLM and GSM, which involved using contextual measures of crime and NSES to predict residents' perceptions of neighborhood problems. The seventh section then identifies gaps in that literature, and presents the hypotheses for

the study. Finally, the literature review closes with a brief summary of the goals and objectives of the study and a short discussion of its limitations. Following that, the methods employed in this study are described, and then the results are presented. The document concludes with a discussion of the findings and their implications for neighborhood research.

LITERATURE REVIEW

Neighborhoods and Multilevel Research

Studying the connections between human behavior and the ecological contexts within which it unfolds has always been a prominent theme in community psychology (Anderson, et al., 1966; Livert & Hughes, 2002; Maton, Perkins, & Saegert, 2006). This ecological perspective contributes to community psychologists' enduring interest in research that spans multiple levels of analysis (Dalton, Elias, & Wandersman, 2001) because it positions contexts as phenomena to be studied right along with human behavior. Researching contextual phenomena requires careful attention to identifying and conceptualizing the units of analysis that constitute the context of interest, how to measure the theoretical constructs at each level of analysis (Linney, 2000), and the application of methods suited to answering questions about contextual effects (Luke, 2005; Shinn & Rapkin, 2000).

Neighborhoods as meaningful contexts for residents. A multilevel, ecological perspective lies at the heart of the most basic premise in neighborhood research, which is that neighborhoods are meaningful contexts for their residents. So, why do neighborhoods merit study—what makes them meaningful contexts? People, especially children, spend substantial amounts of time in neighborhood settings, allowing ample opportunity for environmental conditions, events, and social processes in these settings to influence individuals. Because the world is not uniform and has different characteristics from place to place, neighborhoods offer varying levels of access to material, institutional, and social resources (e.g., housing, public services, schools, social interaction, etc.) that may affect residents' welfare and prospects in life (Galster, 2001).

Neighborhoods often take on symbolic identities and importance, as evidenced by the way people attach names to their neighborhoods (Lee & Campbell, 1997). Residents can form deep emotional attachments to these places and to their neighbors (Manzo & Perkins, 2006; Unger & Wandersman, 1985); they also sometimes form voluntary associations to express their identification as members of a shared neighborhood or to advocate for their collective interests (Unger & Wandersman, 1983).

Yet another reason to study neighborhoods is that government agencies, philanthropic foundations, and other organizations set policies, develop programs, offer services, and engage in other forms of intervention with respect to neighborhoods, treating them as identifiable units for planning and action related to social change (Chaskin, 1998). Finally, empirical research has provided extensive evidence that neighborhood conditions influence outcomes such as school readiness and achievement among children, teen pregnancy, physical and mental health, perceptions of crime and disorder in the neighborhood environment, and rates of violent crime (Franzini, Caughy, Spears, & Esquer, 2005; Gephart, 1997; Kruger et al., 2007; Leventhal & Brooks-Gunn, 2000; Quillian & Pager, 2001; Sampson et al., 2002; Sampson, Raudenbush, & Earls, 1997; Shinn & Toohey, 2003; Wyant, 2008). For all of these reasons, neighborhoods are contextual settings that merit our attention.

Theoretical mechanisms underlying neighborhood effects. Neighborhood researchers have described numerous theoretical mechanisms through which neighborhoods may influence residents' welfare, behavior, and development including collective socialization, institutional resources, contagion, competition, and relative deprivation theories, among others (Leventhal & Brooks-Gunn, 2000; Sampson, et al.,

2002). Authors categorize and label these mechanisms differently and some write about mechanisms not discussed by others. A comprehensive review of these theoretical mechanisms is beyond the scope of this study, so what follows is only a brief discussion of some key mechanisms that have been widely discussed. For example, collective socialization has been described as a mechanism through which social groups exert influence on residents' attitudes, values, and behavior by providing a structured social environment with role models, parental supervision and monitoring, routines, and deviation-counteracting social interactions, all of which tend to produce conformity with group norms (Galster, 2001; Leventhal & Brooks-Gunn, 2000; Shinn & Rapkin, 2000).

Institutional resources provide another mechanism through which neighborhoods can influence residents (Sampson, et al., 2002). Because neighborhoods vary in the availability, accessibility, and quality of resources such as libraries, community centers, public services, and recreational programs that promote learning and development, institutional resources provide a mechanism for explaining some neighborhood effects (Leventhal & Brooks-Gunn, 2000). For example, adolescent girls living in neighborhoods with more parks engage in more physical activity, suggesting that parks (particularly those that have more amenities conducive to walking and physical activities) are institutional resources that promote exercise among residents (Cohen, et al., 2006).

Contagion models posit that the behavior of a resident might directly influence the same behavior in his or her neighbors, leading to (typically negative) behaviors that spread like epidemics within neighborhoods (Leventhal & Brooks-Gunn, 2000). With respect to outcomes such as attitudes or perceptions, neighbors engaging in the social

construction of reality might share information and mutually influence each other's perceptions (Shinn & Rapkin, 2000), leading to a contagion effect.

Competition for access to and control over scarce resources, in which one person's gain necessarily comes as a loss to others, provides another mechanism through which neighborhoods may influence residents (Dietz, 2002; Galster, 2008; Leventhal & Brooks-Gunn, 2000). When neighborhoods influence resident outcomes through residents' comparison of their own situation to that of their neighbors, the concept of relative deprivation may explain neighborhood effects (Dietz, 2002; Galster, 2008; Leventhal & Brooks-Gunn, 2000). For example, relative deprivation might explain why the presence of homeowners in a neighborhood could have a detrimental effect on nearby renters (Haurin, Dietz, & Weinberg, 2003).

Finally, attraction, selection, and attrition processes (Shinn & Rapkin, 2000) account for the fact that the composition of neighborhoods is rarely a random sample of the larger population. Processes that affect who is attracted to a particular neighborhood, opts to live there, or decides to leave may contribute to the geographical clustering of similar people within neighborhoods. If the individual-level characteristics on which residents are similar are also related to the outcome of interest, neighborhood effects may be present because of varying composition rather than varying contextual conditions. Therefore, controlling for the composition of neighborhoods when testing for contextual effects (Bingenheimer & Raudenbush, 2004; C. Duncan, et al., 1998; Merlo, Yang, et al., 2005) is important to avoid confounds.

Given the variety of theoretical mechanisms for neighborhood effects already identified, it is clear that the specific mechanisms that could explain an effect will depend

on the specific constructs involved in a given study. For example, institutional resource theory may be particularly well-suited to explaining why the presence of youth- service organizations in a neighborhood can contribute to the social development of the youth who live there (Quane & Rankin, 2006). However, authors such as Papachristos and Kirk (2006) have argued that theoretical mechanisms such as collective efficacy and informal social control play a strong role in controlling how much gang violence occurs in a neighborhood. It is also important to recognize that multiple theoretical mechanisms may be working in concert to produce neighborhood effects. Thus, a researcher might incorporate several predictors into a statistical model, each of which may represent the influence of a distinct theoretical mechanism.

In the present study, contextual effects of crime and NSES represent two different theoretical mechanisms through which neighborhoods might affect residents' perceptions of neighborhood problems. According to broken windows theory (J. Q. Wilson & Kelling, 1982), visible signs of physical and social disorder in a neighborhood exert a direct contextual influence on residents' perceptions of neighborhood problems (Quillian & Pager, 2001; Sampson & Raudenbush, 2004). Exposure to higher levels of actual crime, which is an extreme form of social disorder (Sampson & Raudenbush, 1999), should lead to higher levels of perceived problems. In contrast, NSES may exert a contextual influence on residents' perceptions because poor neighborhoods have become stigmatized as disorderly places where problems are rampant (Sampson & Raudenbush, 2004). Thus, residents living in places with low NSES may perceive more problems than they would in a wealthier, but otherwise similar, neighborhood.

Because the present study focused on comparing HLM and GSM with respect to testing the effects of contextual characteristics of neighborhoods, it was important to control for neighborhood effects that could be explained by individual-level factors associated with residents' perceptions working in concert with theoretical mechanisms that might lead to geographical clustering of similar individuals. To do that, several individual-level predictors were used in to control for the effects of unobserved attraction, selection, and attrition processes that might generate neighborhood effects on perceived neighborhood problems through their influences on neighborhood composition.

Finally, contagion processes resulting from social interaction and information sharing among residents (Shinn & Rapkin, 2000) could easily lead to mutual influence on residents' perceptions of the neighborhood. Such a mechanism might account for residual autocorrelation remaining in residents' perceptions after accounting for neighborhood composition, crime, and NSES effects. Thus, the statistical models in the present study incorporate variables representing several different theoretical mechanisms that can explain neighborhood effects.

Multilevel assumptions in neighborhood research. Inherent in the premise that neighborhoods are meaningful contexts that exert influence on resident outcomes are some key multilevel assumptions. Those include: (a) neighborhoods and individuals are separate kinds of observable units, with the former at a higher level of analysis than the latter because individuals live within neighborhoods, (b) neighborhoods differ from each other in important ways such that their characteristics define, at least in part, the nature of the environmental context affecting their residents, and (c) neighborhood characteristics can directly or indirectly influence resident-level processes and outcomes. Subjecting that

last assumption to inquiry is the point of studying neighborhood effects, which also requires grappling with the second assumption because identifying dimensions on which neighborhoods vary from one another is a prerequisite to explaining contextual effects. But before we can measure neighborhood characteristics we need a conceptual definition of neighborhoods as units of analysis and a way to operationally define them that is aligned with that conceptual definition.

Defining Neighborhoods

Before delving into the conceptual definitions of neighborhoods adopted by researchers, it is useful to explore residents' colloquial definitions of neighborhoods. When Guest and Lee (1984) asked residents of Seattle, Washington to define the word neighborhood and the boundaries of their neighborhoods, over 76% of the residents they interviewed defined neighborhoods in terms of a geographic area or territory, although only 30% relied on solely physical definitions. Almost 60% of the residents defined neighborhood in terms of nearby people; 39% endorsed social definitions based on sense of community and social cohesion. Finally, about 10% defined neighborhood in terms of local institutions (e.g., schools, shopping centers, parks, and so on). Guest and Lee concluded that one major dimension in their data was a contrast between geographic and social definitions of neighborhood. They also found that residents who provided institutional definitions reported having larger neighborhoods than people who provided other kinds of neighborhood definitions.

More recent work by Lee and Campbell (1997) in Nashville, Tennessee, further clarifies how residents think about the nature of neighborhoods. Nearly 87% of their sample endorsed a territorial definition, showing that the notion of neighborhoods as

geographic places is widespread. The social dimension, tapping both definitions based on nearby people and definitions based on sense of community, was endorsed by a little over 40% of the residents. Most of the residents (59%) gave egocentric definitions wherein their own home served as a spatial referent. The final dimension in this study was a structural one that defined neighborhood in terms of physical structures, similar to Guest and Lee's (1984) institutional definition. Lee and Campbell also found that the perceived size of a neighborhood varied considerably even among people who agreed on the name.

Another interesting set of findings about how people think about places comes from work by Montello, Goodchild, Gottsegen, and Fohl (2003), who were investigating how people spatially define named places. Comparing maps drawn by different participants, they found strong evidence that people varied in where they drew the boundaries of downtown Santa Barbara, California despite the fact that this place has a strong symbolic identity. While there was a core area where most, if not all, of the maps overlapped, they concluded that it may be better to think of places as having fuzzy or probabilistic boundaries, rather than sharply defined edges because fewer maps overlapped locations farther out from the core area.

The findings discussed above provide insight into what typical residents mean when they talk about neighborhoods, which very often includes an element of place, but frequently has social elements too. Both HLM and GSM approaches ultimately define neighborhoods in geographic terms, but they differ in how they do that because they make different assumptions. In HLM, geographic space is treated as a discontinuous phenomenon, so neighborhoods are places with fixed, non-overlapping boundaries that apply equally to all neighborhood characteristics. In GSM, geographic space is treated at

a continuous phenomenon, so neighborhoods are places defined relative to where residents live and may have different boundaries depending on what neighborhood characteristic is being measured. As a result, neighborhoods in GSM may overlap and do not have fixed boundaries.

The findings above suggest that the HLM assumption that neighborhoods are clearly bounded places may oversimplify the phenomenon. Residents appear to have very idiosyncratic and egocentric notions about what constitutes their neighborhoods. That suggests that neighborhood-level constructs for residents living at the center of discrete geographic units like census tracts or block groups may contain less measurement error than they do for residents living on the edges of those units. The residents along the edges may be more likely to consider areas outside that unit as being part of their neighborhood, but the HLM approach would exclude those additional areas from the measurement of neighborhood characteristics, thereby introducing additional error into the measurement. Fortunately, GSM is compatible with a wider range of ways to define neighborhoods than HLM, including egocentric definitions that treat the space within a certain distance of one's home as the neighborhood. In other words, GSM methods may fit resident conceptions of neighborhoods better than HLM. This serves as a useful point of reference as the review now moves on to discuss the conceptual definition of neighborhoods.

Defining neighborhoods: Social versus geographical units. Scholars from a wide array of scientific disciplines have written about how to conceptualize neighborhoods (cf., Chaskin, 1997; Galster, 2001; Nicotera, 2007). Comparing the many definitions reveals a consensus that neighborhoods are complex, multidimensional entities comprised of a combination of objective physical and environmental

characteristics tied to geographic places and subjective, socially constructed characteristics that emerge from social interactions and lived experience. However, different authors emphasize different aspects of neighborhoods. One strand in the literature primarily views neighborhoods as communities (social units) that have developed naturally through the processes involved in the growth of cities (Chaskin, 1997; Suttles, 1972), while another strand emphasizes viewing them as geographic places or territories (geographic units). As will be shown below, these are not mutually exclusive perspectives (Chaskin, 1997).

Neighborhoods as social units. Viewing neighborhoods as natural communities focuses attention on things like the presence of social networks, a sense of community, shared culture and values, place attachment and identity, social cohesion, and other social processes such as economic exchange relationships (Chaskin, 1997; Forrest & Kearns, 2001). While neighborhoods conceptualized as local communities may not be strictly confined within small geographic areas, they are often spatially concentrated in ways that anchor them to particular places (Chaskin, 1997). According to a recent community psychology textbook, “Neighborhoods might be defined as local communities that are bounded together spatially where residents feel a sense of social cohesion and interaction, homogeneity, as well as place identity” (Duffy & Wong, 2002, p. 18). Specifying that neighborhoods are spatially bounded acknowledges that even definitions emphasizing the social aspect of neighborhoods (as Duffy and Wong’s does) must also recognize that local communities occupy identifiable geographic places.

Neighborhoods as geographic units Meanwhile, viewing neighborhoods primarily as geographic units focuses attention more on identifying the boundaries that

demarcate the geographical areas they occupy. Boundaries can be delineated by asking residents to identify them (Coulton et al., 2004; Coulton et al., 2001; Guest & Lee, 1984; Lee & Campbell, 1997), looking at physical features of the environment like the layout of the street network (Grannis, 1998; Guo & Bhat, 2007), or by relying on boundaries set by government agencies and other external organizations to facilitate their own activities (Chaskin, 1997). Of course, local stakeholders often use their knowledge about the social aspects of neighborhoods to inform the selection of those geographical boundaries (Chaskin, 1997; U.S. Census Bureau, 1994, 2002).

Consistent with the approach taken in most studies of neighborhood effects (Burton, Price-Spratlen, & Spencer, 1997; Coulton et al., 2004; Diez Roux, 2001; Gephart, 1997; Lee, 2001; Leventhal & Brooks-Gunn, 2000; Sampson et al., 2002), this study defines neighborhoods geographically (i.e., as places that occupy areas within geographic space). This was crucial to operationalizing neighborhoods for the purpose of measuring contextual characteristics that are known to vary across geographic space (e.g., crime and NSES). But, defining neighborhoods requires considering an aspect of their geographical representation that is rarely discussed in the community psychology literature on neighborhood effects: the relationship between place and space.

Relating places to geographic space. Geographical space stretches across the entire surface of the earth: it is the physical environment within which nearly all human activity is embedded. Maps show where various features of the world can be found within geographic space. Locations within geographical space can be precisely identified by spatial coordinates (e.g., latitude and longitude), but places tend to be larger than simple point-referenced locations. Instead, places are subsets of geographic space (i.e.,

contiguous collections of locations). Perhaps more importantly, places are portions of space that have been imbued with identity, meaning, or purpose through human activities, experiences, cognitions, and intentions (Relph, 1976). For example, political maps illustrate how geographic space is divided into various countries, which are large places controlled by sovereign national governments.

Residential neighborhoods are important places because they are settings in which people engage in routine activities of life. There are two major options for linking places to space, each of which has implications for how we study neighborhoods. Geographic space can be conceptualized as discontinuous and fragmented into mutually exclusive places that have fixed boundaries, or as a continuous field in which places may overlap and may have variable or indeterminate boundaries. The next two sections deal with these contrasting conceptualizations of the relationship between place and space in neighborhood research, describing the implications of each, particularly with respect to how neighborhood characteristics are measured.

Neighborhoods as places in discontinuous geographic space. The vast majority of multilevel neighborhood effects studies use administrative areas such as census tracts or block groups as the geographical boundaries of neighborhoods (Burton et al., 1997; Coulton et al., 2004; Diez Roux, 2001; Gephart, 1997; Lee, 2001; Leventhal & Brooks-Gunn, 2000; Sampson et al., 2002), although some studies have used larger geographical units comprised of multiple census tracts grouped together (Browning & Cagney, 2002; Browning, Feinberg, & Dietz, 2004; Sampson et al., 1997). Using these administratively defined areas as neighborhoods makes it easy and cost effective to measure some neighborhood characteristics with census data, which may contribute to the prevalence of

this practice (Lebel, Pampalon, & Villeneuve, 2007). However, few of those studies discuss the conceptual definition of neighborhoods or the conceptualization of geographic space underlying their neighborhood definitions. So, a closer look at how census geographic units are constructed and what using them implies about the relationship between geographic space and place is warranted.

Census geography. The Census Bureau works extensively with local governments and other stakeholders to develop a hierarchical system of boundaries that divides the nation into many small geographic units for use in collecting and tabulating decennial census data (U.S. Census Bureau, 1994, 2002). The foundation of that system is a discontinuous conceptualization of geographic space in which boundaries divide space into distinct geographic units that occupy mutually exclusive areas: census units at the same level of the hierarchy never overlap. As shown in Figure 1 above, the three lowest levels in the hierarchy of geographic units used by the Census Bureau (in order of increasing size) are blocks, block groups, and census tracts. Census data are most routinely tabulated at the tract level.

Units like block groups and tracts are useful to neighborhood researchers who want to use HLM. Using them to represent neighborhoods allows researchers to adopt a well-known boundary system that is grounded in a discontinuous conceptualization of space. Studies employing HLM for neighborhood research rely on that conceptualization because they must be able to identify which residents to group together. At any given level of analysis (e.g., block groups), the key issue is that each neighborhood must be a place with an unambiguous boundary demarcating the division between the portion of space that belongs to it and the portion that belongs to other neighborhoods to ensure

unambiguous sorting of residents into neighborhoods. Census geographic units clearly fulfill that requirement and ensure that each resident will belong to one and only one neighborhood at any given level of analysis.

Both in the census boundary system and in custom boundary systems developed by researchers, visible geographic features (e.g., major streets, waterways, railroads, parks, etc.) often define most of the boundaries between neighborhoods, but some boundaries are selected to ensure that the resulting neighborhoods are internally homogenous and externally heterogeneous with respect to important demographic and socioeconomic characteristics (Sampson et al., 1997; U.S. Census Bureau, 2002; Weiss, Ompad, Galea, & Vlahov, 2007). For neighborhood researchers, the overarching goal is to obtain neighborhood units that are ecologically meaningful units. Because all space outside a neighborhood's boundary belongs to some other neighborhood, the discontinuous view of space implies that measuring a neighborhood characteristic over any area stretching beyond that boundary would contaminate the measurement with data from a different neighborhood and lead to measurement error. Thus, the discontinuous view of space fosters defining neighborhoods as discrete entities with fixed boundaries that apply to all neighborhood attributes measured at a given level of analysis.

The MAUP. Unfortunately, there are many alternative ways to subdivide any geographic region, each of which might group residents differently and each of which can result in different values for the neighborhood attributes that might be associated with a resident. This is what causes the MAUP. Because of the MAUP, the specific set of boundaries chosen by a researcher to divide space into neighborhood units influences the results of analyses (Bailey & Gatrell, 1995; Downey, 2006). So if two researchers set out

to test the same theoretical model using data from the same study region, their studies can yield different statistical results if they use different sets of neighborhoods boundaries.

The sensitivity of statistical inferences about contextual effects to the specific boundary system used to define neighborhoods may lead researchers to draw inaccurate conclusions or make poor policy recommendations (Bailey & Gatrell, 1995; Guo & Bhat, 2007). For example, Kruger (2008) found that the correlation between residents' satisfaction with their neighborhood and the number of deteriorated residential buildings in the neighborhoods varied depending on whether deterioration was measured at the ZIP code level ($r = .034$) or the census tract level ($r = .137$). Because ZIP codes are larger than census tracts, they are more likely to be internally heterogeneous with respect to the levels of deterioration. Because tracts are closer in size to what residents think of as their neighborhoods than ZIP codes (Coulton, et al., 2001), that heterogeneity may increase measurement error for the neighborhood-level construct, thereby weakening its correlation with residents' satisfaction.

Lack of meaningful boundaries. Defining neighborhoods as geographic units within discontinuous space assumes neighborhoods have sharply defined, fixed boundaries that are meaningful and recognizable. If that were true, the boundary problem (an aspect of the MAUP) would not be as pressing because there would be a good foundation for identifying which of the many possible boundary systems yielded the neighborhood units most relevant to residents. But if neighborhoods are indeed such discrete entities, people should agree on where the boundaries between them actually lie.

As stated previously, even residents who live close together often disagree about the size and boundaries of their neighborhood and resident defined boundaries rarely

match census boundaries (Coulton et al., 2004; Coulton et al., 2001). For example, Coulton et al. (2001) found that neighborhood maps drawn by residents were roughly similar in size to census tracts, but typically contained parts of two census tracts and three block groups. They also found that different residents drew maps with unique boundaries, in part because most residents thought of their home as the center of their neighborhood.

Coulton et al.'s (2004; 2001) findings are the not the only ones that conflict with the HLM assumption that census-derived boundaries correspond to meaningful boundaries for geographic neighborhoods. A neighborhood's perceived size can vary considerably even among people who agree on its name, indicating that the presence of a shared symbolic identity does not induce agreement on boundaries (Lee & Campbell, 1997). The results of cognitive mapping research show that geographic places do not have the kind of sharp, fixed boundaries assumed to exist under a discontinuous view of space (Montello, et al., 2003). So part of the boundary problem is that we need to group residents and measure neighborhood characteristics within boundaries that are psychologically meaningful to residents, but administratively defined units like census tracts and block groups impose artificial boundaries that may not be relevant.

These problems with the validity of census-based neighborhood boundaries have serious implications for neighborhood research because of the boundary problem. Without a meaningful natural boundary system, researchers have little basis for deciding which of the many alternative ways to divide a study region into neighborhoods generates the most appropriate set of neighborhood units and every option can conceivably lead to different answers about what effects (if any) neighborhoods have on residents.

Part of the problem is that the boundaries are used to group residents, presumably with each group representing a set of people who all consider each other neighbors and who all consider people outside the group non-neighbors. Using boundaries that are misaligned with residents' notions of the geographic area of their neighborhoods may group together residents that do not consider themselves neighbors and separate residents who do think of each other as neighbors. Similarly, it may increase measurement error in neighborhood characteristics because measures would be based on only part of the geographic area relevant to a resident while other relevant parts might be excluded. That is particularly likely to occur with people living on the edges of the neighborhood units. These are serious conceptual problems with defining neighborhoods as fixed geographic areas. In contrast, defining neighborhoods as buffers surrounding each resident's location, as may be done in GSM, never leaves anyone living at the edge of his or her own neighborhood because the neighborhood is defined relative to the individual's location, rather than via an absolute position in space.

The lack of support for neighborhoods being discrete entities calls into question the assumption that all neighborhood characteristics should be measured within the same boundary. While some contextual constructs probably do naturally have fixed, known boundaries, such as social policies that apply within the jurisdiction of a governmental agency (assuming that the jurisdiction boundaries match those of the neighborhood units), there is little reason to expect that the most relevant boundaries for all contextual conditions will match those of the researcher's chosen set of neighborhood units or of any other available fixed boundary system (Guo & Bhat, 2007; O'Campo, 2003). For example, block groups are poorly suited to capturing the spatial patterns of crime because

census boundaries often run along the centerlines of streets that are the loci for crime hotspots, which means that hotspots can get bisected such that crimes on one side of the street get assigned to one block group and those on the other side to a different block group despite the fact that they are part of coherent spatial grouping of crimes (McCord & Ratcliffe, 2007). Yet, a person living on or near that street might legitimately be affected by or concerned about all of the crimes in a hotspot.

Similarly, the impact of pollution generated by a factory is unlikely to be confined to the census tract where the factory is located and unlikely to affect everyone in either the host tract or other nearby tracts equally due to the location of factories along major transportation routes that serve as tract boundaries and because prevailing wind patterns affect the dispersal of pollutants (Downey, 2006). Thus, tracts would be a poor approximation of local neighborhoods for the purpose of measuring pollution levels, even if they are excellent for measuring other contextual characteristics.

Ignoring spatial proximity. Another major problem with defining neighborhoods as places in discontinuous space is that this ignores the broader spatial context within which residents live. One of the implicit assumptions in many neighborhood studies is that neighborhoods are self-contained, independent settings representing “intact social systems, functioning as islands unto themselves” (Sampson, 2004, p. 164). This meshes neatly with the HLM assumption that the units at the highest level of analysis are statistically independent of each other (Hofmann et al., 2000; Raudenbush & Bryk, 2002), but it means that most HLM analyses ignore the spatial arrangement of neighborhoods with respect to each other. In effect, it asserts that only the context in one’s own neighborhood unit influences outcomes.

However, there are important social, economic, and institutional ties that link residents from different neighborhood units and can create forms of spatial dependence that argue against this idea that neighborhoods are independent (Sampson, 2004). For example, the fact that some neighborhoods are close together while others are far apart matters because physical proximity is an important factor in predicting the number of social trips between different census tracts: people make more frequent trips to tracts that are close to their own tract than to more distant tracts (Wheeler & Stutz, 1971). Furthermore, assuming that only the conditions within a resident's own census tract or block group matter ignores the fact that people frequently cross the boundaries between such units when commuting to work, shopping, or attending religious services (Sastry, et al., 2002). That challenges the idea that discrete neighborhood units represent the best approximation of the relevant neighborhood setting for any given resident because residents may be exposed to conditions, events, and social processes from other nearby neighborhood units in addition to the those from their own unit.

There are statistical methods that permit modeling spatial influences between neighborhoods (Bailey & Gatrell, 1995; Haining, 2003), but they are designed for studies where all the data come from the neighborhood level of analysis, not for multilevel studies. The few HLM-based neighborhood effects studies that have considered whether outcomes in one neighborhood are influenced by conditions in adjacent neighborhoods have found that surrounding neighborhoods do indeed matter (Caughy, Hayslett-McCall, & O'Campo, 2007; Morenoff, 2003; Morenoff, Sampson, & Raudenbush, 2001; Swaroop & Morenoff, 2006). But, researchers have mostly had to apply GIS methods to the neighborhood-level residuals from HLM analyses to test those hypotheses because HLM

lacks a method for incorporating spatial autocorrelation between neighborhoods at the highest level of analysis (O'Campo, 2003). The scarcity of HLM studies that have looked for such spatial effects suggests that the discontinuous view of geographic space required by HLM leads researchers to treat neighborhoods as independent places, which de-emphasizes thinking about spatial relationships between them.

Ignoring spatial variability in contextual conditions. When a contextual characteristic is measured at the neighborhood level and used in HLM analyses, it implies that contextual conditions are identical for all residents of that unit. But the spatial distribution of contextual characteristics within a neighborhood is often not that uniform (Roosa, et al., 2003). As an example, consider racial composition, which is often measured by the percentage of residents who belong to a particular minority group, such as African Americans (Quillian & Pager, 2001; Sampson & Raudenbush, 2004). If 20% of the population in a particular census tract are African Americans, that does not mean that this is true on every individual face block in that tract: residential segregation occurs even on a relatively small spatial scale (Grannis, 1998), so there is likely to be within-tract variation in racial composition that would be ignored in studies using tracts as neighborhood units. Such within-neighborhood spatial variability in contextual characteristics may be important when the spatial scale on which residents are sensitive to that characteristic is smaller than the neighborhood units selected by the researcher. Thus, this problem is tied to other issues related to spatial scale, to which we now turn.

Poor handling of spatial scale issues. There is a conceptual aspect to the scale problem associated with the MAUP. Recall that the scale problem is that as the size of the neighborhood units used to subdivide a region changes, the variances of the

contextual characteristics change, as do their correlations with other variables. The net result is that the statistical conclusions about neighborhood effects may depend on the size of the neighborhood units chosen. Using a single, fixed definition of neighborhood boundaries suffers from a scale problem because it assumes that all contextual conditions vary on the same geographical scale and that the neighborhood units are in fact the right size to best capture each and every contextual effect (Guo & Bhat, 2007).

If the chosen neighborhood units are too small or too large relative to the actual geographical scale on which a particular construct actually matters, the spatial patterns may be obscured and estimates of the relationships between contextual conditions and outcomes may be biased, possibly leading to erroneous statistical inferences (Lery, 2008). Using the example of crime hotspots and pollution, it is reasonable to expect that a smaller geographical scale would be more appropriate for considering effects of crime hotspots on nearby residents whereas a larger scale may be more appropriate for the effects of pollution given the spread via prevailing winds.

Clearly, using a single set of neighborhood units (e.g., block groups) does not permit researchers to measure different contextual conditions at different spatial scales. The obvious solution to that dilemma would be to use multiple levels of neighborhood units with different contextual characteristics measured at each level. Several researchers have argued that neighborhood is a multilevel concept and that residents can and do distinguish between multiple spatial scales at which their neighborhoods could be described (Galster, 2001; Kearns & Parkinson, 2001; Suttles, 1972). For example, Suttles (1972) proposed a multilevel conceptualization of neighborhood that integrates social and spatial aspects to define neighborhoods at four distinct spatial scales. Starting with the

face-block⁵ on which one lives then moving up to successively larger spatial scales he called the defended neighborhood, the community of limited liability, and the expanded community of limited liability. Kearns and Parkinson (2001), simplified Suttles' conceptualization by trimming it down to three scales (home area, locality, and urban district or region) and argued that each is loosely coupled with a predominant function for residents. So, there is conceptual support for the idea that neighborhoods exist at multiple spatial scales and that each spatial scale may be important to residents, but for different reasons. However, some of the spatial scales in these two conceptualizations are vaguely defined and little research has been done that would allow researchers to argue that specific census-based geographic units correspond to the different spatial scales described by these authors.

Despite this conceptual support for a multilevel representation of neighborhoods, most neighborhood studies still use only a single level of geographic units to represent neighborhoods (Beyers et al., 2003; Caughy et al., 2008; Caughy & O'Campo, 2006; T. E. Duncan et al., 2003; Franzini et al., 2005; Rankin & Quane, 2000; Sampson et al., 1997; Sunder et al., 2007). So, while HLM could easily represent neighborhoods with multiple levels of geographic units (e.g., block groups nested within census tracts) so that different characteristics could be measured at each level, this is simply not common practice. Instead, all of a given neighborhood's characteristics are usually measured within the same geographic boundary. Few authors discuss why they choose not to use multiple levels of geographic units, but feasibility issues probably influence that decision. One such issue is that using multiple levels of geographic units would increase sample

⁵ A face-block is comprised of both sides of a street bounded on either end by cross-streets.

size requirements, making it more costly to collect data. Finally, even if the use of multiple levels of geographic neighborhood units increased, HLM-based studies would still be limited in their ability to examine the spatial scale on which neighborhood effects operate because they would still rely on the dubious assumption that at least one of the available levels of units is the right size to capture the effect of interest.

Ideally, theory and findings reported in the literature should guide researchers in selecting the spatial scales on which specific neighborhood-level factors should be measured (Messer, 2007), but there is a dearth of research that directly addresses this issue. Careful consideration of the pathways through which neighborhood conditions are believed to influence resident outcomes might allow researchers to extract clues about the relevant spatial scale from studies of the spatial aspects of related phenomena such as social networks and urban social travel (Greenbaum, 1982; Greenbaum & Greenbaum, 1985; Stutz, 1973; Wellman, 1996; Wheeler & Stutz, 1971) or travel to activities such as grocery shopping and commuting to work (Sastry, et al., 2002). However, research that directly investigates the spatial scale on which contextual factors operate by comparing the results of statistical models which differ only in the way neighborhoods are operationalized for measurement purposes would be far more valuable (Chaix, et al., 2006; Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian, et al., 2005; Kruger, 2008; Meersman, 2005).

Summary. The problems noted above (the MAUP, the boundary problem, and the scale problem) are interrelated. They derive from the fact that a narrow conceptualization of neighborhoods that presumes they are adequately represented by a single set of spatial boundaries leads to operational definitions that are sometimes poorly aligned with the

ac

ac

er

m

ne

pi

Bu

ma

con

con

up

As

con

con

mai

imp

Dov

scri

usef

even

decli

actual nature of the phenomena under study. However, to use HLM, researchers must adopt a discontinuous conceptualization of space that treats neighborhoods as discrete entities and constrains how neighborhoods are operationalized for the purpose of measuring neighborhood characteristics.

The discontinuous view of space also encourages researchers to view neighborhoods as independent of one another by focusing narrowly on neighborhoods as places and de-emphasizing the fact that they are embedded in a larger spatial context. But, if the phenomena we are modeling are not so neat and tidy, we need to adopt modeling tools that fit better with empirical reality and accommodate a more flexible conceptualization of neighborhoods. Conceptualizing neighborhoods as places within continuous space offers researchers a way to begin addressing these problems by opening up new ways to operationalize neighborhoods when measuring contextual characteristics. As shall become clear below, GSM methods are compatible with this more flexible conceptualization of neighborhoods, but HLM is not.

Neighborhoods as places in continuous geographic space. Adopting a conceptualization of geographic space that emphasizes its continuous, connected nature makes distance between locations (and sometimes direction) relevant and decreases the importance of potentially arbitrary boundaries between discrete units like census tracts. Downey (2006) argued that while a discrete view of space is sometimes practical, a more sophisticated perspective recognizes that continuous representations of space are also useful because the social impact or sphere of influence for various goods, objects, or events is often not confined to the boundaries of units such as census tracts and usually declines continuously as distance from them increases. Similarly, when contextual

characteristics exhibit substantial spatial variability within the boundaries of units like census tracts, treating them as continuous, contoured surfaces stretching across the study region may be useful (Chaix et al., 2006; Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian et al., 2005; Downey, 2006).

Loosely speaking, a resident's neighborhood is a place that occupies some subset of the geographical space surrounding his or her home. The relationship between neighborhoods and space is less constrained when space is conceptualized as continuous rather than discontinuous. Discarding the notion that neighborhood boundaries subdivide space into mutually exclusive areas allows neighborhoods to partially overlap (Coulton, et al., 2004), to have boundaries that are somewhat elastic or fuzzy and depend on the purpose for drawing a boundary (Coulton et al., 2004; Montello et al., 2003; Sastry et al., 2002), and to be defined in ways that are more consistent with findings suggesting that residents tend to see their own homes as the center of the neighborhood (Coulton, et al., 2001). Another advantage of adopting a continuous view of geographic space is that we can think of residents as belonging to multiple overlapping neighborhoods, such that their homes are simultaneously located at the center of some neighborhoods and more peripherally located with respect to other neighborhoods. That lets us use a dramatically different method of grouping residents for the purpose of characterizing spatial variation in outcomes than is used in HLM studies, where each resident is a member of only a single neighborhood.

These points deserve further attention because they relate to how well a researcher can align the definition of neighborhoods as units of analysis with both the nature of real world phenomena and formal statistical representations in HLM and GSM. As will be

clarified further below, they are closely tied to why the GSM approach may sometimes be more appropriate than HLM for testing hypotheses about neighborhood effects.

Conceptual definition of neighborhood. George Galster proposed the most comprehensive and useful conceptual definition of neighborhood for the purposes of the present study: “Neighbourhood is the bundle of spatially based attributes associated with clusters of residences, sometimes in conjunction with other land uses” (Galster, 2001, p. 2112). Galster elaborates on that definition, listing many different types of neighborhood attributes that can be tied to geographical places. He makes it clear that this definition encompasses any and every construct that can be measured within a spatially bounded area, but that it does not restrict those attributes to all use the same set of boundaries and does not require neighborhoods to occupy mutually exclusive geographic areas. Accordingly, it may include elements of the physical environment such as the structural characteristics of nearby buildings and roads, local levels of pollution, noise, or traffic. It can also include measures of the demographic composition or aggregate socioeconomic characteristics of the resident population within that area, the kinds of public services or programs available, the quality of schools, proximity to employment opportunities, or the social policies that may be in place. This definition also extends to levels of citizen participation, sense of community, collective efficacy, place attachment, the presence and attributes of social networks, and more.

Making boundaries more meaningful. Galster’s (2001) definition is fully consistent with multilevel conceptualizations of neighborhoods as described above, but more useful for GIS-based statistical models of neighborhood effects that consider both place and space than other available definitions because it recognizes the potential for

ambiguity in the geographic boundaries of a neighborhood. Under this conceptualization, the geographic area of a neighborhood would be unambiguous only if all its spatial attributes happened to vary on the same spatial scale and also happened to follow identical boundaries (Galster, 2001; Guo & Bhat, 2007). This definition is useful because it accommodates the possibility that different attributes of a resident's neighborhood may need to be measured within different boundaries, perhaps with some measured over larger areas than others or within areas of different shape, but all of which encompass the resident's home. The way neighborhoods are defined in HLM studies is simply a special case within this broad definition in which the boundaries produce mutually exclusive, non-overlapping geographic areas containing residents' homes and where the same boundaries are applied to measure multiple neighborhood attributes. Galster's definition encompasses that option, but also allows additional possibilities and, as a consequence, researchers can better address the boundary problem described in the previous section.

To elaborate on that point, consider the catchment areas for local public schools: they are relatively fixed bounded areas defined by governmental agencies, so children from families living within any given catchment area would all attend the same public school. As such, a measure of the quality of the local public school in a resident's neighborhood should be the same for everyone in that catchment area. In contrast, there is no clear, inherent fixed boundary that defines the geographic area over which many other contextual variables should be measured. So, for a contextual variable like crime, which is known to be very unevenly distributed over space (Block, 2000; Ratcliffe & McCullagh, 1999; Taylor, 1998) and where people may be affected by crime occurring near their home regardless of which block group or tract it occurs in, it might make more

sense to measure crime within a circular buffer centered on each person's home. Galster's definition and GIS-based modeling techniques allow these distinct bounding strategies to co-exist in the same study.

Using buffers as neighborhood boundaries. Using a buffer is consistent with the research suggesting that residents tend to think of their homes as being located at the center of the neighborhood (Coulton, et al., 2001) and thus also with the fact that different residents tend to report different neighborhood boundaries. In addition, it allows the researcher to flexibly increase or decrease the size of the area over which crime would be measured, which is necessary if one wishes examine the spatial scale on which crime matters (see Figure 2 for an illustration of this point). This would accomplish two other things as well. First, it would allow a researcher to assign an individual resident to different but overlapping geographical neighborhoods for purposes of measuring different contextual variables such as crime and school quality. Second, it would also allow researchers to assign residents in the same school catchment area to different neighborhoods for the purpose of measuring crime, though those neighborhoods might overlap to some degree (or not at all) depending on the distance between the residents and the size of the buffer.

This notion of using circular buffers, sometimes referred to as "sliding neighborhoods" (Guo & Bhat, 2007) or "bespoke neighborhoods" (Galster, 2008), is consistent with Galster's conceptualization of neighborhood and can be implemented in a GSM analysis but not within an HLM analysis. To continue with the crime example, a resident living inside a school catchment area that has low crime but close to the boundary between that catchment area and one that has high crime might end up with a

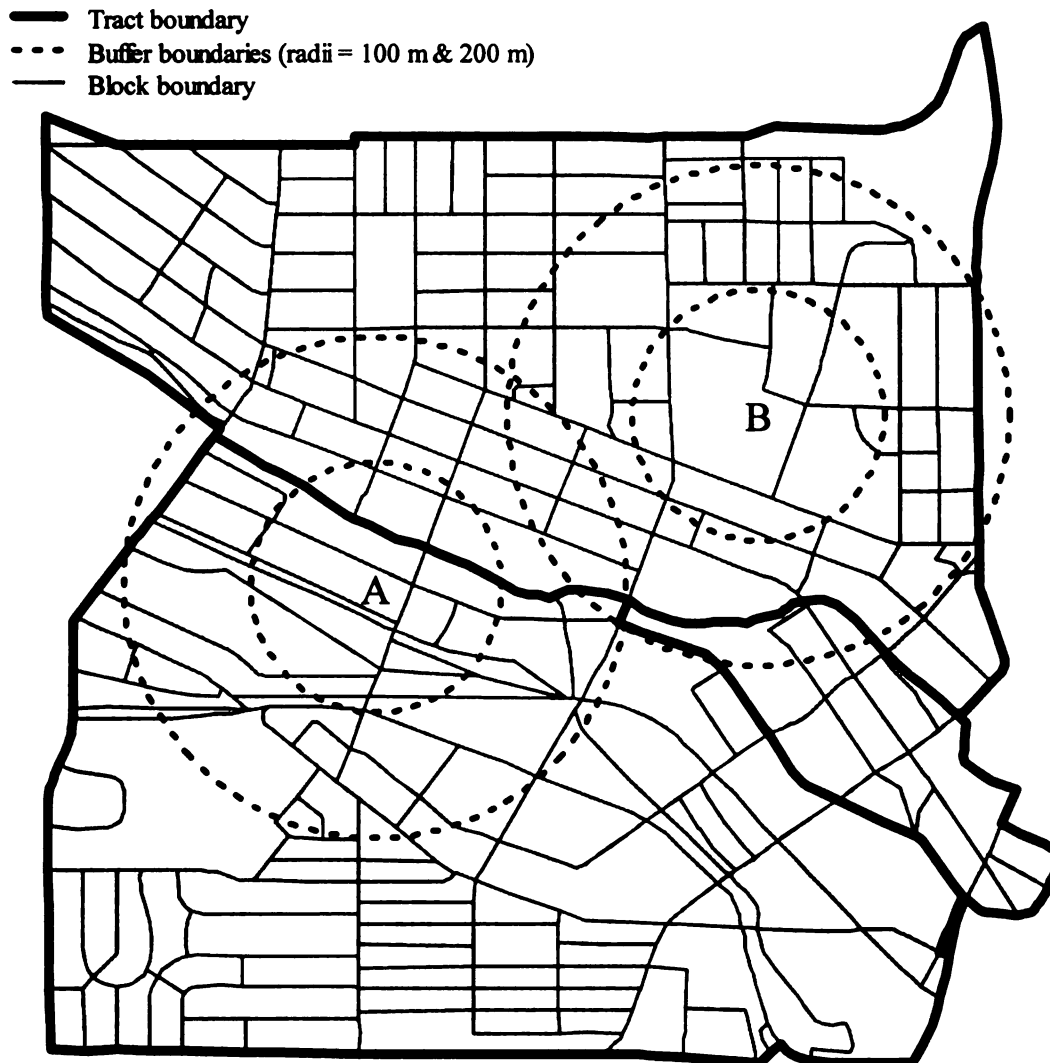


Figure 2: Illustration of using buffers as neighborhood boundaries. Contextual characteristics could be measured within circular buffers of different sizes (e.g., with radii of 100 m and 200 m) centered on points A and B, which are shown in relation to census tract and block boundaries in Battle Creek, MI. For point A, both buffers overlap portions of multiple tracts. For point B, only the larger buffer overlaps portions of more than one tract. The 200 m buffers for these two points partially overlap. Source: Map produced by the author from GIS files prepared by the U.S. Census Bureau (U.S. Census Bureau, 2007a, 2007c).

buffer-based measure of crime that captures some of the crimes occurring on the other side of the border that may well affect that resident. Using fixed boundaries for neighborhoods does not permit modeling such boundary-spanning contextual effects, but a buffer approach does. So while adopting a continuous view of space permits using

buffer approaches to defining neighborhoods for some variables and fixed boundaries for others, the discontinuous view of geographic space is incompatible with buffer approaches to measuring neighborhood context because the buffers for residents might sometimes only partially overlap.

Addressing spatial proximity. Taking a continuous view of space enables researchers to take the spatial arrangement of both neighborhoods and residents into account in their analyses. Because residents are viewed as members of multiple, overlapping neighborhoods, GSM analyses do not assume that neighborhoods are statistically independent of one another. Instead, physical proximity becomes a key element in detecting and modeling spatial variation in outcomes. GSM analyses effectively treat neighborhoods as places that have fuzzy boundaries for the purpose of grouping residents; the farther a resident is from the center of a particular neighborhood, the less similar his or her outcomes are likely to be to those of another resident located at the center of the neighborhood. This is one of the ways that GSM improves on what HLM analyses can do with respect to accounting for both place and space.

Addressing spatial variation in contextual conditions. Another important advantage offered by buffer-based techniques for measuring neighborhood contextual variables is that it enables GSM techniques to better handle spatial variation in contextual conditions within things like census tracts. Residents on different ends of a large tract might easily be assigned different values for contextual characteristics like crime because the buffers centered on their homes may not even overlap, or may only overlap partially, leading data aggregated within them to yield different values for the two residents. One can even think of buffer-based techniques as yielding estimates of a smooth, continuous

surface describing the spatial distribution of a contextual variable if we simply imagine estimating the value of that variable at a grid of densely-packed points covering the entire study region. This offers a much more informative view of how contextual conditions vary over space than just measuring conditions within a single set of geographic units like block groups or census tracts.

Addressing issues of spatial scale. The ability to use different boundaries to define the relevant neighborhood area for measuring different neighborhood characteristics also helps to deal with the scale problem described earlier. Rather than being limited to selecting a single spatial definition of a neighborhood (or a narrowly defined hierarchical representation like block groups nested within census tracts), thereby perhaps forcing the researcher to measure one or more contextual characteristics at a geographic scale that is not well suited to the actual construct, Galster's (2001) conceptualization recognizes that different neighborhood characteristics might need to be measured at different scales. If the relevant characteristic is associated with some known and meaningful pre-defined areal unit (like a school catchment area), that can be used as the boundary for that construct. However, for other constructs, one can use buffers and can even vary the size of the buffers used for the constructs. Researchers can also use buffers of different sizes for the same construct and compare models to empirically examine which size exhibits the best statistical performance. That then opens the door to generating theories that can explain why particular constructs operate on specific geographical scales of measurement.

Summary. Conceptualizing neighborhoods as places within continuous space is useful because it creates new options for operationalizing neighborhoods. Those options

are more compatible with GSM than with HLM. Adopting this more flexible conceptualization of space and neighborhoods will allow us to adopt modeling tools that may be better aligned with the phenomena we wish to investigate. This study proposes applying a statistical method that does just that – and to the extent that GSM outperforms HLM in modeling the data, it implies that treating neighborhoods as discrete, non-overlapping entities for the constructs studied in this project may be less effective or accurate than an approach that recognizes that neighborhoods are not so precisely bounded. We need to align the methods with the phenomena, rather than allowing the methods to drive the definition of the units of interest. To continue developing the rationale for this study, the next section of the literature review describes how HLM is applied to neighborhood research and how the discontinuous conceptualization of space and neighborhoods affects HLM analyses.

Hierarchical Linear Modeling Methods for Testing Contextual Effects

To build the argument for why GSM may be a useful alternative to HLM, it is essential to understand how HLM works and some of its limitations associated with the assumptions made in order to apply it to neighborhood research. There is a large literature on HLM, so this section focuses only on the aspects and issues most relevant to the present study. For simplicity, the examples are drawn from the neighborhood effects literature or framed in terms of neighborhood research, although HLM has also been applied to many other content areas.

While the statistical methods for investigating contextual effects have evolved considerably in recent decades, interest in multilevel research questions is hardly new. One of the early statistical approaches to this was simply to merge contextual variables

into an individual-level dataset by assigning the same value of each contextual variable to everyone from a given setting. Then, that contextual variable was added to an ordinary least squares (OLS) regression model as a “cross-level operator” to test the effect of interest (James & Williams, 2000). However, the OLS regression model assumes that all observations are independent from one another (Fox, 1997), so this initial approach was eventually criticized for failing to take account of the fact that the observations from people in the same setting are in fact not independent (Raudenbush & Bryk, 2002; Roosa et al., 2003). The most serious effect of violating the independence assumption underlying the OLS model is that it results in overly optimistic estimates of the significance of contextual effects (Raudenbush & Bryk, 2002; Roosa, et al., 2003).

To better illustrate that point, consider a hypothetical study focusing on whether or not neighborhood poverty affects educational achievement among youths. To do such a multilevel study, one might choose a set of neighborhoods and then collect data about multiple youths from each neighborhood. Simply using neighborhood poverty as a predictor in an OLS regression model on the full sample in that study will yield inaccurate estimates of the effect of neighborhood poverty if educational outcomes for youths in the same neighborhood tend to be more similar to each other than they are to outcomes among youths from other neighborhoods. This phenomenon is referred to as autocorrelation; it indicates that the residuals are correlated rather than independent.

When the assumption of independent errors is met, each person’s data contributes unique statistical information to the analysis, but when it is not met that is no longer true: part of the information gained from each observation overlaps with information obtained from other observations in the same neighborhood. The net result is that the effective

sample size is really smaller than the number of persons in the sample. Because OLS regression does not correct for that, the standard errors for the regression coefficients end up being too small and Type I error rates are inflated. The Type I error rate gets worse as the degree of autocorrelation increases.

Why HLM is useful. Part of HLM's appeal lies in the ease with which we can match the units and levels of analysis from our theories to formal statistical representations—there is a very clear mapping from the terminology of theory onto the terminology used in the analysis (Luke, 2005). Indeed, HLM is a statistical method that was tailor-made for doing multilevel research. The techniques that fall under the broad umbrella of HLM were developed to deal with situations where data are grouped hierarchically, with the units of analysis at one level nested within larger, more inclusive units that represent higher levels of analysis (Gelman & Hill, 2007; Raudenbush & Bryk, 2002). In the HLM framework, the levels of analysis are usually numbered, starting with level 1 at the lowest or most micro level of analysis then incrementing the level number at each higher level of analysis added to the research design. For example, Browning and Cagney (2002) used HLM to analyze data from a sample of 8,782 residents (level 1) nested within 343 neighborhoods (level 2) in Chicago, Illinois.

There are both statistical and theoretical reasons why researchers use HLM. A key statistical reason for its adoption is that HLM extends the OLS regression model to allow for non-independence (i.e., autocorrelation) between the data from people nested within the same higher level sampling unit (Raudenbush & Bryk, 2002). For neighborhood research, that would mean that outcomes among people from the same neighborhood can be autocorrelated without violating the HLM assumptions, which addresses one of the

key criticisms of simply adding a cross-level operator to an OLS regression model. While correcting coefficient standard errors for the influence of autocorrelation is a key feature of HLM, there are also compelling theoretical reasons to use it in neighborhood research. HLM provides a way to explore the multilevel structure of the data and examine a wide variety of substantive hypotheses associated with multilevel theories.

One set of reasons that HLM may be useful is that it allows one to detect how much variance in outcomes can be attributed to neighborhoods as opposed to individuals and interpret the implications of that variance. First, the amount of autocorrelation present in the data has substantive meaning: it quantifies how much variation in the outcome can potentially be attributed to neighborhood-level as opposed to individual-level differences (Merlo, 2003; Merlo, Chaix et al., 2005a). Second, comparing results from alternative HLM models can help sort out whether neighborhood-level variation is a result of compositional or contextual effects by showing how much neighborhood-level variability can be explained by geographical clustering of similar people within the neighborhoods (compositional effects) and how much variability can be explained by neighborhood-level contextual characteristics (C. Duncan, et al., 1998; Merlo, 2003). That helps researchers avoid attributing variability to contextual effects that can be adequately explained by individual-level effects. Third, quantifying the relative amounts of within- and between-neighborhood variability in an individual-level outcome makes it easier to interpret the substantive meaning and policy implications of the regression coefficients associated with contextual variables (Merlo, 2003).

Like other forms of regression modeling, HLM allows one to test hypotheses about the effects of specific predictors. For example, HLM allows researchers to test

theoretical propositions about whether contextual characteristics have direct (or indirect) cross-level effects on outcomes for residents (Merlo, Chaix, et al., 2005b). HLM can also be used to control for neighborhood effects in order to obtain more accurate estimates of the effects of individual-level predictors. Consistent with the idea in HLM that variability is of primary interest, HLM allows one to test whether the effect of an individual-level predictor varies across neighborhoods, indicating that an as-yet unidentified characteristic of the neighborhood as a whole moderates the relationship (Merlo, Chaix, et al., 2005b; Merlo, Yang, et al., 2005). Finally, researchers can add cross-level interactions to a model to test whether specific contextual moderator variables explain variability in individual-level regression coefficients (Merlo, Chaix, et al., 2005b).

The HLM statistical model. Although there are several different ways to write out the statistical model underlying an HLM analysis (Gelman & Hill, 2007), the most intuitive form is specified by writing multiple equations (Raudenbush & Bryk, 2002). For example in a model where level 1 units are residents and level 2 units are neighborhoods, there would be a level 1 sub-model showing the relationships between individual-level predictors and the outcome, plus one or more level 2 equations describing how specific coefficients in the level 1 sub-model (including the intercept) are themselves outcomes predicted by level 2 variables. For example, a simple HLM with one predictor at each level might be written as shown in Equations 1-3 below:

$$\text{Level 1 sub-model: } Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij}. \quad (1)$$

$$\text{Level 2 sub-model for the intercept: } \beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j}. \quad (2)$$

$$\text{Level 2 sub-model for the slope of } X: \beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j}. \quad (3)$$

In Equation 1, Y_{ij} is the outcome for person i in neighborhood j , which is shown as the sum of a neighborhood-specific intercept (β_{0j}), plus the neighborhood-specific effect (β_{1j}) of a level 1 predictor (X_{ij}), plus a level 1 residual (r_{ij}) for person i in neighborhood j . The fact that the level 1 intercept and slope are estimated separately for each neighborhood is important because that means one can now construct a new outcome variable from each of them, then use the level 2 sub-model in Equations 2 and 3 to predict those coefficients with neighborhood-level predictors. Thus, Equation 2 shows that the intercept in each neighborhood (β_{0j}) can be modeled as the sum of a fixed intercept that is the neighborhood-level mean (γ_{00}), plus the systematic effect (γ_{01}) of some neighborhood-level contextual variable (Z_j), plus a neighborhood-level residual (u_{0j}) that represents random error at level 2. Changing from using β to using γ to represent the regression coefficients at level 2 simply reminds readers that one has now moved to a different level of analysis.

As illustrated by Equations 2 and 3, one can write a separate level 2 sub-model for every parameter in the level 1 sub-model; in each of these level 2 sub-models, there is a unique level 2 residual term. Taken together, Equations 1-3 propose a multilevel model in which (a) the individual-level variable X has a main effect, (b) the neighborhood-level variable Z also has a direct main effect through Equation 2, and (c) Z moderates the

effect of X through Equation 3. The moderator effect is easier to recognize when Equations 2 and 3 are substituted into Equation 1 so that the entire model is represented by a single regression model (Equation 4), then simplified and rearranged with all three residual terms collected inside the parentheses in Equation 5:

$$\text{Combined model: } Y_{ij} = (\gamma_{00} + \gamma_{01}Z_j + u_{0j}) + (\gamma_{10} + \gamma_{11}Z_j + u_{1j})X_{ij} + r_{ij}. \quad (4)$$

$$\text{Combined model: } Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}Z_jX_{ij} + (u_{0j} + u_{1j}X_{ij} + r_{ij}). \quad (5)$$

The term $\gamma_{11}Z_jX_{ij}$ in Equation 5 represents the cross-level interaction between the neighborhood-level variable Z and the individual-level variable X (γ_{11} is the coefficient for that effect). This interaction functions as a moderator term, exactly paralleling how moderators are represented in OLS regression models (see Aiken & West, 1991).

The combined model in Equation 5 is also useful for illustrating how HLM is simply an extension of the more familiar OLS regression framework. The major difference is the addition of the neighborhood-level residuals for the intercept (u_{0j}) and slope ($u_{1j}X_{ij}$), and the change in notation from β to γ for the coefficients. Of course, adding the neighborhood-level residuals and allowing them to potentially correlate with each other makes estimating the model more complex, but iterative maximum likelihood methods make that tractable (Raudenbush & Bryk, 2002) and modern software packages such as SPSS, SAS, R, and WinBUGS can easily handle these models (Gelman & Hill,

2007; Hayes, 2006; Lunn, Thomas, Best, & Spiegelhalter, 2000; Peugh & Enders, 2005; West, Welch, Galecki, & Gillespie, 2007).

Assumptions in HLM. Three of the major methodological assumptions underlying HLM relate directly back to the broader assumptions in multilevel research. The first of those is that every level 1 unit is nested within a specific, known higher level unit (Hofmann, et al., 2000); in neighborhood research that simply means we need to know which neighborhood each resident lives in. Defining neighborhoods as places that occupy mutually exclusive portions of space simplifies determining who belongs in each neighborhood. The second assumption is that level 1 units are exposed to and potentially affected by processes and conditions within the higher level units to which they are linked (Hofmann, et al., 2000). In neighborhood research, residents are assumed to be affected by neighborhood processes and/or neighborhood conditions. Finally, HLM assumes that outcomes can potentially vary both within and between higher level units (Hofmann, et al., 2000; Raudenbush & Bryk, 2002).

An important, but frequently unstated, assumption in HLM as applied to neighborhood research is that contextual conditions are assumed to be homogenous within each neighborhood (Roosa, et al., 2003). This shows up in the statistical model by assigning the same value on each neighborhood-level predictor to every individual in a given neighborhood.

Naturally, there are also formal statistical assumptions in HLM, most of which follow from the fact that HLM simply extends OLS regression. The following discussion draws primarily from Hofmann et al. (2000), though similar points are made by Raudenbush and Bryk (2002). For every level 1 unit within each level 2 unit, the level 1

residuals (r_{ij}) are assumed to be independent and normally distributed with a mean of zero and a variance of σ^2 . Because there can be multiple level 2 residuals, each associated with a different level 1 parameter, each kind of level 2 residual is assumed to follow a normal distribution and be independent across level 2 units. The set of level 2 residuals are collectively assumed to follow a multivariate normal distribution, which means that the variance components associated with those level 2 residuals can be arranged in a variance-covariance matrix whose elements are labeled with the Greek letter tau (τ_{qq}). It is also assumed that the level 1 residuals are independent of any and all level 2 residuals and that neither level 1 nor level 2 residuals are correlated with any predictors at their respective levels in the hierarchy.

Autocorrelation in HLM. Quantifying the amount of autocorrelation detected by an HLM analysis is quite easy. One simply runs an empty or null model in which neither the level 1 nor the level 2 sub-models contain any substantive predictors (Raudenbush & Bryk, 2002). Instead, they contain only intercept and residual terms at each level. This provides estimates of two variance components: σ^2 , which represents within-neighborhood or individual-level error variance, and τ_{00} , which represents between-neighborhood variance. The sum of these two variance components is the total variance in the outcome measure. If there is a tendency for people in the same level 2 unit to have similar outcomes, then there must be some variance attributable to between-neighborhood differences ($\tau_{00} > 0$). To measure the amount of autocorrelation present, one calculates the ICC (ρ), which is the ratio of the between-neighborhood variance to

the total variance from an empty HLM model, as shown in Equation 6 (Raudenbush & Bryk, 2002).

$$\text{ICC: } \rho = \tau_{00}/(\tau_{00} + \sigma^2). \quad (6)$$

The simplest way to interpret the ICC is to think of it as the *proportion of variance in the outcome that is attributable to neighborhoods*. Because variance components cannot be negative numbers, the ICC ranges from zero to one (Merlo, Chaix, et al., 2005a). Obviously, when τ_{00} is equal to zero, then the ICC is also zero, indicating that the neighborhood a resident lives in is unrelated to the outcome in any way; in those cases, HLM produces results identical to OLS regression (Roosa, et al., 2003). The ICC indicates the amount of autocorrelation present: larger values indicate more autocorrelation, but even small amounts of autocorrelation can compromise the accuracy of OLS regression. At the upper end, an ICC of one indicates that there are no individual differences within neighborhoods and that all differences in outcomes can be explained by the neighborhoods in which residents live.

Because HLM approaches neighborhoods as spatial units, the autocorrelation it models can be conceptualized as a form of spatial dependence (Bass & Lambert, 2004) that is *strictly hierarchically structured*. The level 2 residuals in an HLM are the mechanism for introducing autocorrelation between residents (level 1 units) within the same neighborhood (level 2 unit) into the model. The standard HLM statistical model does not permit autocorrelation between residents located in different neighborhoods unless the entire hierarchy is extended to include a third level.

The HLM statistical model also does not permit the amount of autocorrelation that may exist between level 1 units within the same level 2 unit to vary. So, the requirement that the nesting of residents inside neighborhoods must be known combines with the operational definition of neighborhoods as units of analysis in HLM to determine whose outcomes are allowed to be autocorrelated and whose are not.

Overall, HLM treats neighborhoods as places that are disconnected from and independent of one another, unless they are connected via common membership in a still higher level of spatial unit. Even when one adds additional levels to the hierarchy, the boundaries of the units at the highest level will always define sharp discontinuities in whether outcomes for people on either side of the border will be autocorrelated despite being quite close together in space.

Controlling for composition. A question that frequently arises in multilevel neighborhood research is whether evidence that outcomes differ between residents of different neighborhoods represents a contextual effect or whether it is instead attributable to differences in the compositions of the neighborhoods' populations (Bingenheimer & Raudenbush, 2004; C. Duncan, et al., 1998). Upon detecting the presence of autocorrelation, it is tempting to conclude that a contextual property of the neighborhood is the only possible explanation. However, an alternative explanation may be that the neighborhood-level variability can be explained by the geographical clustering of similar types of people into neighborhoods—that neighborhood is effectively confounded with one or more individual-level characteristics (C. Duncan, et al., 1998; Merlo, Yang, et al., 2005). A variety of social processes might result in that sort of clustering, some voluntary (wealthy individuals choosing to live in certain neighborhoods) and others involuntary

(structural racism may restrict the housing options available to minorities, leading to residential segregation).

The ICC calculated from an empty HLM model does not by itself distinguish contextual from compositional effects that may explain that autocorrelation. To control for composition, it is necessary to run another HLM analysis that expands the empty model by adding only individual-level predictors known or believed to be related to the outcome (Merlo, Yang et al., 2005). Doing that controls for the composition of the population in each neighborhood, at least insofar as those particular variables are concerned and the variance components will now reflect that the effects of those individual-level variables have been removed.

The adjusted ICC based on the revised model represents the amount of autocorrelation remaining in the data that may reflect the influence of contextual factors at the neighborhood level (Merlo, Yang, et al., 2005). If the adjusted ICC still indicates the presence of autocorrelation, then controlling for composition has not eliminated the possibility of contextual effects. However, if the ICC is effectively zero after controlling for composition, then there is no neighborhood-level variance left to explain and adding contextual characteristics will not be fruitful. Formulas for calculating the proportional change in variance at each level of the model are available (Merlo, Yang et al., 2005), allowing the researcher to calculate level-specific analogues to R^2 .

Neighborhoods as level 2 units in HLM. There are two important ways that the conceptualization and operationalization of neighborhoods as units of analysis influences an HLM analysis. The first is that HLM is only compatible with a definition of neighborhoods that presumes there are sharply defined, non-overlapping boundaries that

make it possible to unambiguously assign neighborhood membership for each resident. Without that, HLM's mechanism for modeling autocorrelation breaks down.

One problem with this approach is that the boundaries chosen by the researcher (which are one option among many) may not be valid and may not group people in a meaningful way. For example, a researcher might choose to use census tracts to group people into neighborhoods, but this may group the wrong people together and split up people who should be grouped with each other. Consider a hypothetical case where a researcher assigns two people who consider themselves neighbors to different neighborhood units: social interactions and information exchange between those people might lead to autocorrelation in their perceptions of neighborhood sense of community, norms, or safety, thus violating the HLM assumption that outcomes for residents in what the researcher calls different neighborhoods are independent. Furthermore, in spatial datasets, the degree of autocorrelation observed is often a function of the distance between observations (Bailey & Gatrell, 1995; Haining, 2003). This form of spatial autocorrelation is succinctly described by Tobler's First Law of Geography, which states "Everything is related to everything else, but near things are more related than distant things." (Tobler, 1970, p. 236). So, assuming that autocorrelation is hierarchically structured may be inaccurate, depending on how well the researcher's boundary system captures the actual patterns in the data.

Another side-effect of this need to unambiguously group people into neighborhoods is that findings will be subject to the MAUP through its implications for the measurement of neighborhood-level constructs. The discontinuous view of geographic space implies that researchers should use the same boundaries when

determining the geographic area within which all neighborhood-level constructs should be measured. The frequent use of census tracts or block groups as neighborhood units often derives from a desire to use readily available census data to measure structural characteristics of the neighborhoods so that they can be used as contextual predictors. However, doing that assumes that size, shape, and boundaries of the neighborhoods are fixed and equally appropriate for all of those measures, thus opening the door to the MAUP because of the boundary and scale problems.

Considering space in HLM. As previously noted, the standard software packages for HLM provide very few options for trying to take account of the spatial arrangement of neighborhoods. One option is to add an additional hierarchical level to the model, then model regional effects with variables at the new level. However, that is not a very flexible approach for considering spatial issues, so some neighborhood researchers are now beginning to perform alternate kinds of analyses that attempt to take the spatial arrangement of neighborhoods into consideration. One approach involves using results from HLM analyses as inputs to spatial regression models that are entirely conducted at the neighborhood level (Morenoff, 2003; Morenoff et al., 2001; Swaroop & Morenoff, 2006). In these studies, the question typically being asked is whether the mean level of the outcome in a focal neighborhood is influenced by the contextual characteristics of adjacent neighborhoods.

The modeling required to answer that question proceeds in two stages. In the first stage, an HLM is run in the normal fashion to link contextual and individual-level predictors to some level 1 outcome, yielding estimates of the neighborhood-level residuals. Those residuals represent each neighborhood's mean on the outcome (adjusted

for the composition of the sample in each neighborhood). The neighborhood means are expressed as deviations from the grand mean across all neighborhoods.

In the second stage of these analyses, the neighborhood-level residuals become the dependent variables in a subsequent “spatial lag” regression model of the general form illustrated in matrix notation by Equation 7 (Morenoff, 2003; Morenoff, et al., 2001; Swaroop & Morenoff, 2006).

$$Y = \rho WY + X\beta + \varepsilon. \quad (7)$$

In Equation 7, the parameter ρ is a spatial autoregressive parameter that represents the effect of a one unit change in the weighted average of the dependent variable in surrounding neighborhoods. The W in this equation is a weights matrix defining how much each other neighborhood’s value on the dependent variable contributes to the weighted average that is denoted WY . One typically specifies weights so that only the neighborhoods that share a common border or corner with the focal neighborhood directly affect the spatial lag term for the focal neighborhood. Naturally, $X\beta$ represents the effects for contextual variables associated with the focal neighborhood and ε is a typical regression residual that is assumed to be independent and normally distributed.

Because the Y values from surrounding neighborhoods used to construct the WY term are themselves functions of the contextual conditions in their respective neighborhoods, the spatial lag model essentially says that the outcomes in a focal neighborhood depend not only on the contextual characteristics within its own boundaries but also on the contextual characteristics of other neighborhoods. More distant neighborhoods often exert increasingly weaker influence mediated through their effect on

m

c

b

e

to

at

cc

De

ap

in

th

So

th

P

a

e

th

in

E

S

C

more proximal neighborhoods, though that depends on exactly how the weight matrix is constructed in this model.

While this approach does allow researchers to build on the foundation provided by **HLM** to start considering spatial issues, it would be better to integrate these spatial **effects** directly into the original level 2 portion of the HLM model. One way to do that is to **replace** the **WY** term in Equation 7 with a single variable **Y'** that represents a weighted **average** of the outcome in surrounding neighborhoods that has been purged of any **correlation** with the error term, then treat the new variable as simply another **neighborhood-level** predictor (Land & Deane, 1992; Wyant, 2008). An advantage of that **approach** to enhancing HLM with additional spatial information is that it can be **implemented** even with standard multilevel modeling software. The disadvantage with **this approach** is that implementing it relies on an instrumental variables framework that is **somewhat** difficult to understand and it still requires running additional models prior to the **main** analysis in order to construct the **Y'** variable.

Taking a fully Bayesian approach and switching to more specialized software **packages** like WinBUGS (Spiegelhalter, Thomas, Best, & Lunn, 2007) allows one to take a **more** sophisticated approach to enhancing an HLM model with spatial information. For **example**, some authors have proposed relaxing the assumption of independence among the **neighborhood-level** intercepts by adding a conditional autoregressive (CAR) structure into the HLM model (Beard, 2008; W. Browne & Goldstein, in press). A CAR-HLM **model** is simply a slightly more general version of the HLM model where one assumes **that that** the proximity between neighborhoods is important because the average resident **outcome** from one neighborhood is similar to the average outcomes in other nearby

neighborhoods. That makes the CAR-HLM model conceptually similar to a spatial lag model. Using CAR-HLM models, researchers can allow the correlation between the intercepts of adjacent neighborhoods to depend on the distance between neighborhoods (Beard, 2008; W. Browne & Goldstein, in press). One advantage of moving to a CAR-HLM model instead of following up on a standard HLM model with a separate spatial lag model is that the entire CAR-HLM model is estimated at one time.

Like the spatial lag approach, adding a CAR structure to an HLM requires a weight matrix (W) that defines which neighborhoods affect each other and how much they do so. A CAR-HLM analysis is implemented by adding a spatially structured residual term to the level 2 model to supplement the regular unstructured residual term. As Equation 8 below shows, each of these new spatial residuals (denoted S_i) is assumed to be drawn from a normal distribution with a mean equal to the weighted average of the spatial residuals from the surrounding neighborhoods (denoted S_j), which are defined by the W matrix values that have non-zero values. This also adds a second level 2 variance component to the model.

$$S_i | S_{-i} \sim \text{Normal}(\sum_j S_j / n_i, v / n_i), \text{ where } n_i = \sum_j W_{ij}. \quad (8)$$

Only one study has used a CAR HLM approach to study residents' perceptions of their neighborhoods (Fagg, Curtis, Clark, Congdon, & Stansfeld, 2008). Unfortunately, that study did not contrast the CAR HLM results with those of HLM models without the CAR structure. While it did not consider many of the other spatial issues related to the definition of neighborhoods discussed above, Fagg and colleagues' study shows that it is possible, though certainly very far from common, to enhance HLM to better account for the spatial arrangement of neighborhoods.

However, the GSM approach also offers a simple, direct approach to modeling the spatial patterns in outcomes and to representing the possibility that contextual conditions in what HLM would call different neighborhoods may matter for outcomes in a focal neighborhood. With GSM, there is no need to pull the results out of HLM and feed them into another statistical procedure because it has mechanisms for representing these concepts. In addition, GSM offers greater flexibility than HLM with respect to defining the boundaries to be used for measuring neighborhood-level variables. Thus, the review now turns to a discussion of GSM.

Geostatistical Modeling Methods for Testing Contextual Effects

The GSM methods used in this study belong to a family of related statistical techniques that have their origin in the earth sciences, namely geology. This study uses the term *geostatistical model* to honor that origin and to maintain the link back to the larger literature where the method is most often used and described, which is usually called geostatistics (Banerjee, Carlin, & Gelfand, 2004; Chilès & Delfiner, 1999; Diggle & Ribeiro, 2007; Goovaerts, 1997; Isaaks & Srivastava, 1989). Retaining that link should help other researchers interested in exploring the range of GSM methods available to locate relevant materials. Other authors who have begun to use GSM techniques outside of its original application areas have also retained this terminology (Bass & Lambert, 2004; Chaix, et al., 2006; Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian, et al., 2005)

GSM is one of three major types of statistical approaches to spatial data analysis that fall under the broad umbrella of GIS (Bailey & Gatrell, 1995). It is informed by an explicit spatial perspective on analysis where place, space, and spatial dependence are of

primary interest (Haining, 2003) and relies on point-referenced data—each observation is associated with a location defined by a pair of spatial coordinates (Bailey & Gatrell, 1995; Haining, 2003). GSM focuses on modeling the spatial distribution of an attribute or **variable** attached to those locations, generally conceptualizing the observations as **samples** from some underlying, continuous surface (Bailey & Gatrell, 1995; Chilès & Delfiner, 1999; Goovaerts, 1997; Haining, 2003; Isaaks & Srivastava, 1989).

Geostatistical techniques were developed to study the spatial distributions of **minerals** and natural resources (Chilès & Delfiner, 1999; Goovaerts, 1997; Isaaks & Srivastava, 1989). They are often used to predict the amount of a particular mineral **expected** at unsampled locations on the basis of both the large-scale spatial trends and **small-scale** spatial autocorrelation evident in the data from sampled locations. Such **models** often use the spatial coordinates (or polynomial functions of them) as predictors in **regression** models to represent the large-scale spatial trends, but they may also use **substantive** predictors like the concentration of another mineral that was also measured at the **sampled** locations and is believed to predict the levels of the target mineral.

In traditional applications of GSM like those discussed above, the point is to **predict** values at unsampled locations, not to interpret the substantive meaning of the **coefficients**. However, GSM, like HLM, is an extension of the OLS regression model (Banerjee et al., 2004; Diggle & Ribeiro, 2007). It can be used in an explanatory capacity **because** the distinction between prediction and explanation is tied to the purpose of the **research**, not how a regression model works (Diggle & Ribeiro, 2007; Myers, 1990). To **use** GSM for explanatory purposes one must replace spatial coordinates as predictors

with substantive predictors selected on the basis of theory because the former merely *describe* spatial patterns in the dependent variable, while the latter *explain* them.

Why GSM is useful. This study focuses on GSM as an alternative to HLM for **studying** neighborhood effects because GSM offers greater flexibility to incorporate and **model** spatial aspects of those phenomena. Furthermore, GSM is compatible with a **multilevel** conceptualization of neighborhoods as entities that surround each resident's **home** and have fuzzy, sometimes overlapping boundaries that may vary in size and shape **depending** on the neighborhood attribute being measured (Galster, 2001).

Despite its origins in modeling physical phenomena, GSM can be applied to study the **social** phenomena that are of interest in neighborhood research. GSM has recently **been applied** in epidemiological studies of place effects on health and health care **utilization** (Boyd, Flanders, Addiss, & Waller, 2005; Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian et al., 2005). Geostatistical methods have even been applied **in a limited** way to examine spatial autocorrelation in urban youths' perceptions of **neighborhood** disorder (Bass & Lambert, 2004).

As with HLM, there are both statistical and theoretical reasons why GSM is **useful** for neighborhood research. On the statistical side, GSM is designed to explicitly **model** spatial autocorrelation between observations, allowing the regression coefficients **associated** with predictors to be accurately estimated in situations where the OLS **regression** assumption of independence would be violated (Banerjee et al., 2004; Diggle & Ribeiro, 2007).

On the theoretical side, GSM allows the researcher to test many of the same **multilevel** hypotheses that HLM can test. For example, GSM can partition the amount of

variance in outcomes into components attributable to neighborhood-level spatial variation and individual-level non-spatial variation (Banerjee, et al., 2004). By adding individual-level predictors, GSM can account for population composition effects and yield revised neighborhood- and individual-level variance estimates in a manner parallel to that in HLM (Chaix, Merlo, Subramanian, et al., 2005).

As an extension of the basic regression model, GSM also allows one to test hypotheses about the effects of predictors, whether those predictors are located at the neighborhood or individual levels of analysis. Any hypothesis about direct or indirect cross-level effects of contextual predictors that can be represented in a regression model or in HLM can be tested in similar manner in GSM, as can hypotheses about cross-level interactions between contextual and individual-level predictors.

GSM relies on a continuous representation of space rather than on one fragmented into spatial units of arbitrary size, shape, and boundaries (Bass & Lambert, 2004; Chaix et al., 2006; Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian et al., 2005). Thus, in GSM places are naturally embedded in the larger spatial context because spatial proximity, rather than membership in the same neighborhood unit, is the basis for modeling autocorrelation (Banerjee et al., 2004; Chaix et al., 2006; Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian et al., 2005; Finley, Banerjee, & Carlin, 2007). The advantages of that approach are explained further below in the section about how autocorrelation is handled in GSM. The continuous representation of space is also useful because it allows a researcher to independently vary the size and shape of the area over which each neighborhood-level predictor is measured. The significance of that

flexibility is discussed further below in the sections on neighborhoods as level 2 units in GSM and on considering space in GSM.

The GSM statistical model. The statistical model underlying GSM is an **extension** of the standard regression model that relaxes the independence assumption associated with ordinary least squares models. Several other statistical models also relax that **assumption**. For example, where the units of observation are geographic areas, **spatial** regression models such as simultaneous autoregressive (SAR) models and **conditional** autoregressive (CAR) models are frequently used to incorporate non-independence into what is otherwise a regular regression model (Haining, 2003). With **point-referenced** data, generalized least squares (GLS) models can be used to add a **covariance** structure to the model's error term, thereby modeling residual spatial **dependence** (Bailey & Gatrell, 1995). In fact, GLS models are closely related to the GSM **method** discussed here. The GSM model used in this study is shown in Equation 9 below. Throughout the model, the (s) attached to various terms denotes that they are associated with **known** spatial locations—it is like a subscript indexing the data by spatial location.

$$Y(s) = X^T(s)\beta + W(s) + \varepsilon(s). \quad (9)$$

As shown in Equation 9, the dependent variable Y at a particular location (s) is **predicted** from an intercept and a set of spatially referenced predictors from the same **location**, denoted in matrix notation as $X^T(s)\beta$, plus a normally distributed residual term **called** $\varepsilon(s)$ that represents pure random error at that location. The only difference between **the** GSM model and the standard OLS regression model is that GSM adds a spatial **random** effect residual called W(s) that represents spatial autocorrelation (Banerjee et al., 2004; Finley et al., 2007). The value of W(s) for any given observation depends on the

observation's location and hence also on its distance from the other observations. Unlike *HLM*, GSM does not use different notation for the coefficients associated with predictors located at the neighborhood versus individual levels of analysis.

The notation used to identify the variance components in GSM is different than in *HLM* because it switches the meaning of the symbols τ^2 and σ^2 . In *HLM*, τ_{00} refers to between-neighborhood variance and σ^2 refers to individual-level within-neighborhood variance. In GSM, the situation is reversed because τ^2 (also called the *nugget*) traditionally refers to the non-spatial, pure error variance that is basically individual-level variance and σ^2 (also called the *partial sill*) refers to spatial variance that is analogous to neighborhood-level variance (Banerjee, et al., 2004).

Assumptions in GSM. As with all statistical methods, GSM makes both methodological and statistical assumptions. One of the key methodological assumptions is that the location associated with every observation is known. Galster's (2001) definition of neighborhoods implies that every residential location exists inside a neighborhood, so contextual characteristics measured at such locations provide the neighborhood-level information required to use GSM as a multilevel analysis technique. Neither Galster's neighborhood definition nor GSM methods require assuming that neighborhoods have a constant size or shape that should be used for measuring all neighborhood characteristics. GSM can also accommodate neighborhoods that overlap.

GSM, like *HLM*, assumes that residents are exposed to and potentially affected by neighborhood processes and conditions (Chaix et al., 2006; Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian et al., 2005). It also assumes that outcomes can vary

both within and between neighborhoods, but it does not rely on neighborhood boundaries to determine how much autocorrelation may exist between observations. Instead, GSM assumes that autocorrelation is a spatial phenomenon operating over continuous space.

There are several formal statistical assumptions in GSM. Because many of the assumptions in GSM are similar to those in OLS regression models, this discussion focuses on the specialized assumptions that are pertinent to understanding how GSM works. As in OLS regression, the random error term $\epsilon(s)$ is assumed to follow a normal distribution with a mean of zero (Banerjee et al., 2004; Finley et al., 2007), though the symbol for the variance component of this distribution is τ^2 in GSM rather than the σ^2 traditionally used in OLS regression and HLM. Though GSM and OLS regression use different symbols for the error term, there is no substantive difference in what they represent. As usual, the errors denoted with $\epsilon(s)$ are assumed to be independent of other terms in the model, just like the individual-level residuals from an HLM.

The spatially autocorrelated term $W(s)$ in the GSM model is assumed to be the result of a stationary Gaussian spatial process (Banerjee et al., 2004; Finley et al., 2007). That means that the GSM approach is conceptualizing these residuals as values sampled from a joint multivariate normal (Gaussian) distribution where each observed location is associated with a separate, normally distributed variable that has a mean of zero. Stationarity refers to the assumption that the mean value of the process is zero everywhere in the study region.

Describing the full multivariate normal distribution of the spatial process requires a covariance matrix with rows and columns that correspond to the observed locations. Each element in the matrix is therefore associated with two locations and represents the

covariance between the two random variables from which the residuals associated with those two locations are drawn. GSM assumes that the covariance matrix has been generated by an underlying *covariance function* that specifies the shape of a smooth theoretical curve that models the amount of covariance between observations at any two locations as a function of the physical distance separating them. GSM authors also sometimes refer to this as the correlation function because the covariance function can be converted into a correlation metric that is more interpretable.

Typically, the correlation function associated with the $W(s)$ term is assumed to be stationary and isotropic (Banerjee et al., 2004), which means that strength of spatial autocorrelation depends only on distance between observations and not on the direction one would have to travel to move from one location to reach the other. The correlation function is said to be anisotropic when the direction from one point to another affects the level of autocorrelation observed (Banerjee et al., 2004; Chilès & Delfiner, 1999; Diggle & Ribeiro, 2007; Isaaks & Srivastava, 1989).

Autocorrelation in GSM. In contrast to HLM, where the autocorrelation in the data is represented by a single number, autocorrelation is not a single value in GSM. Instead, the correlation function describes how much autocorrelation there is between points as a function of the distance between them. It is estimated by grouping pairs of observations separated by certain distances, then estimating the variance in each group. Each observation contributes to multiple groups because it lies at different distances from various other observations. Observations that are close together are usually more highly autocorrelated than observations that are far apart.

Figure 3 shows examples of several alternative correlation functions that can be used in GSM (Banerjee et al., 2004; Chilès & Delfiner, 1999; Diggle & Ribeiro, 2007; Isaaks & Srivastava, 1989), each of which is described by a different mathematical formula that has a few parameters, usually consisting of a partial sill parameter (σ^2), a range parameter (ϕ), and the nugget (τ^2). In Figure 3, the dot at distance = 0 indicates that each model assumes perfect autocorrelation between observations at the exact same location (distance = 0), whereas the lines indicate the level of autocorrelation between observations that are at different locations (distance > 0). The vertical space between the dot and the left end of the line in each semivariance panel illustrates the nugget parameter. The effects of the range and partial sill parameters are easiest to see in the semivariance panel for the spherical model: the vertical space between the left end of the line and the level at which the line turns flat is the partial sill, whereas the distance at which that line first turns flat is the range⁶. Details on alternative correlation functions are presented in geostatistics textbooks (Banerjee et al., 2004; Chilès & Delfiner, 1999; Diggle & Ribeiro, 2007; Isaaks & Srivastava, 1989), but a full discussion of the types of correlation functions available is beyond the scope of this study, other than to note that it is common practice to fit alternative GSMs with different correlation functions and assess which one produces the best empirical results.

To quantify the amount of autocorrelation detected by a GSM analysis, one can use the parameters of the correlation function estimated from an empty GSM model (i.e.,

⁶ Technically, the parameter that determines the range often actually measures the rate of decay in the spatial covariance and the range corresponds to the distance at which that covariance has become negligible. So, the range is actually a value calculated by transforming the true parameter value.

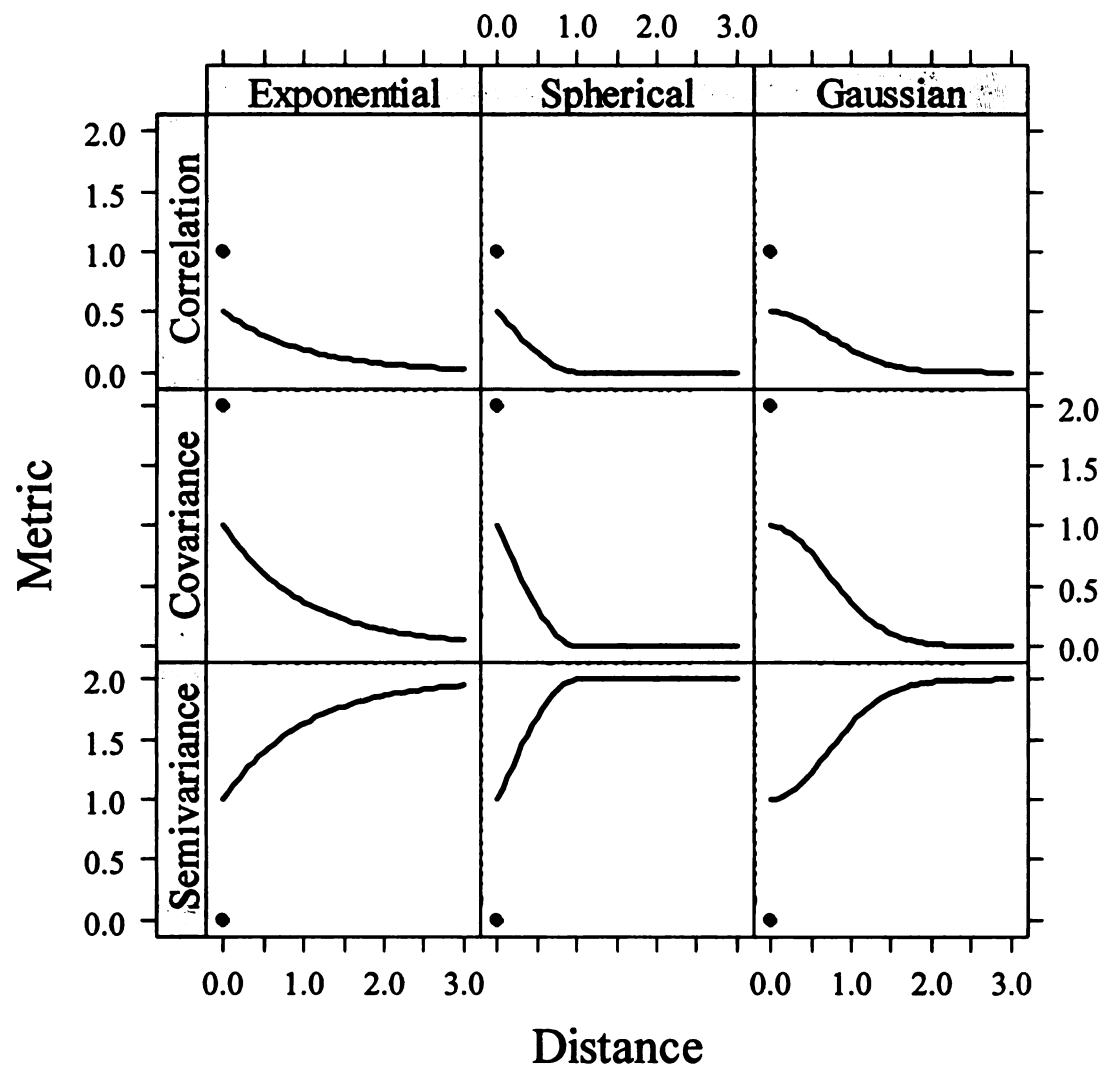


Figure 3: Exponential, spherical, and Gaussian variogram models displayed in three different metrics. Each panel illustrates the autocorrelation between observations as a function of the distance between them. The top two rows show these models in correlation (top) and covariance (middle) metrics, which are measures of similarity; the bottom row shows them in a semivariance metric, which is a measure of dissimilarity. All three models have the same partial sill ($\sigma^2 = 1$), range ($\phi = 1$), and nugget ($\tau^2 = 1$) parameters, but differ in shape.

one with no substantive predictors, only an intercept term). Recall that the partial sill (σ^2)

represents the amount of neighborhood-level, spatial variance, while the nugget (τ^2)

represents the individual-level, non-spatial variance. The range parameter in a GSM

analysis allows the researcher to identify the actual distance beyond which data are no longer spatially autocorrelated (or only negligibly so). Both the partial sill and the nugget are variance components, so their sum is the total variance in the outcome. Thus, we can use Equation 10 to construct a partial sill ratio (PSR) that is conceptually similar to the ICC measure used in HLM.

$$\text{PSR: } \rho = \sigma^2 / (\sigma^2 + \tau^2). \quad (10)$$

The PSR is the maximum level of autocorrelation observed in the GSM, which occurs at very short distances between observations. Like an ICC, it varies between zero and one, with one indicating perfect spatial autocorrelation. Thus, the PSR estimated in a GSM is directly comparable to the ICC estimated in an HLM: The simplest way to interpret it is to think of the PSR as the *proportion of variance in the outcome variable that is attributable to neighborhoods*. In addition to looking at the PSR, one can plot and examine the entire correlation function associated with the GSM (Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian et al., 2005).

Incorporating the spatially autocorrelated residual term $W(s)$ in the GSM model corrects for the autocorrelation in the data and accounts for the spatial arrangement of both residents and neighborhoods, resulting in more accurate standard errors for the regression coefficients. This parallels how HLM corrects for autocorrelation, but makes different assumptions about how autocorrelation is structured because it does not use neighborhood boundaries to inform the statistical representation of autocorrelation.

In substantive terms, the range associated with the correlation function in a GSM analysis identifies the geographic scale on which the spatial autocorrelation in the residuals exists. Models with large ranges indicate that residual spatial autocorrelation

exists over long distances, while small ranges indicate spatial autocorrelation is restricted to relatively short distances. This is critical information because it clarifies the conditions under which HLM may be a reasonable method to deal with spatial autocorrelation, which is probably when the range of spatial autocorrelation roughly spans the entire length of the typical HLM-based neighborhood unit and the units are far enough apart that spatial autocorrelation is not spilling over from one to another.

Consider some contrasting hypothetical situations. In one, the neighborhood units used in an HLM are all far enough apart that any distance-based spatial autocorrelation that could have been detected by GSM does not reach from one neighborhood to another (e.g., neighborhoods drawn from different cities or states). In that case, HLM's assumption that outcomes for residents located in different neighborhoods are independent is met and modeling the within-neighborhood autocorrelation as a single neighborhood-level residual shared by all residents of the neighborhood may not be sacrificing too much information if the individual-level predictors account for any remaining patterns in the spatial distribution of the outcome within the neighborhood.

In another hypothetical situation, the neighborhood units used in the HLM are close enough together that the range of spatial autocorrelation detectable by GSM reaches from some neighborhoods into other neighborhoods, indicating that outcomes among residents in different neighborhood units are still autocorrelated. That would decrease the between-neighborhood variability detected by the HLM and underestimate neighborhood effects. This is where GSM has the most potential to provide a better method for dealing with that autocorrelation because it models an aspect of the data that HLM cannot.

In the final hypothetical scenario, the range of spatial autocorrelation might be short compared to the typical size of the HLM neighborhood units. In this situation, HLM may not detect much autocorrelation because the neighborhoods are too large and thus effectively pool the data for people who are far enough apart to be uncorrelated with data for people who are close enough to be correlated. That would increase the within-neighborhood variance detected by the HLM, thereby decreasing the ICC. In short, GSM might outperform HLM whenever the range of the spatial autocorrelation present is not well matched to the size of, and spacing between, the neighborhoods used in HLM.

Controlling for composition. Because variations in population composition can explain spatial variability in outcomes that might otherwise be attributed to contextual effects, composition effects are of concern in GSM for the same reason they are a concern in HLM. Fortunately, the approach to controlling for composition is the essentially the same between the two methods: one simply adds individual-level predictors to the model then examines the adjusted PSR to see how much spatial variability remains that might be related to contextual factors (Chaix et al., 2006; Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian et al., 2005). If there is still autocorrelation after controlling for composition, then adding contextual predictors may explain the remaining spatial variability. Because the PSR and the ICC are comparable measures, the formulas used to calculate the proportional change in variance between alternative HLM models (Merlo, Yang et al., 2005) should also be applicable in GSM and allow calculation of level-specific analogues to the R^2 statistic used in OLS regression modeling.

In the case of GSM, one might also look at whether the range of spatial autocorrelation has changed after accounting for composition. Plotting the correlation function for both the empty model and the model including individual-level effects (Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian et al., 2005) may be useful.

As in HLM, each regression coefficient in a GSM model is adjusted for all other predictors in the model. Therefore, contextual predictors should be added after relevant individual-level predictors when there are theoretical or empirical reasons to expect that residents' personal characteristics are related to the outcome being studied.

Neighborhoods as level 2 units in GSM. One of the major ways in which GSM differs from HLM is in how neighborhoods are represented. Above, two ways in which the conceptualization and operationalization of neighborhoods affected HLM were discussed. Next, we revisit those issues to describe how they affect GSM.

First, GSM does not rely on grouping residents into neighborhoods to model autocorrelation. Each resident is associated with a location by the spatial coordinates that identify his or her position in geographic space. Then, autocorrelation is modeled as a function of the distance between residents' locations, thereby treating geographic space as a continuous phenomenon. To the extent that Tobler's First Law of Geography (Tobler, 1970) holds for a particular outcome measure, this may be a better way to represent spatial autocorrelation in resident outcomes than the one used in HLM.

Some authors have argued that GSM is not subject to the MAUP because it does not rely on grouping residents into bounded neighborhood units (Bass & Lambert, 2004; Chaix et al., 2006; Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian et al., 2005). Unfortunately, this is not entirely true because neighborhood boundaries are also

used for measurement purposes. Simply put, the second issue is that to measure neighborhood characteristics, one still must select the geographic area over which they should be measured—and that requires setting boundaries (though the boundaries used can vary for different neighborhood characteristics in GSM).

GSM offers more flexibility than HLM for this because while it can use things like census tract boundaries to operationalize neighborhoods for the purpose of measuring constructs, it is not restricted to doing so. Unlike HLM, GSM permits researchers to use buffers (sliding or bespoke neighborhoods, Galster, 2008; Guo & Bhat, 2007) to measure contextual conditions in areas centered on residents' homes, which is more consistent with the idiosyncratic and egocentric way residents think about their own neighborhoods (Coulton et al., 2004; Coulton et al., 2001; Lee & Campbell, 1997; Montello et al., 2003). Because different contextual predictors do not need to be measured within the same size buffer, GSM provides great flexibility to customize how neighborhoods are operationalized for the purpose of measuring each neighborhood-level construct. That makes GSM highly compatible with the conceptual definition of neighborhoods adopted for this study, which emphasizes that different neighborhood characteristics may need to be measured within different boundaries (Galster, 2001).

One advantage of GSM's ability to use buffers is that each resident can have unique values on neighborhood-level measures. Only residents at the same location would have identical buffer boundaries and therefore identical values on contextual variables. Aggregating data to measure a contextual variable for partially overlapping buffers would lead to similar but not necessarily identical values for that variable (with

greater overlap producing more similar values), while values for buffers that do not overlap at all could be quite different indeed.

Another advantage of buffers is that they can be easily adapted to measure conditions at different geographic scales. For example, the studies by Chaix and colleagues varied the size of the buffer used for their neighborhood income measure to capture the mean income of the nearest 100, 200, 500, 1000, and 1500 residents (Chaix, Merlo, Subramanian et al., 2005); they also tried buffers with radii of 500 m, 750 m, and 1000 m to measure crime (Chaix et al., 2006). Similarly, other researchers are also exploring the use of buffer methods, coupled with varying the size of the buffer, to measure contextual conditions (Guo & Bhat, 2007; Kruger, 2008; Kruger, et al., 2007; Meersman, 2005).

Measuring conditions within buffers may be particularly useful when (a) contextual conditions exhibit spatial variability within the administrative units that are typically used as proxies for neighborhoods in HLM studies, (b) there is no reason to believe that the boundaries of units like census tracts are relevant to the contextual characteristic being measured, or (c) none of the administrative units available for use in HLM match the geographic scale on which a particular contextual condition matters.

Considering space in GSM. There are several important spatial issues that GSM can address in neighborhood research. First, it can quantify the geographical scale on which spatial autocorrelation exists in the data, which is reflected in the estimated range associated with the correlation function. Second, GSM takes the spatial arrangement of the residents and neighborhoods into account by using the location of each observation to estimate a neighborhood-level spatially autocorrelated residual for each resident. Third,

running GSM analyses that differ only in the size of a buffer used to measure a particular contextual factor enables researchers to empirically determine what geographic scale of measurement is most appropriate because statistics like the deviance information criterion (DIC) can be used to select the model with the closest fit to the data (Chaix et al., 2006).

The ability to expand or shrink a buffer measure of contextual conditions may provide a more elegant solution to the question of whether outcomes in a focal neighborhood are affected by conditions in other nearby neighborhoods. Recall that HLM cannot directly address this question, so some researchers have pursued such questions by extracting the neighborhood-level residuals from HLM analyses and used them in “spatial lag” regression models conducted entirely with neighborhood-level data (Morenoff, 2003; Morenoff, et al., 2001; Swaroop & Morenoff, 2006). That approach may be viewed as an indirect method of asking whether the spatial scale on which the contextual condition of interest operates is really larger than the size of the neighborhood units that were used in HLM. The GSM approach to answering such a question is very straightforward: just measure those conditions over a larger area by increasing the size of the buffer, then compare the GSM results to those from a model with a smaller buffer. This would retain the multilevel nature of the data, reduce the number of different statistical procedures that need to be applied, and more directly answer the question about the geographic scale on which the targeted characteristic matters most.

Comparing HLM and GSM Approaches

The sections above described HLM and GSM, highlighting major features of each approach. As described above, the view of space underlying these two techniques sets the stage for a number of key differences between these methods that affect how we can

apply them in neighborhood research. Although some of those differences have been discussed briefly by other authors, the prior literature has not offered a comprehensive conceptual comparison of HLM and GSM. Therefore, Table 1 synthesizes material from previous sections of this literature review to present a concise, side-by-side comparison of HLM and GSM along several conceptual dimensions that are relevant to studying neighborhood effects.

Clearly, there are important conceptual differences, but what is known about empirical differences in their performance and the scientific findings they yield? So far, only three studies have directly compared HLM and GSM analyses by applying both techniques to a single dataset (Boyd et al., 2005; Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian et al., 2005). All three were epidemiological studies, with one

Table 1: Conceptual comparison of HLM and GSM

Dimension	HLM	GSM
View of space	Discontinuous.	Continuous.
Type of neighborhood boundaries	Only fixed boundaries are possible.	Either fixed or buffer-based boundaries can be used.
Overlapping neighborhoods	Hierarchical overlap allowed, if using ≥ 3 levels of analysis.	Any form of overlap is allowed.
Spatial proximity	Proximity effects are generally ignored in this approach. They can be added, but doing so takes extra effort.	Proximity effects are intrinsic to and explicitly modeled in this approach.
Structure of autocorrelation	Hierarchical. Observations are grouped using neighborhood boundaries. Each observation can only belong to one group at any given level of analysis.	Spatial. Pairs of observations are grouped into bins as a function of the distance between them. Each observation contributes to many bins through its pairing with other observations at varying distances.
Measure of autocorrelation	Intraclass correlation (ICC).	Partial sill ratio (PSR).
Spatial scale of autocorrelation	Spatial scale is only indirectly quantified (if authors describe the size of the neighborhood units).	Spatial scale is directly quantified by variogram parameters.
Options for varying spatial scale of neighborhood-level measures	Options are limited by the size of the available neighborhood units. Measures usually all use a shared boundary for each neighborhood unit. Values can vary only between units.	There are many options (buffers can be defined at arbitrary sizes). Boundaries can be customized for each measure. Values can vary continuously over space when using buffers.

focusing on infectious disease among Haitian students (Boyd et al., 2005), one focusing **on** healthcare utilization in France (Chaix, Merlo, & Chauvin, 2005), and the other **focusing** on substance abuse diagnoses in a city in Sweden (Chaix, Merlo, Subramanian, **et al.**, 2005). Those studies show that GSM, like HLM, can be adapted to analyze binary **outcomes**; the resulting models are related to the GSM model described above in the **same** way that logistic regression is related to OLS regression.

Comparing models of infectious disease in Haiti. Boyd et al. (2005) were **interested** in spatial variations in the prevalence of a mosquito-borne parasitic infection in **a Haitian** community. Their sample consisted of 5- to 11-year old students from 57 **schools** in a contiguous geographic area covering approximately 400 km². They used **school** tuition (as a measure of local area's SES), whether or not the school offered a **nutrition** program, altitude, and topographic zone (plains, foothills or mountains) as **contextual** measures. To illustrate that HLM does not fully control for spatial **autocorrelation**, Boyd et al. compared results from non-spatial and spatial variations of **hierarchical** logistic models. While their spatial models are not identical to the GSM **model** described above, they are part of the larger category of GSM techniques whose **key** feature is the inclusion of spatially autocorrelated residuals.

The key finding in the Boyd et al. study (2005) was that the spatial models **Produced** coefficients that were somewhat smaller than those from the corresponding **HLM** analyses, but they did not report indices of overall model fit. The attenuated **coefficients** in their spatial models may be an artifact of the way spatial autocorrelation **was** handled. In particular, initial variogram modeling suggested that their outcome **variable** was spatially correlated up to about 2.15 km, but a limitation in WinBUGS (the

software they used to estimate their spatial models) required them to adopt a conditional autoregressive structure in their spatial models that treated data from schools within 4.35 km of one another as spatially correlated because every school had to have at least one “neighbor” and this longer distance was the minimum at which every school met that criterion. This may have diluted the value of spatial modeling by treating schools that were relatively far apart as if they were just as highly correlated as schools that were close together. They presented a figure showing that the neighborhood-level residuals from their HLM showed some residual spatial structure, but did not quantify how much spatial autocorrelation remained in the HLM results.

Comparing models of health care utilization in France. The second study that has directly compared HLM and GSM examined whether a nationwide sample of residents from France had regular primary care physicians and whether they had used specialist physicians at more than half of their doctor visits in the last year (Chaix, Merlo, & Chauvin, 2005). Their sample represented over 3000 municipalities nested within 340 larger units called broad areas that were scattered throughout France. The contextual factors of interest were a measure of local SES (percentage of residents with minimal education), the supply of primary care physicians, and the supply of specialist physicians.

They measured contextual factors by aggregating data within municipalities in one HLM, and within broad areas in another HLM. They calculated buffer-based measures of contextual factors for their GSM analyses (Chaix, Merlo, & Chauvin, 2005). For local SES, they used a buffer with a radius of 37.5 km; for the supply of physicians, they used a buffer with a 50 km radius. These buffers were somewhat larger than the broad area units. In both cases, the contextual measures were weighted such that data

closer to the center of the buffer contributed more to the measure than data further out toward the edge of the buffer. They used the scaled deviance statistic to compare model fit among the empty versions of the HLM and GSM models, finding that while the broad area HLMs fit better than the municipality HLMs, the GSMs uniformly fit better than either of the corresponding HLMs (Chaix, Merlo, & Chauvin, 2005). In a series of GSM analyses, they also found that the contextual variables had consistently stronger effects when measured within buffers than when measured within municipalities or broad areas. So, their comparison between GSM models that differed only in how the neighborhoods were defined for measurement purposes showed that using buffers produced better statistical results than using HLM-style neighborhood units.

In addition, testing with Moran's *I* indicated that the level 2 residuals in the HLM models were spatially autocorrelated (Chaix, Merlo, & Chauvin, 2005). They argue that HLMs systematically overestimated the significance of contextual effects because the residual spatial autocorrelation in the HLM results leads to inappropriately small standard errors for exactly the same reason that ignoring hierarchical autocorrelation leads to inappropriately small standard errors in OLS regression.

Comparing models of substance abuse disorders in Sweden. Chaix, Merlo, Subramanian et al. (2005) compared HLM and GSM techniques by studying the relationship between neighborhood mean income and risk of substance abuse disorders in a city in Sweden. Their sample consisted of all persons aged 40-59 living in the city they were studying. The city was divided into 100 administratively defined neighborhoods with a median area of 0.5 square km, which became the level 2 units in their HLM analyses. They measured neighborhood income within those administrative units for their

HLM and GSM analyses. For their GSM analyses, they also measured neighborhood income within spatially adaptive buffers. The buffers were centered on residents' homes, but did not have a constant spatial radius. Instead, they were scaled to contain a constant number of residents (the nearest 100, 200, 500, 1000, or 1500 residents). This meant that the buffers were physically larger in more sparsely populated areas.

They conducted HLM and GSM analyses in three steps, starting with empty models that contained no substantive predictors, then adding individual-level predictors, and then adding the neighborhood income measure (Chaix, Merlo, Subramanian, et al., 2005). The GSM models consistently fit their data better than the corresponding HLM models, as indicated by lower DIC values (DIC is a Bayesian measure of model fit, see Spiegelhalter, Best, Carlin, & van der Linde, 2002). This was true when neighborhood income was measured within the same administrative areas used in the HLM, but the effect of neighborhood income was stronger when measured within buffers that were smaller than the administrative areas. Indeed, the strength of the effect was inversely proportional to size of the buffers (using the smallest buffer yielded the strongest effect). The odds-ratios for neighborhood income were quite similar between the two techniques, though the GSM model had slightly wider confidence intervals. Adding individual- and neighborhood-level predictors explained substantial amounts of the neighborhood-level variance in both the HLM and GSM models. In the HLM models, adding the substantive predictors caused a decrease in the residual spatial autocorrelation detected in the neighborhood-level residuals, while adding those predictors to the GSM models reduced both the level and range of spatial autocorrelation.

Summary. The literature comparing HLM and GSM approaches to modeling neighborhood effects on the health and behavior of residents is quite small. The studies reviewed above suggest that GSM can, at least with some kinds of data, produce statistical models that fit better than HLM even when contextual variables are measured within the same boundaries for both techniques (Chaix, Merlo, Subramanian, et al., 2005). Furthermore, they also suggest that GSM analyses based on measuring contextual variables within buffers of appropriate size can yield stronger effects than are observed in HLM analyses based on measuring those same variables within discrete geographic units (Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian, et al., 2005). Finally, they also suggest that when the spatial scale on which a contextual measure operates is larger than the units used in a corresponding HLM, level 2 HLM residuals still contain unmodeled spatial autocorrelation (Boyd, et al., 2005; Chaix, Merlo, & Chauvin, 2005).

Previous studies have not fully discussed how comparing HLM and GSM analyses can help us update how we conceptualize and think about neighborhoods. Broadly speaking, the way we conceptualize neighborhoods informs two aspects of neighborhood studies, (1) how we group residents in order to detect spatial variability and model autocorrelation in outcomes, and (2) how we define the geographic area that should be used when measuring neighborhood context. While these aspects are nearly inextricably intertwined in HLM, the GSM approach offers the possibility of dissociating and examining them separately. Both of these aspects can and should be explored when comparing these two methods, but previous work has focused more on the latter aspect.

The next section of the literature review describes the nature of the substantive phenomenon being used to compare HLM and GSM in this study. Introducing the

substantive constructs at this point sets the stage for the following section, which describes how the present study fills specific gaps in the literature, presents arguments for why GSM may be a better alternative than HLM, and then links the research questions for the study to specific hypotheses that can be tested to inform our thinking about neighborhoods and how to test neighborhood effects.

Background on the Substantive Constructs

Because this study aimed to compare GSM and HLM, it was necessary to select an example application in which the same data could be used with both approaches and there were clear theoretical links between the contextual characteristics to be tested and the outcome variable. It was also useful to choose an outcome known to be influenced by individual-level characteristics, as this allowed a comparison of how the two methods handle issues of context versus composition. The study tested whether crime and NSES exert contextual effects on residents' perceptions of neighborhood problems, after controlling for a variety of individual-level characteristics.

The remainder of this section first describes how the substantive constructs were selected and the theoretical links between the predictors and the outcome. Then it describes the constructs in more detail. After introducing the outcome of interest, this section reviews literature surrounding the contextual predictors, then briefly describes individual-level variables that were incorporated into the analysis because they were expected to be related to the outcome.

Selection of constructs. As noted above, the outcome modeled in this study was perceived neighborhood problems. This outcome was selected because prior HLM research found that it exhibits neighborhood-level variance that can be modeled as

hierarchically structured autocorrelation (Coulton, et al., 2004), while prior research using GIS methods found that it shows a substantial amount of spatial autocorrelation that decays as a function of distance (Bass & Lambert, 2004; Pierce, 2006). Thus, it is a candidate for use in both HLM and GSM. Although no prior literature directly addresses which of these two methods is more appropriate for modeling this outcome, there are some findings and possible theoretical mechanisms that suggest GSM may be more appropriate than HLM.

Meanwhile, crime and NSES were selected because (a) they are frequently used contextual characteristics in neighborhood research, (b) there are clear theoretical links between them and the outcome variable, and (c) the GIS data available for the example dataset permitted both to be measured by aggregating data within any set of neighborhood boundaries. In addition, the spatial distribution of crime was unlikely to be captured well by census-based geographic units (McCord & Ratcliffe, 2007), which were used to construct the neighborhood units in the example dataset⁷ so it was reasonable to expect that using buffers around residents' homes to measure crime might yield different results than using crime aggregated within discrete neighborhood units.

While NSES may be somewhat better aligned with census geography than crime, it too may yield different values for a contextual measure when aggregated within buffers rather than discrete neighborhood units. Using multiple contextual characteristics also allowed the study to explore whether there were differences in the spatial scale on which different contextual characteristics influenced resident outcomes. Overall, these two contextual characteristics possessed essential qualities for pursuing whether the

⁷ The data are a clustered sample originally collected for use in HLM analyses (see the Method section).

differences in how neighborhoods were defined for the purpose of measuring neighborhood conditions in HLM and GSM made a difference in the obtained results.

Theoretical mechanisms. This study assumed that several sources might contribute to the observed spatial variation in perceived neighborhood problems. This section elaborates on the theoretical mechanisms associated with each of those potential sources of neighborhood effects on residents' perceptions.

First, spatial variation in neighborhood crime and NSES could produce contextual effects that explain some of that spatial variation. The theoretical link between actual crime and perceived neighborhood problems derives from broken windows theory (J. Q. Wilson & Kelling, 1982). Nearby crime is an observable sign of social disorder in the neighborhood (Sampson & Raudenbush, 1999) that residents interpret as a social problem (Sampson & Raudenbush, 2004). So, exposure to higher levels of actual crime should lead residents to perceive and report higher levels of neighborhood problems. The mechanism linking NSES to resident's perceptions is different. Sampson and Raudenbush (2004) argue that neighborhoods experiencing concentrated poverty have historically also been afflicted by extensive physical and social disorder, so now poor neighborhoods have become stigmatized as disorderly places. Environmental cues that provide information suggesting that NSES is low (such as low median housing value) may exert a contextual effect on resident's perceptions of neighborhood problems because this stigma primes residents of poorer neighborhoods to perceive more problems than they would in wealthier, but otherwise similar, neighborhoods.

Second, geographical clustering of similar individuals could explain some of the spatial variation in residents' perceptions of neighborhood problems, particularly if

individual-level resident characteristics are good predictors of those perceptions. While this is still a kind of neighborhood effect, it is a compositional effect rather than a contextual effect. Because this study focuses on comparing HLM and GSM for testing the effects of neighborhood-level predictors, this theoretical mechanism is only relevant to the extent that the two methods differ in their ability to control for neighborhood effects resulting from unobserved attraction, selection, and attrition processes that might affect perceived neighborhood problems indirectly through their influence on neighborhood composition.

Finally, residents often exchange information about neighborhood events and conditions with their neighbors (Unger & Wandersman, 1985), so social interactions and social construction of reality (Shinn & Rapkin, 2000) may also shape their perceptions of neighborhood problems. That suggests that spatial autocorrelation remaining in residents' perceptions after accounting for composition and contextual effects could be generated by contagion processes (Leventhal & Brooks-Gunn, 2000) operating through social networks. Such contagion effects can be modeled using distance as a proxy for network connections. Because members of neighborhood networks are more likely to know and interact with others who live nearby (Greenbaum, 1982; Greenbaum & Greenbaum, 1985; Stutz, 1973; Wheeler & Stutz, 1971), spatial autocorrelation should decrease with increasing distance between observations. GSM techniques would model that residual spatial autocorrelation explicitly, while HLM would ignore it.

Perceived neighborhood problems. Perceived neighborhood problems refers to the degree to which a resident thinks undesirable physical conditions and deviant social behaviors are present at unacceptable levels in his or her neighborhood. This construct

appears frequently in the neighborhood research literature, though its name varies.

Comparing definitions across studies reveals that perceived disorder, perceived incivilities, and perceived neighborhood problems refer to essentially the same phenomenon (Bass & Lambert, 2004; Coulton, Korbin, & Su, 1996; Dupéré & Perkins, 2007; Foster-Fishman, et al., 2007; Foster-Fishman, et al., 2009; Franzini, Caughy, Spears, & Esquer, 2005; Franzini, et al., 2008; Perkins, Meeks, & Taylor, 1992; Perkins, Wandersman, Rich, & Taylor, 1993). For example, perceived disorder has been conceptualized as “exposure to deviant behavior in the neighborhood” (Bass & Lambert, 2004, p. 283), “perceptions of deleterious conditions in neighborhoods” (Coulton, et al., 1996, p. 16), and “visible cues indicating a lack of order and social control in the community” (Ross & Mirowsky, 1999, p. 413). This latter definition matches how Perkins and colleagues’ (Perkins, et al., 1992; Perkins, et al., 1993) conceptualize incivilities as symbols of physical and social disorder that signal that an area is poorly supervised.

Regardless of the name, perceived neighborhood problems is typically measured by asking residents to rate the degree to which various conditions or activities are problems in their neighborhood. Such ratings are almost perfectly correlated with the volume of disorder residents report having observed (Sampson & Raudenbush, 2004). The specific items used often include questions about forms of social disorder such as crime, gang activity, prostitution, or drug-dealing, or about signs of physical disorder such as litter, graffiti, abandoned buildings, and poorly maintained homes and yards (Bass & Lambert, 2004; Coulton, et al., 1996; Dupéré & Perkins, 2007; Foster-Fishman,

et al., 2007; Foster-Fishman, et al., 2009; Franzini, et al., 2005; Franzini, et al., 2008; Perkins, et al., 1992; Perkins, et al., 1993).

Residents' perceptions of whether or not things like crime, drugs, prostitution, and abandoned buildings are problems in their neighborhoods are important for several reasons. On one hand, seeing the neighborhood as beset by problems can motivate residents to become active citizens (Chavis & Wandersman, 1990; Greenberg, 2001). For example, Peterson and Reid (2003) found that residents who were aware of substance abuse problems in their neighborhood were more likely to participate in substance abuse prevention activities. Other research has also shown that residents reporting high levels of neighborhood problems were more likely to engage in both individual and collective forms of activism (Foster-Fishman et al., 2007). Although perceived levels of problems may not be important in predicting citizen participation among self-identified neighborhood leaders, they are related to participation among residents who do not see themselves as leaders (Foster-Fishman, et al., 2009).

On the other hand, if residents perceive that neighborhood crime problems have grown too severe, they may fear retaliation and refrain from intervening when local youth are misbehaving (Korbin & Coulton, 1997), thereby weakening informal social control processes. Indeed, residents who perceive severe problems in their neighborhood may simply exit the neighborhood altogether (Orbell & Uno, 1972). Perceived neighborhood problems may also influence residents in other ways. For example, residents of neighborhoods characterized by high average levels of perceived problems tend to report being in poorer health than residents of neighborhoods with lower levels of problems (Pampalon, Hamel, De Koninck, & Disant, 2007) perhaps because perceived problems

are sources of chronic stress that increase the risk of poor health and impair physical functioning (Stephoe & Feldman, 2001).

Previous HLM research has found evidence of neighborhood-level variability in residents' perceptions of neighborhood problems at both the block group and census tract levels (Coulton, et al., 2004; Franzini, et al., 2008; Quillian & Pager, 2001; Sampson & Raudenbush, 2004). Of these, only one study examined the data at multiple spatial scales. Coulton et al.'s (2004) research showed that perceived neighborhood disorder and incivilities varies on a relatively small spatial scale, with larger ICCs observed in smaller neighborhood units. Though it originates from an HLM study, this is potentially more consistent with the kind of distance-based spatial autocorrelation assumed in GSM than with the strictly hierarchical autocorrelation assumed in HLM.

More direct empirical support for spatial autocorrelation in perceived disorder comes from Bass and Lambert (2004), who collected perception data from adolescents in Baltimore via face-to-face interviews, then used variograms to model the spatial autocorrelation in those data. Although they do not specifically report range parameters for their variograms, visual inspection of their plots suggests that the range of autocorrelation may be between 200-400 m in the raw data, and perhaps as high as 1000 m after accounting for census-tract level crime and poverty measures. Preliminary research by the present author found that the range of spatial autocorrelation in residents' levels of perceived neighborhood problems was approximately 600 m (Pierce, 2006). Together with the HLM studies mentioned above, these studies suggest that there is indeed spatial variation in this outcome, but none of them directly test whether hierarchical or spatial structure better describes that spatial variation.

Crime. Crime represents the extreme end of the continuum of social disorder (Sampson & Raudenbush, 1999). Victims of crime often experience serious adverse consequences such as injury, death, financial losses, property damage, psychological distress, and mental health problems. The salience of crime as a social problem is underscored by the tremendous amounts of time, money, and other resources devoted to defining crime and the legal consequences of committing it, catching and prosecuting persons accused of crimes, and sequestering and rehabilitating convicted offenders. It should come as no surprise then that crime is often viewed as an important contextual characteristic of neighborhoods that poses a serious problem for residents. This assumption about crime is apparent in measures of residents' perceptions of neighborhood problems. Questions about crime in general or about specific criminal activities such as burglary, drug-dealing, or prostitution frequently appear in measures of those perceptions (Foster-Fishman, et al., 2007; Foster-Fishman, et al., 2009; Franzini, et al., 2005; Franzini, et al., 2008; Meersman, 2005; Perkins, et al., 1992; Perkins, et al., 1993; Quillian & Pager, 2001; Sampson & Raudenbush, 2004).

Because levels of crime represent real variations in the local environment, it is quite natural to expect that residents' perceptions of neighborhood problems will be sensitive to this reality rather than divorced from it (Quillian & Pager, 2001). Crime is salient and threatening, so residents are alert for signs of its presence. If observed in sufficient quantity or severity, then residents perceive crime as a problem in the neighborhood (Sampson & Raudenbush, 2004). Residents may directly witness crimes, or they may indirectly perceive crime via physical cues such as bullet holes or smashed storefront windows left at the scene of a crime (Sampson & Raudenbush, 1999).

Information about local crime may also be obtained indirectly from discussions with **neighbors** or through local news media (Perkins & Taylor, 1996).

Official police records are the primary source of data for measuring crime in **neighborhood** research (Chaix, et al., 2006; Franzini, et al., 2008; Quillian & Pager, 2001; Sampson & Raudenbush, 2004). Many crimes, particularly less serious ones, are never **reported** to the police and sometimes police do not record minor crimes that are reported **to them**, so police data almost certainly underestimate actual crime (Quillian & Pager, 2001). Despite that limitation, they may be the best available source for many studies.

However, other measurement issues must still be addressed. Perhaps the most **important** issue is which crimes should be counted. So far, all three HLM-based studies that **have** used crime as a predictor of neighborhood problems have operationalized crime in **terms** of rates of violent crime (all studies include assault, homicide, rape, and robbery, but **one** study also included burglary, theft, and arson) that were log-transformed to **reduce** skew (Franzini et al., 2008; Quillian & Pager, 2001; Sampson & Raudenbush, 2004).

However, a community psychologist might have a theoretical interest in testing **whether** two or more different kinds of crime independently affect resident perceptions. Such **interests** might include testing hypotheses about whether different kinds of crime **operate** on different spatial scales. For instance, one might want to test whether residents' **perceptions** are only sensitive to property crime occurring quite close to their homes, but are **sensitive** to violent crimes occurring over a larger geographic area. Three different **crime** variables based on the major categories (crimes against persons, crimes against **property**, and crimes against society) used by the Federal Bureau of Investigation to

classify crimes (Uniform Crime Reporting Program, 2000) were considered for use in **this** study, but multicollinearity problems ultimately prevented using more than one crime **variable** at a time (see Method section).

That forced a methodological choice about which type of crime to use in the **analyses** reported below. Crime against persons (i.e., violent crime) was selected for three **reasons**. First, the presence of violent crime in the neighborhood is an unambiguous **threat** to residents' safety and is undoubtedly a sign of very serious social disorder. It **should** therefore be more salient in shaping residents' perceptions than property crime or **crimes** against society. Second, this is consistent with how crime has been measured in **previous** research as noted above. Third, crime against persons exhibited the strongest **relationships** with the outcome in preliminary analyses.

Another measurement issue is whether crime should be measured by the raw **number** of crimes occurring in a neighborhood, by a crime rate (number of crimes per **capita**), or by crime density (number of crimes per unit area). Raw crime counts are **generally** not used in neighborhood research because they do not adjust for the variation in **either** the size or population of the neighborhoods. Neighborhood crime is frequently **operationalized** with crime rates (Franzini et al., 2008; Quillian & Pager, 2001; Sampson & **Raudenbush**, 2004), though crime density has been used occasionally (Brodsky, O'**Campo**, & Aronson, 1999). Crime rates reflect the fact that a given number of crimes may **be** felt more acutely in sparsely populated areas and are mostly intended to measure **victimization** risk (Bowes & Ihlanfeldt, 2001), but it can also be argued that people may **be more** affected by the absolute number or spatial density of crimes regardless of **population** density (Chaix, et al., 2006). Given this study's focus on spatial analysis, it

made sense to focus on crime density rather than crime rates because (a) density measures are preferred by researchers who study the spatial distribution of crime (Chainey, Tompson, & Uhlig, 2008), (b) crime density may be more closely associated with the average resident's knowledge of nearby crimes (Bowes & Ihlanfeldt, 2001), and (c) crime density is a better measure of exposure to neighborhood crime for people who wish to avoid either witnessing a crime or being victimized personally (Bowes & Ihlanfeldt, 2001).

No GSM studies have yet linked crime to perceived problems, but one did find that higher numbers of violent crimes in a 500 m radius around residents' homes increased the risk of substance abuse disorders (Chaix, et al., 2006). Three HLM-based studies have demonstrated that there is indeed a link between actual crime and perceived neighborhood problems (Franzini et al., 2008; Quillian & Pager, 2001; Sampson & Raudenbush, 2004). Using census block-groups to represent neighborhoods, both Franzini et al. and Sampson and Raudenbush found that high crime rates were associated with higher levels of perceived problems among residents. Quillian and Pager found similar results using census tracts as neighborhoods. Unfortunately, none of these HLM-based studies tried varying the size of the neighborhood units within which crime was measured, so little is known about sensitivity of the contextual effect to changes in the spatial scale on which crime is measured.

Relying on crime data aggregated to neighborhood units such as block groups ignores the fact that crime is very unevenly distributed over space. The literature on detecting crime "hot spots", which relies on using GIS tools to map and analyze spatial point patterns in the locations of crimes, shows that crimes cluster together in small

geographic areas (Block, 2000; Ratcliffe & McCullagh, 1999; Taylor, 1998) and hot spots may span the borders between adjacent tracts or block groups (McCord & Ratcliffe, 2007). Thus, exposure to crime depends on location because crime is neither uniformly nor randomly distributed over space (Block, 2000). This suggests that there may be substantial spatial variability in crime within individual census tracts or block groups.

In crime mapping studies, measuring crime within administratively defined geographic neighborhood units is considered to be particularly vulnerable to the MAUP, so instead researchers employ buffer techniques to summarize the spatial point pattern of crimes by calculating estimates of crime intensity at a high-resolution grid of points across the study region (Ratcliffe & McCullagh, 1999). The intensity of a crime point pattern is the number of crimes per unit area within the buffer centered on each grid point (often after weighting individual crimes so that those far from the center of the window count less than those close to the center), so it measures crime density relative to spatial area (Ratcliffe & McCullagh, 1999), whereas per capita crime rates measure crime density relative to population size. In these buffer approaches, the grid points are often close enough together that windows centered on adjacent grid points overlap substantially. That is useful because it allows the construction of relatively smooth maps of the crime intensity surface within the study region.

With geocoded crime incident data, researchers can aggregate crime either within the fixed, mutually exclusive neighborhood units used in HLM, or in buffers centered on residents' homes. With GSM, one can use either of those methods for measuring crime and directly compare how changing the neighborhood definition used to measure crime influences the strength of the association between crime and perceived neighborhood

problems. HLM-based analyses may underestimate the strength of that relationship because the arbitrary neighborhood boundaries do not correspond well with what residents think of as their neighborhoods and therefore the crime measure may not include the crimes that are most salient to the resident. Perhaps using buffers to measure crime, paired with GSM's ability to vary the scale on which that contextual factor is measured, will produce better models than HLM-based analyses by more accurately capturing the crime occurring in the areas residents think of as their neighborhoods.

As a supplemental analysis in an HLM study that used census tracts to define neighborhood units (without describing the physical size of those tracts), Quillian and Pager (2001) investigated whether actual crime rates in adjacent tracts influenced perceptions of crime, after controlling for crime levels in the individuals' own tract. They found little evidence that crime in adjacent tracts mattered, which suggests that the geographic scale on which crime may matter is equal to or smaller than the size of census tracts. However, their results are subject to all the limitations associated with adopting fixed boundary systems for defining neighborhood units, so this is relatively weak evidence about the spatial scale on which crime may impact resident perceptions.

Neighborhood SES. Many researchers have pursued questions about how the socioeconomic context in residential neighborhoods affects residents, often focusing on outcomes among children and youth (Leventhal & Brooks-Gunn, 2000; Sampson, et al., 2002). Wilson's (1987) observation that poverty was increasingly concentrated in inner-city neighborhoods over the 1970s and 1980s spurred renewed interest in poverty as a contextual phenomenon rather than simply an individual-level problem, resulting in a wave of studies focusing on the consequences of living in high-poverty neighborhoods

(Gephart, 1997; Leventhal & Brooks-Gunn, 2000; Sampson & Morenoff, 1997). There is now a substantial body of research showing that neighborhoods vary substantially with respect to various indicators of NSES and that these contextual variations are linked to many outcomes for children and youth, including school achievement, cognitive problem solving skills, behavior problems, delinquency, sexual activity, and teen pregnancy (Caughy & O'Campo, 2006; Gephart, 1997; Leventhal & Brooks-Gunn, 2000; Pebley & Sastry, 2003; Ramirez-Valles, Zimmerman, & Juarez, 2002; Sampson et al., 2002).

Census data are often used to obtain neighborhood poverty rates (Brooks-Gunn, Duncan, Leventhal, & Aber, 1997), which represent the proportions of residents living in households with annual incomes below the poverty threshold defined by the federal government. Although poverty rate is a frequently used measure of NSES, other measures have also been used. For example, some studies looked at concentrated affluence rate (percent of residents living in households with annual incomes exceeding a researcher-defined threshold such as \$75,000) in addition to poverty rate (Beyers et al., 2003; Pebley & Sastry, 2003; Sampson, 2001), while others have operationalized NSES with mean income (Rountree & Land, 1996) or median housing values, which measure the value of residential property (Cozier et al., 2007; Gee, 2002; Laraia et al., 2006). Although hardly surprising, it is important to note that different measures of NSES are often strongly correlated. For example, poverty rates are inversely correlated ($r = -.62$) with median housing value at the level of census tracts (Gee, 2002). One explanation for that lies in the fact that local land use regulations (zoning ordinances) promote spatial segmentation of cities into neighborhoods with similar residential property values,

leading to corresponding spatial segmentation of the population by income because families tend to seek better housing as their incomes rise (Schill & Wachter, 1995).

Measuring NSES via median housing value is attractive because property value data are often available from local tax assessors' offices and are usually more current than census data (Coulton & Hollister, 1998; Kingsley, Coulton, Barndt, Sawicki, & Tatian, 1997). In addition, such data are often available as GIS files (Kingsley, Coulton, Barndt, Sawicki, & Tatian, 1997), making it feasible to estimate the median residential property value within any desired geographic area (e.g., fixed neighborhood units, buffers, or both), regardless of whether it will be used in HLM or GSM analyses.

There is extensive evidence that poverty, crime, physical and social disorder, and other social problems tend to co-occur in the same geographic places (Sampson, 2001; Sampson, et al., 2002). For example, recent research has shown that owner occupied median housing value is negatively correlated ($r = -.56$) with observed physical incivilities when both measures are aggregated to the block group level (Laraia, et al., 2006). Given that objective levels of neighborhood disorder are generally good predictors of perceived disorder (Franzini et al., 2008; Perkins et al., 1992; Sampson & Raudenbush, 2004), it may be that residents of low-SES neighborhoods perceive more problems simply because there are indeed more present. This link between NSES and perceived problems might also be mediated by other variables: in a multilevel study of contextual effects on youth alcohol and drug problems, structural equation modeling demonstrated that high levels of neighborhood poverty led to decreased social cohesion, which was in turn associated with greater perceived problems with youth alcohol and drug use (S. C. Duncan, Duncan, & Strycker, 2002).

However, there may be another way in which NSES influences residents' perceptions. Sampson and Raudenbush (2004) used HLM models to show that that neighborhood racial composition and neighborhood poverty both predicted perceived disorder even after controlling for observed levels of physical and social disorder. They interpret those findings as support for their contention that perceived disorder is in part socially constructed, arguing that "Neighborhoods with high concentrations of minority and poor residents are stigmatized by historically and structurally induced problems of crime and disorder" (Sampson & Raudenbush, 2004, p. 337). Essentially, the stigma associated with poverty primes residents of poor neighborhoods to perceive more disorder than can be explained by observed disorder alone. Additional empirical support for this link between neighborhood poverty and perceived disorder comes from work by Franzini et. al. (2008) who used methods similar to those of Sampson and Raudenbush, but sampled from a different city.

Two of the studies linking NSES to perceived problems discussed above operationalized neighborhoods with census block groups (Franzini, et al., 2008; Sampson & Raudenbush, 2004), while another did not describe what geographic units were used to operationalize neighborhoods, though it does say that census data were used to measure poverty (S. C. Duncan, et al., 2002). None of those studies specifically explored the spatial scale on which NSES is most closely linked to perceived problems. In addition, none of them applied GIS-based spatial analysis approaches: they all relied on HLM (Franzini et al., 2008; Sampson & Raudenbush, 2004) or related methods like multilevel structural equation models (S. C. Duncan, et al., 2002).

Only one study has used GIS methods to explore the link between NSES and perceived neighborhood problems. Meersman (2005) measured poverty and other indicators of NSES (percent college educated, percent residential stability, percent unemployment) within a set of buffers of varying sizes ($0.25 \text{ mile} \leq \text{radii} \leq 1.50 \text{ miles}$, in 0.25 mile increments) centered on residents' homes. At each window size, GIS tools were used to determine which census tracts were overlapped by the window around a resident's home, then the poverty rate in the window area was set to the weighted average of the poverty rates from those census tracts. The tract weights were the proportions of the buffer's area that belonged to each census tract. Meersman argued that this method allows one to take into account a resident's precise location within a census tract, plus proximity to other census tracts, but this method still suffers from the MAUP.

Measuring poverty in a series of concentric, circular buffers allows one to compare the effects of neighborhood poverty measured on different spatial scales (Meersman, 2005). This technique naturally allows the buffers for different residents to overlap to different degrees depending on how far apart they live. Using OLS regression, Meersman found that poverty had the largest standardized coefficient as a predictor of perceived neighborhood problems when measured over a 1.50 mile radius. Other NSES indicators had their strongest effects at other buffer sizes (residential stability at 0.25 mile, unemployment at 0.75 mile). There are several problems with Meersman's study. First, using OLS regression to analyze the data ignores the likely presence of spatial autocorrelation. Second, weighted versions of census tract poverty rates are crude measures of NSES in a buffer around individual homes because the aggregation that had already occurred to create tract level measures eliminates any spatial variability within

tracts. For a superior buffer measure of NSES, it would be far better to start from point-referenced data or data that represent geographic areas small enough to be treated as point-referenced data (e.g., parcel-level property value data). Third, Meersman always measured different indicators of NSES at the same geographical scale: while he did note the scale on which each measure had the strongest effects, he did not go beyond that to combine measures associated with different size buffers in the same model.

Individual-level predictors. People who live close together may not necessarily experience or perceive the neighborhood in the same way, so contextual conditions are not the sole influence on people's perceptions of neighborhood problems. Several studies show that individual-level factors (e.g., sex, age, race, etc.) also predict those perceptions (Franzini, et al., 2008; Meersman, 2005; Quillian & Pager, 2001; Sampson & Raudenbush, 2004), indicating that residents' perceptions are not pure reflections of external conditions (Quillian & Pager, 2001).

Residents who are more physically or socially vulnerable to crime tend to report higher levels of fear of crime (Rountree & Land, 1996), suggesting that some residents, such as the elderly or women, may have lower thresholds for deciding that the conditions they observe constitute a problem. While the empirical data show that women do consistently report higher levels of perceived crime and disorder (Quillian & Pager, 2001; Sampson & Raudenbush, 2004), the evidence for age effects is somewhat mixed. Older residents report higher levels of perceived crime in one study (Quillian & Pager, 2001), but lower levels of perceived disorder in others (Meersman, 2005; Sampson & Raudenbush, 2004).

Research appear to consistently find that Black residents report lower levels of perceived disorder than White residents (Franzini, et al., 2008; Meersman, 2005; Sampson & Raudenbush, 2004). In addition, the extent to which a resident's race predicts perceived disorder varies across neighborhoods (Sampson & Raudenbush, 2004) indicating that it may be fruitful to explore cross-level interactions between race and neighborhood-level factors.

Other personal characteristics also might affect residents' perceptions of neighborhood problems. Marital status effects on perceived neighborhood problems have been examined in a couple studies, but the results are inconsistent. Compared to widowed residents, Franzini et al. (2008) found that married and separated or divorced residents perceive less disorder than widowed residents, but Sampson and Raudenbush (2004) found that separated or divorced residents perceive more disorder than widowed residents. Similarly, higher levels of education are sometimes associated with less perceived disorder (Franzini, et al., 2008), but other research did not find an education effect (Quillian & Pager, 2001). Another characteristic that might be important is the presence of children in the home. Although none of the available studies address this factor, residents who are raising children may be particularly concerned about the quality of the neighborhood environment and therefore more likely to view a given situation to be a problem than people who are not raising children.

Identifying individual-level factors that may be related to the outcome under study is important in this study primarily because it controls for neighborhood composition, thereby permitting a better test of the importance of contextual factors (C.

Duncan, et al., 1998; Merlo, Yang, et al., 2005). Thus, in the present study, several personal characteristics were incorporated into both the HLM and GSM analyses.

Linking Gaps in the Literature to Hypotheses for the Present Study

This section describes how the present study fills specific gaps in the literature on comparing HLM and GSM. Along the way, it presents arguments for why GSM may be a better alternative than HLM, and links the research questions for the study to specific hypotheses that can be tested to inform our thinking about neighborhoods and how to test neighborhood effects.

Taking full advantage of spatial information. The first gap in the literature is that previous comparisons of HLM and GSM have not taken full advantage of GSM's tools for representing spatial autocorrelation. None of them have used the precise locations of the residents in the sample both for constructing buffer-based measures of contextual factors and for estimating the variogram in the GSM. This is due to either a lack of precise location data (Boyd, et al., 2005; Chaix, Merlo, & Chauvin, 2005), or to computational difficulties associated with the size of the dataset and software limitations (Chaix, Merlo, Subramanian, et al., 2005). However, new software makes it possible to run GSM analyses with large datasets while taking full advantage of precise location data (Finley et al., 2007). Location data for residents was available, so this study used that software to better model the actual pattern of spatial autocorrelation in the data than has been possible in previous studies.

Detecting autocorrelation. A second gap in the literature is that previous studies have not directly compared the amounts of neighborhood-level variance and autocorrelation detected by HLM and GSM. This prompted the first research question for

this study, which simply asked: how do GSM estimates of neighborhood-level variance and autocorrelation compare to HLM estimates? Both methods can estimate neighborhood- and individual-level variance components that can be converted into directly comparable measures of autocorrelation (ICC for HLM, PSR for GSM).

Recall that GSM ignores the boundaries of the discrete neighborhood units used in HLM and can therefore potentially account for autocorrelation both within and between them, while HLM will only account for within-neighborhood autocorrelation. That suggests that GSM may be the more sensitive method for detecting neighborhood-level variability in outcomes, particularly if some neighborhood units are close enough together that spatial autocorrelation may spill over between them. Furthermore, Coulton et al. (2004) found successively larger ICCs for perceived neighborhood disorder and incivilities when they examined smaller and smaller neighborhood units, indicating that neighborhood-level variances were getting larger as the neighborhood units got smaller. Because the correlation functions built into GSM models assume that neighborhood-level variances decay with increasing distance and they start at distances far smaller than any neighborhood unit adopted for HLM analyses, Hypothesis 1 (H1) is:

H1: GSM estimates of neighborhood-level variance and the amount of autocorrelation for perceived neighborhood problems will be higher than the corresponding HLM estimates, both before and after controlling for neighborhood composition.

The correlation function in a GSM model also provides a way to compare the spatial scale of autocorrelation in perceived neighborhood problems to the size of the neighborhood units used in the HLM models. The neighborhood units in this study are

substantially smaller than block groups or census tracts (Van Egeren, Huber, Foster-Fishman, Pierce, & Law, 2007), which are the units typically used in HLM studies. Some of them are also quite close together. This is one of the situations in which spatial autocorrelation with a long enough range might spill over between neighborhood units. Given the evidence that spatial autocorrelation in perceived neighborhood problems may extend several hundred meters or more (Bass & Lambert, 2004; Pierce, 2006),

Hypothesis 2 (H2) is:

H2: The range of spatial autocorrelation in perceived neighborhood problems detected by GSM will be long enough to reach across the borders between at least some of the neighborhood units used in the HLM analyses.

While testing H1 and H2 provides useful information about the potential importance of neighborhoods and spatial scale of autocorrelation, it does not directly tell us whether the autocorrelation in the data is hierarchically or spatially structured. The third gap in the literature is that previous comparisons of how HLM and GSM handle autocorrelation have been incomplete and one-sided. Previous work is incomplete because while it has examined the neighborhood-level residuals in HLM analyses for evidence of residual spatial autocorrelation (Boyd, et al., 2005; Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian, et al., 2005), it has not looked for evidence of spatial autocorrelation in the level 1 HLM residuals. If they contain spatial autocorrelation, this will be another indication that HLM is not fully accounting for the spatial variability in the data. Previous studies have been one-sided because none have yet reported any attempt to examine the residuals from GSM analyses for evidence of hierarchically structured autocorrelation.

Modeling autocorrelation. That prompted the second research question for the study, which asked: which method (HLM or GSM) is more effective at modeling the autocorrelation actually observed in data from neighborhood residents? This is ultimately a question about which conceptualization of neighborhoods as places within geographic space provides a heuristic for grouping residents that is more consistent with the empirical data. Put another way, it gets at whether only place matters, or both place and spatial proximity matter: while HLM assumes that neighborhoods are independent and thus only a resident's own neighborhood matters, GSM assumes that neighborhoods are embedded in a larger spatial fabric and that residents are influenced by multiple neighborhoods, with the amount of influence each exerts depending on spatial proximity.

Several findings from the literature suggest that GSM will be superior to HLM for modeling spatial variability in outcomes because its assumptions are more consistent with what we know about neighborhoods. Neighborhood research has shown that daily life frequently takes people across the borders of traditional neighborhood units like block groups (Sastry, et al., 2002) and fixed neighborhood boundaries are quite artificial (Coulton, et al., 2001; Montello, et al., 2003). Residents tend to think of their own home as the center of their neighborhoods (Coulton, et al., 2001; Lee & Campbell, 1997), are more likely to be acquainted with other residents who live close to them than with people who live farther away (Greenbaum & Greenbaum, 1985), and tend to visit nearby census tracts more often than distant ones (Wheeler & Stutz, 1971) indicating that urban social travel exhibits proximity effects. In addition, similarity in outcomes as a function of spatial proximity is a common feature in spatial data (Bailey & Gatrell, 1995; Haining, 2003; Tobler, 1970).

Two lines of empirical evidence suggest that spatial rather than hierarchical structure may better describe the autocorrelation in perceived neighborhood problems. **First**, one of the few HLM studies to have systematically varied the size of the neighborhood units used observed larger ICCs as smaller and smaller neighborhood units were tested with this outcome (Coulton, et al., 2004); this is precisely what one would expect to see if there really was spatial rather than hierarchical structure in the actual data, but one tried to model the data with HLM rather than GSM. Second, both Bass and Lambert (2004) and Pierce (2006) found direct evidence for distance-based spatial autocorrelation by using variogram models with survey-based measures of perceived neighborhood problems. Therefore, Hypothesis 3 (H3) is:

H3: An empty GSM will fit the perceived neighborhood problems data better than an empty HLM. Similarly, a GSM model of perceived neighborhood problems containing only individual-level predictors will fit better than a corresponding HLM model containing only individual-level predictors of perceived neighborhood problems.

Support for H3 would indicate that the conceptualization of neighborhoods associated with GSM provides a better basis for grouping residents than the one associated with HLM. In statistical terms, it would suggest that autocorrelation is spatially structured, not hierarchically structured.

Testing assumptions by examining residuals. Both HLM and GSM assume that the individual-level residuals they produce are fully independent of one another and of the neighborhood-level residuals. On the other hand, the neighborhood-level residuals are assumed to be independent of each other only in HLM because GSM explicitly assumes

there will be different degrees of similarity among neighborhood-level residuals depending on the distance between the locations associated with them. To the extent that a statistical model makes accurate assumptions about the autocorrelation structure in the data, the resulting residuals will meet the model assumptions. If either statistical method is not effectively modeling the autocorrelation in the actual data, that should be evident in its residuals. Therefore, inspecting the residuals from each method may reveal clues about which method is performing better. For example, if the data contain spatial autocorrelation, but they are modeled with HLM, then there will still be residual spatial autocorrelation in the level 1 and/or level 2 HLM residuals. Given the argument that led to H3, it follows that Hypotheses 4 and 5 (H4 and H5, respectively) are:

H4: HLM will not fully control for spatial autocorrelation in perceived neighborhood problems, so there will be evidence of residual spatial autocorrelation remaining in both the Level 1 and Level 2 residuals from HLM models.

H5: GSM will fully control for within-neighborhood spatial autocorrelation in residents' perceptions of neighborhood problems, so there will be no evidence of hierarchical autocorrelation remaining in the individual-level residuals from GSM models.

HLM assumes that the amount of autocorrelation between residents of the same neighborhood is the same no matter where they are located within the neighborhood. If the underlying pattern of autocorrelation is a function of distance, then HLM would still detect what appeared to be hierarchical autocorrelation in neighborhood-level residuals from GSM. That would happen because HLM would be grouping people who are close

together (high spatial autocorrelation) with those who are farther apart (low spatial autocorrelation), which should essentially average out to an amount of hierarchical autocorrelation that lies below the maximum level of spatial autocorrelation (found at short distances), but higher than the minimum level of spatial autocorrelation. Thus, Hypothesis 6 is:

H6: Neighborhood-level GSM residuals from a model predicting perceived neighborhood problems will contain hierarchical autocorrelation when examined with HLM, but the ICC will be lower than the PSR.

Testing contextual effects. A fourth gap in the literature concerns how neighborhood-level factors have been measured when comparing HLM and GSM approaches. The source data for measures of constructs like NSES have usually been derived from census data that were only available at the level of areal units (Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian, et al., 2005). Thus, previous studies have used crude methods to convert areal data originally associated with one set of geographic units to estimates of the values that might be obtained within the boundaries of the buffers used in GSM. A more refined approach would involve starting from a spatial dataset of much higher resolution (preferably point-referenced data for households) that can be aggregated directly to match the boundaries of the units used for HLM analyses or within the buffers used for GSM analyses with equal ease. This study was the first to use such data sources for measuring the contextual factors.

To address whether the buffers we can use in GSM provide a better geographic definition of neighborhoods for testing the effects of specific contextual conditions on residents than the neighborhood units used in HLM, this study varied how neighborhood

boundaries are defined in GSM (fixed neighborhoods vs. buffers centered on residents' homes) and compared the results to corresponding HLM models. A GSM analysis that uses fixed neighborhoods like those used in HLM differs from the HLM only in the assumptions made about how to model autocorrelation, while a GSM analysis that uses buffers to approximate neighborhoods also differs from the HLM in how neighborhood boundaries were set. Comparing both kinds of GSM analyses to an HLM analysis allowed the study to disentangle whether any improvement of GSM over HLM was due to how autocorrelation was modeled, how neighborhood boundaries were defined for measuring neighborhood conditions, or the combination of these aspects of the method.

Given the research suggesting that residents tend to see their own homes as the center of their neighborhoods (Coulton, et al., 2001; Lee & Campbell, 1997), using buffers to represent neighborhood boundaries should produce better GSM models than using fixed neighborhoods because buffers better approximate how residents think about their neighborhoods. Thus, Hypothesis 7 (H7) is:

H7: GSM will yield models that fit better and have larger contextual effects of crime and NSES on perceived neighborhood problems than corresponding HLM models when they use contextual measures calculated within appropriately-sized buffers. Using HLM-style contextual measures of crime and NSES calculated within discrete neighborhood cluster boundaries in GSM analyses will yield models of perceived neighborhood problems that improve on HLM results, but not as much as when buffers are used.

Support for H7 would indicate that the buffers used in the GSM are a better approximation of the geographic areas that are actually relevant to resident's outcomes

than the neighborhood units used in the HLM analyses. Testing H7 depends on using buffers that are the right size to best capture the spatial scales on which particular neighborhood characteristics matter, which may be smaller or larger than the fixed neighborhood. This study investigated the effects of two neighborhood characteristics, namely crime and NSES, on residents' perceptions of neighborhood problems.⁸

Comparing GSM models that varied only in the size of the buffers used for measuring these contextual characteristics allowed the study to select an appropriate buffer size for each of them (Chaix, et al., 2006; Chaix, Merlo, Subramanian, et al., 2005).

Examining spatial scale. There is little research or theory available that directly addresses how varying the spatial scale on which crime and NSES are measured might affect their relationship with residents' perceptions. Meersman (2005) examined the effect of the percentage of the population living in poverty within buffers with radii ranging from 0.40 km (0.25 mile) to 2.41 km (1.50 mile) on perceived neighborhood problems. He found that measuring poverty in the 2.41 km radius buffer provided the strongest effect on perceived neighborhood problems. However, the study by Kruger (2008) offers indirect evidence the spatial scale on which NSES matters may be much smaller: He found that physical decay of residential buildings correlated most strongly with residents' fear of crime when measured in 0.40 km (0.25 mile) buffers. Assuming that residential decay is strongly correlated with NSES and that fear of crime is strongly correlated with residents' perceptions of neighborhood problems, one might expect that NSES may be best measured in buffers of similar size in this study. Clearly, this is a large difference in possible spatial scales, so the study examined a range of spatial scales.

⁸ The rationale for using these variables in the present study is explained in the next section.

These studies provided rough bounds on the range of spatial scales at which NSES might be expected to matter to residents. Unfortunately, there is less literature available to guide expectations about the spatial scale on which crime might matter.

It is possible that different social processes may underlie the importance of these two contextual factors in shaping residents' perceptions of their neighborhoods and that therefore those processes may operate on different geographical scales. There is no a priori reason to believe that the buffers that produce the strongest relationships between crime or NSES characteristics and outcomes will be the same size as the neighborhood units adopted for conducting HLM analyses. Accordingly, Hypothesis 8 (H8) is a simple, exploratory hypothesis:

H8: The geographical scales on which crime and NSES influence resident perceptions of neighborhood problems will differ from one another and from the average size of the neighborhood areas used in the HLM analysis.

In summary, both HLM and GSM provide ways to examine neighborhood effects. The two methods make different assumptions and differ with respect to how compatible they are with certain conceptualizations of neighborhoods. So far, HLM has been used extensively in community psychology, but GSM has rarely been applied. Very little has been done to explicitly compare the two methods. This study tested eight hypotheses by applying HLM and GSM to the same dataset and comparing the results. While there are good reasons to expect that those hypotheses will be supported, they may not be. Therefore, the next section considers some possible reasons why the data might fail to support one or more of those hypotheses.

Alternative possibilities. Naturally, it is worthwhile to consider reasons why the hypotheses above might not be supported in the present study. The most obvious possibility is that the neighborhood units used in the HLM analyses might very well represent distinct, meaningful neighborhoods. Indeed, guided by recommendations and common practices in the multilevel modeling literature on neighborhood effects (Roosa, et al., 2003), the research team who collected the survey data used in this study invested considerable effort in trying to construct ecologically meaningful neighborhood boundaries that would maximize neighborhood-level variance under an HLM framework (Van Egeren, et al., 2007). If that effort was successful, then those units may be highly salient to residents and the spatial variation in residents' perceptions may be more consistent with the assumptions of HLM than of GSM.

So, how could that come about? Unlike other recent studies (Coulton, Chan, & Mikelbank, 2010; Coulton, et al., 2001), this one did not collect a map of each resident's self-reported neighborhood boundaries. The rationale presented above for expecting that the residents might not agree on neighborhood boundaries and that therefore conceptualizing neighborhoods as partially overlapping geographic areas (approximated here by the buffer-based GSM models) was based on this prior work. It is possible that this might not be true in this sample and that the neighborhood unit boundaries selected in the original sampling design do ultimately capture some consensus definition of these residents' local neighborhoods. If so, one might then expect higher neighborhood-level variances under HLM than under GSM, which would fail to support H1. For H2, it is also possible that that practical range of spatial autocorrelation in residents' perceptions of

neighborhood problems may vary considerably from city to city and that, in this sample, it might be too short to reach across neighborhood boundaries.

Before discussing additional hypotheses about the HLM and GSM residuals that can be tested to more fully answer the second research question, we should consider scenarios that conflict with the prediction in H3. As with H1, this hypothesis might not be supported if the neighborhood units used for the HLM really are as meaningful as they were originally intended to be. Alternatively, another reason that H3 might not be supported derives from the fact that both HLM and GSM adopt fairly simple assumptions about how autocorrelation might be structured. It is possible that (a) neither of those assumptions is accurate and the spatial variation in residents' perceptions does not fit either model well, or (b) that the underlying structure in the data is actually a mixture of hierarchical and spatial structures, such that both forms of autocorrelation are present. In either of those scenarios, H3 might not be supported. Because H4-H6 are corollaries of H3, they are unlikely to be supported if H3 is unsupported. So, if a relatively pure distance-decay pattern of spatial autocorrelation does not adequately describe the spatial variability in the data, these three hypotheses may not be supported.

As with previous hypotheses, if the neighborhood unit boundaries selected for the HLM analyses are in fact as meaningful for residents as they were intended to be, then H7 and H8 will probably not be supported. This is especially true if residents' perceptions are not affected by crime or NSES in areas outside of their own neighborhood. One way that could happen is if they are more acutely aware of conditions in their own neighborhoods than they are of conditions in other surrounding neighborhoods. To borrow a term from behavioral geography, residents' "awareness

space” (McCord, Ratcliffe, Garcia, & Taylor, 2007) might not extend much beyond the borders of the neighborhood units used in the sampling design. Even if they are aware of conditions in surrounding neighborhoods, residents might think of those places as sufficiently distinct from their own neighborhood that they ignore the crime or signs of poverty in surrounding neighborhoods when assessing the level of problems in their own neighborhoods.

Several of the hypotheses above depend on the assumption that circular buffers are a good way to represent neighborhoods. However, it is certainly possible that this is not the case and that buffer-based methods should rely instead on some other, more valid method for defining buffer boundaries. Similarly, the current hypotheses assume that, for any given neighborhood-level predictor, the same size buffer is appropriate for measuring the neighborhood area relevant to all residents. This may not be the case because there is some literature suggesting that the size of residents’ self-reported neighborhoods may be related to individual-level characteristics such as age or gender (Lee, 2001). Exploring that possibility was outside the scope of the current study, but it is certainly worth pursuing in future studies.

Summary of the Study

The purpose of the study was to test whether GSM could serve as a useful alternative to HLM in neighborhood research and to respond to the recent call to start applying spatial analysis methods in community psychology (Luke, 2005; Mowbray et al., 2007). To fulfill that purpose, this study used both HLM and GSM to test hypotheses about the effects of two neighborhood-level variables (crime and NSES) on residents’ perceptions of neighborhood problems. By comparing parameter estimates and model fit

indices from HLM and GSM analyses of the same data, the study explored whether the fundamental differences in how neighborhoods are conceptualized and defined in these two methods led to differences in their statistical performance.

How we conceptualize neighborhoods and geographic space informs two aspects of neighborhood studies, (a) how we group residents in order to detect spatial variability and model autocorrelation in outcomes, and (b) how we define the geographic area of the neighborhood that should be used when measuring neighborhood context. This study seeks to answer four research questions:

1. How do GSM estimates of neighborhood-level variance and autocorrelation compare to HLM estimates?
2. Which method (HLM or GSM) is more effective at modeling the autocorrelation actually observed in data from neighborhood residents?
3. How do GSM estimates of contextual effects and model fit compare to HLM estimates?
4. In a dataset originally collected with use of HLM methods in mind, how do the geographical scales on which different contextual factors operate (as estimated with GSM) compare to each other and to the size of the neighborhood units used in HLM?

The first two questions and the attendant hypotheses focus on how we group data in order to detect and model spatial variability in outcomes, while the latter two questions focus on how different ways of defining neighborhood boundaries affect the strength of the relationships between specific contextual characteristics and individual-level outcomes. By pursuing answers to these questions, this study contributes to the literature

on testing neighborhood effects and expands the methodological repertoire available to community psychologists interested in this topic.

Limitations

As the literature review above illustrates, the conceptual issues surrounding the definition and operationalization of neighborhoods and the testing of neighborhood effects are complex and deeply interrelated. The present study addresses some key issues, but no single study can address all of them completely. One of the study's limitations is that it focuses on only a single outcome measure. As a result, the findings will need to be replicated with other outcomes before strong conclusions about the generalizability of the findings to other outcomes can be drawn. This study focused on perceived neighborhood problems specifically because of the existing evidence that suggested it might be a strong candidate for use with GSM instead of HLM.

The study did not use any outcome measures where one might expect HLM to be more appropriate than GSM. This is another limitation, but one that reflects the fact that HLM is the more well-established method in the community psychology literature. Focusing on the situation where it is most plausible that GSM might outperform HLM is a crucial test of whether GSM might be a viable alternative to HLM; it is less critical to show that HLM can sometimes be more appropriate because it is already the de facto standard approach.

Another limitation of the study is that the data come from a single sample located in a small city. This means that the results should be replicated with additional studies drawn from other geographical study regions in order to assess the generalizability of the results beyond the selected study region. The size, shape, and spatial arrangement of the

neighborhood units used for HLM are fixed in this sample, but would certainly vary in other study regions and may play an important role in the comparison of HLM and GSM methods. Furthermore, while many HLM studies of neighborhood effects have been located in large cities, the setting for the present study was a small city. There are many differences between large and small cities, but what influence those differences might have on the use of HLM versus GSM is not known.

One option for comparing HLM and GSM would be to conduct simulation studies where these factors can be directly manipulated by the researcher, as could the location of the individual observations and the actual structure of the dataset. Designing a series of simulations to thoroughly explore the conditions under which HLM and GSM each perform best will be challenging because of the complex spatial issues involved. This study is only the first step toward introducing sophisticated GIS-based spatial statistics into community psychology and the literature on neighborhood effects. Therefore, it was deemed appropriate to use a real dataset for the study as a proof of concept before undertaking that more advanced kind of methodological research.

METHOD

The data selected for this study came from work related to the evaluation of *Yes we can!*, which is a community change effort funded by the W. K. Kellogg Foundation (WKKF) in Battle Creek, Michigan. Several key features of this dataset contributed to its selection for the study.

First, the survey sample comprising the main portion of the dataset was designed with multilevel neighborhood research in mind. The neighborhood units were constructed to have ecologically sensible boundaries (Van Egeren, et al., 2007) and conformed to one of the major suggestions in the HLM literature, which is to sample from neighborhood units that are as small as feasible in order to maximize between neighborhood variance (Roosa, et al., 2003). Because every survey participant's address was known and geocoded with high accuracy, the data were also easy to use in spatial analyses.

Second, the neighborhoods under study were all from a single city, placing them in close proximity to each other, which is important for comparing how HLM and GSM model autocorrelation. To fully do that comparison, the neighborhoods must be close enough together that it is plausible that the range of spatial autocorrelation and/or the geographical scale on which certain predictors are measured might reach across the boundaries between neighborhoods. In this dataset, that is absolutely plausible because there are multiple instances where neighborhood units were located very close together.

Third, additional secondary data sources with high spatial resolution were available to measure contextual characteristics without depending on aggregated survey data. Both crime and residential property value data were available for this study region in point-based GIS shapefiles geocoded to specific addresses—a form ideally suited to

enabling those data to be aggregated to construct contextual measures either within the boundaries of the fixed neighborhood units needed for HLM, or in the buffers centered on each resident's home that were used in GSM. Using predictors that were not aggregated from the survey data itself prevented shared method variance from biasing the study results. Furthermore, the nature of these secondary datasets allowed flexible re-specification of the size of the buffers used in the GSM analyses.

Study Context

The study region comprised a portion of the city of Battle Creek, which is a small city in southwest Michigan with a population of approximately 53,000 residents. During Phase I of *Yes we can!*, the work focused on a set of seven elementary school catchment areas (ESCA) that were selected on the basis of demographic, educational, and economic data. Because *Yes we can!* was being expanded in Phase II to focus on a larger geographic area, the 2005 resident survey collected by the *Yes we can!* evaluation team as baseline data for Phase II sampled residents from the original seven ESCAs, plus residents from several additional ESCAs. However, the ESCAs were large enough to contain areas with considerable heterogeneity in economic conditions and demographic composition, so the team developed a clustered sampling design based around much smaller and more ecologically meaningful neighborhood units (Van Egeren, et al., 2007).

These neighborhood units, shown in Figure 4, were defined by identifying 52 clusters of census blocks within block groups that met at least one of three economic risk criteria according to 2000 U.S. census data (median household income < \$30,568, percent of single-female-headed households living below poverty $\geq 49\%$, or percent of children under age 5 living below poverty $\geq 39\%$). Neighborhood units were only created from

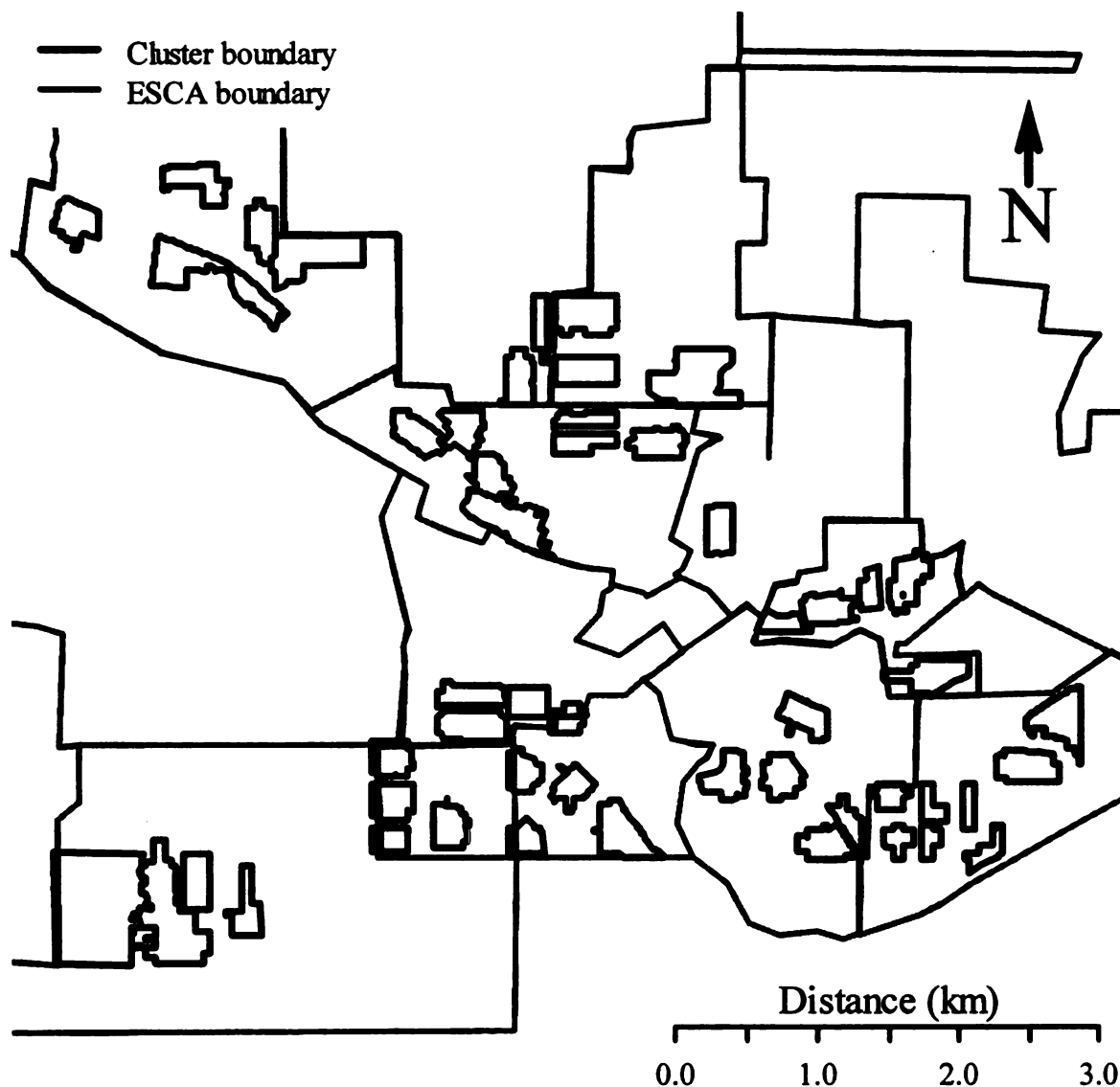


Figure 4: Map of neighborhood cluster boundaries ($N = 52$) and ESCA boundaries. These clusters were used as Level 2 units during the survey sampling and to represent ecologically meaningful neighborhoods for grouping residents in the HLM analyses. Source: Map prepared by the author.

clusters of census blocks that were all from the same block group and were not internally divided by ecological barriers such as major streets, bodies of water, or parks. The units each contained from 1 to 11 census blocks ($M = 5$, counting both whole blocks and partially-included face-blocks equally) and ranged in size from 0.026 to 0.472 km^2 ($M = 0.083 \text{ km}^2$, $SD = 0.069 \text{ km}^2$).

The neighborhood-level sampling was stratified based on whether the neighborhood unit was primed or unprimed to take advantage of *Yes we can!* This priming status variable reflected the evaluation team's expectation that neighborhoods with a history of previous activity focused on creating neighborhood change might be better positioned to benefit from the upcoming intervention activities. The primed neighborhoods either (a) had been identified by the city as having an active neighborhood association or (b) contained at least one active leader according to either *Yes we can!* community organizers or city lists of neighborhood association leaders and neighborhood planning council members. The unprimed neighborhoods had neither active neighborhood associations nor any identified leaders living on any of the included blocks. The 52 neighborhood units ultimately identified were evenly split with respect to priming status (26 primed, plus 26 unprimed).

For simplicity and clarity, the neighborhood units used in the survey sampling will hereafter be called neighborhood clusters (or just clusters). These clusters are the geographic units that were used to define the neighborhood boundaries in all the HLM analyses and some of the GSM analyses reported below.

Data Sources

Survey sample. The initial source for the survey sample frame was a GIS shapefile containing data about parcels of land in Battle Creek obtained from the local tax assessor's office. GIS tools were used to merge the cluster boundaries with a map of the parcels, delete records for parcels outside the cluster boundaries, and assign cluster identification numbers to the remaining records in the draft sample frame database.

Prior to drawing the survey sample, evaluation team members inventoried all the dwelling units on each parcel within each cluster. They identified dwelling units that were vacant, abandoned or uninhabitable, for sale, or advertised as being currently available for rental. Those units were deemed ineligible for selection during the sampling and deleted from the database accordingly. The final survey sample frame was expanded by splitting records for parcels with multi-unit dwellings into separate records for each distinct dwelling unit. Thus, all inhabited dwelling units located in any of the clusters were listed as unique rows in the sample frame database.

The survey sample was clustered. The evaluation team used simple random sampling within each neighborhood cluster, aiming to draw a minimum of 37 households from each neighborhood. This original target sample contained 1,905 households (a few neighborhood units contained fewer than 37 addresses, which prevented reaching the goal of 1,924 households).

In Fall 2005, surveys were mailed to the selected households at three-week intervals until each household either responded or three surveys had been sent without receiving a response. In the third round of mailings, the evaluation team was concerned that people who had already failed to respond twice would again be non-responders. To boost the final sample size, the target sample was augmented at the third mailing by adding one replacement household from the same neighborhood cluster for each non-respondent from the original sample (the original non-respondents still got the third mailing as well). The replacement households received a total of three opportunities to respond to the survey, at three week intervals. In total, surveys were mailed to 2,643 residential addresses, but 184 addresses were later deemed invalid due to vacancy,

undeliverable mail, etc., so the denominator for the response rate is 2459 valid addresses. Comparing the early versus late responders on 13 demographic variables and 44 other survey measures revealed almost differences between these subsets of the survey participants (Pierce, 2008) Prior to the first and third mailings, community residents hired by the evaluation team conducted door-to-door outreach to encourage residents to complete the survey. Each household that returned a completed survey received a \$30 gift card to a local store. Only one survey per household was included in the final sample.

Data collection was cut off in early 2006, 23 weeks after the first mailing. It yielded 1,049 usable surveys (a 42% response rate), which were equally divided between unprimed ($n = 522$) and primed ($n = 527$) neighborhoods. The number of usable surveys per cluster ranged from 8 to 31 ($M = 20.2$, $SD = 5.3$, $Mdn = 21$). Demographic characteristics of the sample are shown in Table 2.

Table 2: Demographic characteristics of survey participants (N = 1049)

Variable	Pre-Imputation			Post-Imputation	
	<i>N</i> or (<i>M</i>)	% or (<i>SD</i>)	Valid %	<i>N</i> or (<i>M</i>)	% or (<i>SD</i>)
Age (in years)	(47.00)	(16.15)		(46.87)	(16.25)
Non-missing	1001	95	100	1049	100
Missing data	48	5		0	0
Age category					
18-35	280	27	28	328	31
36-55	439	42	44	439	42
≥ 56	282	27	28	282	27
Missing data	48	5	0	0	0
Sex					
Male	266	25	26	268	26
Female	775	74	74	781	74
Missing data	8	1	0	0	0
Primary race/ethnicity					
White	640	61	65	672	64
Black or African American	293	28	30	312	30
Hispanic or Latino	42	4	4	45	4
Other	16	2	2	20	2
Missing data	58	6	0	0	0
Marital status					
Single	249	24	24	255	24
Married or cohabitating	479	46	46	483	46
Divorced or separated	217	21	21	217	21

Table 2 cont'd

Variable	Pre-Imputation			Post-Imputation	
	<i>N</i> or (<i>M</i>)	% or (<i>SD</i>)	Valid %	<i>N</i> or (<i>M</i>)	% or (<i>SD</i>)
Widowed	93	9	9	94	9
Missing data	11	1	0	0	0
Education (highest degree obtained)					
Did not graduate from high school	182	17	18	186	18
High school, GED, trade certificate	641	61	63	659	63
Undergraduate college degree	174	17	17	182	17
Graduate degree	21	2	2	22	2
Missing data	31	3	0	0	0
Employment status					
Not employed	444	42	57	446	57
Employed	598	57	43	603	43
Missing data	7	1	0	0	0
Home ownership					
Rent	330	32	33	342	33
Own	685	65	68	707	67
Missing data	34	3	0	0	0
Annual income					
< \$15,000	365	35	37	389	37
\$15,000 – \$25,000	205	20	21	217	21
\$25,000 – \$45,000	268	26	27	288	27
> \$45,000	146	14	15	155	15
Missing data	65	6	0	0	0
No. of children	(1.53)	(1.47)		(1.42)	(1.38)
Non-missing	715	68	100	1049	100
Missing data	334	32	0	0	0
Presence of children					
No children	218	21	31	340	32
Children (≥ 1)	497	47	70	709	68
Missing data	334	32	0	0	0
^a	(30.41)	(19.42)			
Years in BC					
Non-missing	1031	98	100		
Missing data	18	2	0		
^a	(12.19)	(14.24)			
Years at current address					
Non-missing	1005	96	100		
Missing data	44	4	0		

Note. Percentages may not total to 100 due to rounding error. BC = Battle Creek, *M* = mean, *SD* = standard deviation.

^a Post-imputation summaries are not shown for years in BC and years at current address because these variables were excluded from both the imputation model and the analyses reported below.

Crime data. As part of the ongoing *Yes we can!* evaluation, the evaluation team also obtained crime data from the City of Battle Creek Police Department. This secondary dataset contains an electronic list of all the crime incidents (*N* = 8,263) reported to the police in the 12 months prior to the 2005 resident survey, drawn from the

police dispatch records management system. In addition to the address at which each incident occurred, additional attributes of each incident are also available, including up to four offense codes that can be used to determine the type of crime. For this study, offense codes were used to determine whether each incident fell into one or more of the three major categories of crime used in the National Incident-Based Reporting System (Uniform Crime Reporting Program, 2000): crimes against persons, which are all essentially violent crimes such as assault, murder, and rape; crimes against property such as theft, arson, and fraud; and crimes against society such as drug/narcotic offenses, prostitution, and gambling, which are violations of laws that “represent society’s prohibitions on engaging in certain types of activity” (Uniform Crime Reporting Program, 2000, p. 14).

The crime data were converted into a point-based GIS shapefile by geocoding each incident address against a street centerline shapefile. Mapping the locations of the crime incidents showed that 529 of them actually occurred slightly outside the city boundary (mostly along a single highway). Given that most crimes occurring outside the city limits were probably handled by police from other jurisdictions from whom no data had been obtained, only the 7,734 incidents that fell within the City of Battle Creek’s official boundary were used for computing contextual measures in this study.

Property data. The *Yes we can!* evaluation team also obtained property data from the City of Battle Creek’s assessor’s office. This polygon-based GIS shapefile contains data about parcels of land within the local tax assessor’s purview (e.g., within the City of Battle Creek’s official boundaries). The database includes the address and boundaries of each parcel of land, along with information about property class, zoning code, and more.

Among those variables are indicators of whether the parcel is zoned to contain a single family dwelling, a two family dwelling, or a multifamily residential dwelling. Some residential parcels are zoned to contain medium density or high density residential dwellings (e.g., large apartment buildings). Data about the 2005 property value associated with each residential parcel was obtained from the assessor's office, which maintains historical data of these public records for tax purposes, then merged with the GIS shapefile so that residential property values could be aggregated within different neighborhood boundaries as required for the study. To facilitate that aggregation, the polygon representing each parcel was converted to a point located at the parcel's centroid, yielding a point-based shapefile.

Procedures

Survey consent. The 2005 resident survey used a passive consent procedure. As explained in the cover letter sent along with the survey, returning a completed survey served as informed consent to use the survey for the original purpose of the study, which was to evaluate the *Yes we can!* effort and assess conditions in Battle Creek. The evaluation of *Yes we can!* and academic research based on the 2005 survey data are ongoing and have been approved by the institutional review board at Michigan State University. The principal investigator for that work is Dr. Pennie Foster-Fishman.

The present study entailed secondary analysis of that survey data for a new research purpose that posed only minimal risk to the survey participants. It would have been exceedingly difficult to contact all 1,049 of those residents to obtain consent to re-use their data for this new purpose, so the institutional review board at Michigan State University waived the requirement to obtain further consent from the residents.

Geocoding. All three data sources listed above were geocoded and projected into the Michigan State Plane (South) Coordinate System of 1983 (Lusch, 2005), with units for the spatial coordinates set to meters. The resulting point- and polygon-based shapefiles can be plotted on maps using GIS software and were imported into the statistical software used for the analyses.

Surveys. Because the survey sample frame was initially constructed from a GIS shapefile based on the local tax assessor's property database, nearly every survey returned was geocoded (assigned spatial coordinates so that their locations can be plotted on electronic maps) by the evaluation team by simply linking the address of the survey participant back to the GIS files containing the parcel data. The geocoded location for each survey participant is the centroid of the parcel containing the participant's residential address. This resulted in a very high geocoding rate (over 98%). The remaining survey participants' locations were manually geocoded by referring to maps annotated by the *Yes we can!* evaluation team when taking the dwelling unit inventory they used to refine the survey sample frame.

There were 39 survey participants whose spatial coordinates were identical with those of at least one other participant because they lived on parcels containing multiple dwelling units (e.g., apartment buildings or duplexes). Because exact overlap in the locations of the data points causes mathematical problems in GSM, a trivial amount of spatial error (up to 3 m in either direction along each axis) was added to the spatial coordinates for these participants by adding independently drawn random values from a uniform distribution to both the easting and northing coordinates for those 39 cases. This eliminated exact overlap and allowed all cases to be retained in the GSM analyses.

Crime data. Because the crime incident files contained addresses and other location data, 99% of the crime incidents were successfully geocoded by matching the incident addresses to a street centerline GIS file, so each crime is associated with a point in geographic space. This hit rate far exceeds the minimum acceptable hit rate of 85% for geocoding crime data recommended by Ratcliffe (2004). The resulting geocoded crime data constitute a spatial point pattern that can be analyzed with a variety of spatial analysis techniques (Bailey & Gatrell, 1995).

Property data. The property data were available as a polygon-based GIS shapefile showing the precise area occupied by each parcel of land falling under the purview of the local tax assessor's office. These data were geocoded by employees of the City of Battle Creek and are the most authoritative, accurate, and highest resolution spatial data available for property parcels in that city. Because the shapefile contains the entire territorial boundary for each parcel, parcel centroids were easily computed and were used to determine whether or not particular parcels fell within particular geographic areas.

Neighborhood-Level Contextual Measures

The contextual measures for this study were computed from the crime and property datasets. Those datasets were aggregated in several ways to construct contextual measures suitable for the present analyses. For each construct (crime and NSES), data were aggregated into variables representing (a) the geographic area of each neighborhood cluster as defined by the original sampling design, (b) a series of 25 concentric, circular buffers centered on each survey participant's home that varied in size, with radii ranging from 0.10 km (0.06 mile) to 2.50 km (1.55 mile) in 0.10 km increments. For comparison,

a typical face block in Battle Creek is about 0.12 km (0.07 mile or 400 feet) in length.

These buffers, with the modifications noted below, were used in the GSM analyses.

The geographic coverage of the property and crime data was constrained: they were only consistently available within the official City of Battle Creek boundary. The neighborhood clusters all fell entirely within the city boundary, so that did not pose a measurement problem for the HLM analyses. However it posed an edge-effect problem (Bailey & Gatrell, 1995) for measuring neighborhood-level variables in the GSM analyses because buffers for residents living near the edge of the city sometimes covered land outside the city limits, where crime and property value data were not available. This was addressed with an edge-correction procedure inspired by techniques used in spatial point-pattern analyses (Bailey & Gatrell, 1995): only the portion of a circular buffer falling within the city limits was used to measure the neighborhood-level variables. This affected both the shape and the size of the buffers for some residents, particularly those in the southeast corner of the study region, but ensured that only the geographic area over which crime and residential property data were reliably available was considered.

Consistent with recommendations in the multilevel modeling literature, both neighborhood-level measures were grand-mean centered prior to analysis (Enders & Tofighi, 2007). This practice made the coefficients associated with these predictors more interpretable.

Crime. Only crimes that occurred in the 12 months immediately preceding the collection of the survey data were used to calculate crime density figures. The raw neighborhood crime density (crimes/km²) was measured by aggregating the crime data to determine the total number of crime incidents that occurred within the cluster boundaries

(which were dilated 15 m outward to capture crimes geocoded into the middle of streets) and within each of the buffer boundaries, then dividing by the land area enclosed by each of those neighborhood boundaries. To avoid numerical problems in estimating the crime coefficients and to make the units of the variable more sensible, raw crime density values were divided by 10 prior to centering the variable for use in analyses. A one unit difference on the final crime variable therefore represents a difference of 10 crimes/km².

Most neighborhood research uses crime measures based on only violent crime (i.e., crimes against persons such as assault, homicide, and rape) (Franzini et al., 2008; Quillian & Pager, 2001; Sampson & Raudenbush, 2004). Because it was possible that nonviolent crimes might also influence residents' perceptions and that different types of crime might operate on different spatial scales, the utility of using separate measures based on crimes against persons, property, and society was considered for this study. In preliminary analyses, all of these measures were significant predictors of the outcome when entered as the sole neighborhood-level predictor, as was an overall crime measure based on all incidents regardless of type. Crime against persons was the strongest predictor in those analyses.

Further preliminary modeling showed that while individual crime incidents very rarely included offenses falling into more than one crime category, neighborhood-level crime measures for the different types of crime were very strongly correlated. Including more than one of them in a model inevitably caused severe multicollinearity problems, eliminating the effects of all the included crime measures. Given that, it was not feasible to test whether the HLM and GSM yield different patterns of scientific inferences about the effects of different kinds of crime. Therefore, consistent with how crime is

operationalized in other neighborhood research studies (Franzini et al., 2008; Quillian & Pager, 2001; Sampson & Raudenbush, 2004), the crime measure in the analyses reported below is based only on crime against persons (Uniform Crime Reporting Program, 2000).

Neighborhood SES. NSES was measured by aggregating the property data to obtain the median value of all residential property parcels within a given set of neighborhood boundaries (either the cluster boundaries or the various buffer boundaries). While it might have been better to measure NSES in terms of the median value of each dwelling unit, the number of dwelling units per parcel was only available within the cluster boundaries. Such a measure therefore could not be calculated for the full range of buffers used in the GSM modeling. Similarly, relying on the median value of parcels zoned only for single-family dwellings was not feasible because some survey participants lived in areas exclusively zoned for higher density housing arrangements (containing many of apartment building complexes). The resulting missing data for NSES would have reduced the sample size and the generalizability of the results.

The raw NSES values were measured in dollars. To avoid numerical problems in estimating the NSES coefficients and to make the units of the variable more sensible, raw NSES values were divided by 1000 prior to centering the variable for use in analyses. A one unit difference on the final NSES variable therefore represents a \$1,000 difference in the median value of residential parcels in the neighborhood.

Individual-Level Measures

The individual-level measures for the study all came from the 2005 resident survey collected by the *Yes we can!* evaluation team. Because the study focuses on testing contextual effects, the individual-level predictors were grand-mean centered

(Enders & Tofighi, 2007; Paccagnella 2006) to improve interpretation of the coefficients. Although the literature on centering categorical variables in multilevel models focuses only on dichotomous predictors (Enders & Tofighi, 2007), grand-mean centering can also be applied to categorical predictors with more than two categories. Dummy coding them then separately centering each resulting dummy variable yields results that are conceptually comparable to centering dichotomous variables (C. Enders, personal communication, August 17, 2009).

Neighborhood problems. The dependent variable was residents' perceptions of neighborhood problems. This four item scale ($\alpha = .88$) was based on items adapted from Coulton, Korbin, and Su's (1996) disorder scale. Residents were asked how much they agreed or disagreed with a set of statements about whether selected indicators of physical and social disorder were a problem in their neighborhood (e.g., "Crime is a problem" and "Abandoned, vacant, or neglected buildings are a problem") using a 6-point Likert scale (1 = strongly disagree, 6 = strongly agree).

Age. Residents self-reported their birth years in the demographic portion of the survey. Age was calculated by subtracting each resident's birth year from 2006. For the analyses, age was categorized into three groups: 18-35 years (the reference group), 36-55 years, and 56 or more years.

Gender. Residents self-reported their gender in the demographic portion of the survey. For this binary variable (0 = male, 1 = female), males were the reference group.

Race. The survey also collected self-reported racial/ethnic background. For this study, each participant was categorized into one of the following race groups: White (the reference group), Black/African American, Hispanic/Latino, or other.

Marital status. Residents were also asked about their marital status. For this study, marital status was collapsed into the following four groups: single (the reference group), married or cohabiting, separated or divorced, and widowed.

Education. Participants' were asked to report the highest level of education they had completed. For analysis purposes, education was collapsed into four categories: (a) Did not graduate high school, (b) high school diploma, general educational development (GED), trade or training certificates, (c) undergraduate college degree (Associate's, Bachelor's), or (d) graduate degree (Master's or Doctoral). Residents in the second category (i.e., high-school diplomas or similar level of education) served as the reference group because they comprised the largest group.

Employment status. Participants were also asked about their employment status. Unemployed participants were the reference group (0 = not employed, 1 = employed).

Income. Annual household income was collected by asking participants to report which of the nine different income categories included their income. Due to the small numbers of cases in some of the original categories, this variable was recoded into four categories: (a) less than \$15,000, (b) \$15,000 to \$25,000, (c) \$25,000 to \$45,000, and (d) \$45,000 and above. The highest income category was the reference group.

Home ownership. Participants were asked whether they rented or owned their home on the 2005 survey. Renters were the reference group (0 = rent, 1 = own).

Presence of children in the home. Residents were asked to report the number of children (persons under age 18) living in their home. Presence of children in the home was treated as a binary variable (0 = none, 1 = 1 or more children). Residents without children living in their home were the reference group.

Analysis

The analysis began with a thorough inspection of the data and assessment of the amount and patterns of missing data (McKnight, McKnight, Sidani, & Figueredo, 2007). After imputing missing data (see below), exploratory analyses (e.g., univariate and multivariate summaries, screening for outliers, etc.) and graphical methods for visualizing the data (Cleveland, 1993) were employed to check assumptions underlying the statistical models and the nature of the relationships between the variables.

The analyses associated with H1- H6 compared HLM and GSM with respect to how they group residents for detecting and modeling spatial variation in outcomes. They sought to test whether conceptualizing neighborhoods as places in discontinuous geographic space is a practice that should be replaced by conceptualizing neighborhoods as places within continuous space. In contrast, the analyses associated with H7 and H8 focused on comparing HLM and GSM as methods for testing the effects of specific neighborhood-level predictors. These latter analyses sought to inform our thinking about defining neighborhood boundaries for measuring contextual variables.

Imputation of missing data. To maximize the usable sample size and minimize the impact of missing data on the analyses, missing values on all individual-level measures were imputed prior to conducting the analyses (Schafer & Graham, 2002). The amounts of missing data were quite small for most of the variables (see Table 2) and the missing values were scattered throughout the dataset, so single imputation (rather than multiple imputation) was a reasonable strategy for dealing with the missing data.

Because the survey data were originally collected via a clustered sampling design, a multilevel imputation model (Schafer, 1997b, 2001) would normally be used to impute

missing values. However, the present study avoided biasing the results in favor of either HLM or GSM by applying a strictly individual-level imputation model designed for imputing missing values in datasets containing both categorical and continuous measures (Schafer, 1997). This technique ignored both spatial autocorrelation and hierarchically structured autocorrelation, giving neither analysis method an advantage.

The imputation model combined a log-linear model containing all possible main effects and two-way interactions among seven categorical variables (gender, race, marital status, education, employment status, home ownership, and income) with a regression model containing main effects for seven continuous variables. The continuous variables were neighborhood problems, age, and number of children in the home, plus four measures not described above because they were only used in the imputation process: hope (3 item scale, $\alpha = .83$), perceived availability of safe, affordable housing in the neighborhood (1 item), perceived barriers to employment (5 item scale, $\alpha = .90$), and parental support for education (2 item scale, $\alpha = .78$). The latter four measures were useful imputation covariates because they were all correlated with neighborhood problems ($r_s \geq .20$, $p_s < .05$).

Bayesian modeling. The software selected for estimating the GSM models relies on a fully Bayesian approach to statistical inference called Markov chain Monte Carlo (MCMC) estimation via Gibbs sampling (Finley, Banerjee, & Carlin, 2007). The same approach was also used to estimate the HLM models to ensure that differences in estimation methods could not skew the results toward either HLM or GSM.

The Bayesian approach to statistics is deeply grounded in probability theory, so models are specified in terms of joint probability distributions for all observable and

unobservable quantities in the problem at hand (Gelman, Carlin, Stern, & Rubin, 2004).

Observable quantities are the variables contained in the data, while unobservable quantities include the unknown values of parameters associated with specific predictors or other aspects of the model (e.g., error variances). Instead of depending on null hypothesis significance testing for point estimates of unknown model parameters, Bayesian modeling emphasizes describing and drawing inferences from the conditional probability distributions of those parameters given the observed data by examining summary statistics such as credible intervals (Gelman, et al., 2004; Gill, 2008).

Understanding the relationship between the prior distributions for unknown model parameters (often just called the priors) and the corresponding posterior distributions estimated for those parameters during a specific analysis is crucial to Bayesian statistics. The priors specified in a model reflect assumptions about the distributions of the unknown parameters that are made before examining new data (Gelman, et al., 2004; Gill, 2008). One key aspect of defining a prior is choosing the sampling distribution (e.g., normal, binomial, Poisson, etc.) that determines its overall shape and defines what parameters must be estimated (e.g., mean and variance for a normal distribution). Priors may be specified based on knowledge extracted from the relevant substantive literature, previously collected data, or based on methodological considerations. Bayesian modeling then uses the information in new data to update the prior distributions, thereby producing posterior distributions for model parameters that are more informative than the priors and can be used to draw scientific inferences (Gelman, et al., 2004; Gill, 2008).

This study used non-informative prior distributions to minimize the influence of the priors on the posterior distributions (Gelman, et al., 2004; Gill, 2008), ensuring that

the posterior distributions largely reflect information gleaned from the actual data. To rule out the priors as a potential explanation for differences in performance, identical priors were used for corresponding parts of the HLM and GSM models. Flat priors based on the normal distribution ($\mu = 0$, $\sigma^2 = 10,000$) were used for all intercept and slope coefficients. This effectively assumed that the distribution is centered on zero (on average, there is no effect) and that all values for these parameters (even large positive or negative values) were equally likely, but each had a very low probability of occurrence.

While there is an ongoing debate about the best non-informative prior distribution to use for variance components in HLM, inverse gamma priors are widely used (W. J. Browne & Draper, 2006a, 2006b; Gelman, 2006; Van Dongen, 2006) and have also been recommended for variance components in GSM (Banerjee, et al., 2004). Accordingly, inverse gamma priors (shape = 2, scale = 1) were used for the variance components associated with individual- and neighborhood-level residuals in the regular HLM models. This prior distribution is positively skewed (minimum = 0, $\mu = 1$, and $\sigma^2 = \infty$), so adopting it assumed that values at or near zero were most likely to occur, but large positive values were also possible.

For one HLM model, the assumption of independence between neighborhoods was relaxed by supplementing the typical spatially unstructured neighborhood-level residual with an additional, spatially structured residual. This spatial random effect was assigned a Gaussian CAR prior distribution via the `car.normal` function in WinBUGS. Under this prior, the expected value for a neighborhood's spatial residual is the weighted average of the spatial residuals from surrounding neighborhoods (defined here as those

whose centroids were within 2.0 km of the focal neighborhood's centroid). The weight matrix (**W**) for the CAR HLM model used unstandardized inverse distances to make this spatial effect more conceptually comparable to the GSM models. Thus, the CAR HLM model is a hybrid approach that incorporates hierarchical autocorrelation through the spatially unstructured neighborhood-level residuals and their corresponding variance component and a distance-decay form of spatial autocorrelation represented by the spatially structured, neighborhood-level CAR residuals and their variance component.

The conditional variance of the prior distribution for each CAR residual is inversely proportional to the number of surrounding neighborhoods. Per the recommendation in the WinBUGS documentation (Thomas, Best, Lunn, Arnold, & Spiegelhalter, 2004), the CAR variance component was assigned an inverse gamma prior (shape = 0.05, scale = 0.0005) to avoid inducing artificially high levels of spatial autocorrelation. The expected value for the CAR variance under this prior is .0025 and there is a 98% prior probability that it will fall between 0.0001 and 6.25.

It was also necessary to select a shape for the variograms incorporated into the GSM models and set priors for the parameters that define how far spatial autocorrelation extends. There is no corresponding range parameter in the HLM models, so this prior was based solely on recommendations in spatial analysis literature. Isotropic exponential variograms were selected because they were a) reasonable fits to the observed pattern of autocorrelation in these data, b) used in other GSM studies (Chaix, et al., 2006; Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian, et al., 2005; Finley, et al., 2007), and c) recommended in spatial analysis texts (Banerjee, et al., 2004). An exponential variogram has a parameter (ϕ) for the rate of decrease in spatial autocorrelation; its value

determines the practical range beyond which remaining autocorrelation is negligible. Because empirical variograms become unstable at large distances, it is wise to limit their ranges to less than the maximum distance between observed points (Diggle & Ribeiro, 2007). A uniform prior distribution (minimum = 6.49×10^{-4} , maximum = 1) was selected for ϕ ; this constrained the practical range to fall between 3 m and 4622 m because practical range = $-\log(0.05)/\phi$, which is approximately $3/\phi$. That upper limit is half the distance across the study region (i.e., half of the diagonal for a rectangle just large enough to fully enclose all 52 clusters).

MCMC estimation of a Bayesian model consists of simulating an iterative series of random draws from the joint posterior distribution of all the model parameters. This produces a chain of estimates for each parameter that can be summarized with familiar univariate statistics to describe the parameter's posterior distribution (Gill, 2008). For example, the results reported below usually contain the mean and standard deviation of each parameter's distribution, along with the central 95% credible interval (i.e., the interval endpoints exclude the bottom and top 2.5% of the values in the distribution).

Because consecutive draws for each parameter are typically mildly autocorrelated with previous estimates in the same chain, MCMC simulations must be run for many iterations to produce reasonable empirical estimates of the posterior distributions (Gill, 2008). Running multiple MCMC chains for each parameter helps the analyst assess whether they are converging to the same distribution (Gelman, et al., 2004; Gelman & Hill, 2007; Gill, 2008). Accordingly, each HLM model was run in three separate chains with independent, random starting values. Each chain ran for 16,000 iterations, but the first 1,000 iterations in each chain were discarded as a burn-in period to ensure that the

results were no longer influenced by the initial values (Gill, 2008). Final results for each parameter in those models were based on pooling the remaining 45,000 posterior estimates from all three chains.

The main GSM models were also run in three separate chains with independent, random starting values. Those MCMC chains ran for 32,000 iterations each, but the burn-in period discarded the first 7,000 iterations. Final results for parameters in those models were based on pooling the remaining 75,000 posterior estimates from all three chains. Due to the extraordinarily high computational cost of estimating GSM models (1.17 to 2.55 days of computing time per chain), the 50 GSM models used to determine the optimal buffer sizes for the two neighborhood-level predictors were run with only one chain each (these also ran for 32,000 iterations and had 7,000 iteration burn-in periods).

Model building sequence. It was necessary to run a large number of models to test the hypotheses. The analysis required estimating a series of parallel HLM and GSM models of increasing complexity, as illustrated in Table 3. The first four steps in this model building process were identical for both methods and used cluster-based measures of the neighborhood-level predictors, yielding models 1-5 for each method. Empty models without any substantive predictors were estimated first, then all the individual-level predictors were added simultaneously in the second step. Next, crime and NSES were added (each by itself in the third step, then both together in the fourth step). The remaining model-building steps diverged for HLM and GSM. The fifth step in the HLM modeling added a spatial random effect to produce the CAR HLM model (i.e., HLM Model 6).

Table 3: Overview of primary model building sequence

Model No.		Predictors Included	Chains
HLM	GSM		
1	1	None (empty models)	3
2	2	Individual	3
3	3	Individual + Crime ^a	3
4	4	Individual + NSES ^a	3
5	5	Individual + Crime ^a + NSES ^a	3
6		Individual + Crime ^a + NSES ^a + CAR	3
	6-30	Individual + Crime ^b	1
	31-55	Individual + NSES ^b	1
	56	Individual + Crime ^c	3
	57	Individual + NSES ^c	3
	58	Individual + Crime ^c + NSES ^c	3

Note: Neighborhood problems was the dependent variable for all of these models. CAR = conditional autoregressive (spatially-structured) random effect on neighborhood-level intercepts; Chains = number of MCMC chains run per model; NSES = neighborhood socioeconomic status; Individual = all individual-level predictors (age, gender, race, marital status, education, employment status, income, home ownership, and presence of children in the home);.

^a Predictor measured in cluster boundaries. ^b Predictor measured in buffers varying in size (radii ranging from 0.10 km to 2.5 km, in 0.1 km increments); these models were used to identify the optimal buffer size. ^c Predictor measured in optimal size buffer (1.1 km radius for crime, 0.2 km radius for NSES).

Starting in the fifth GSM modeling step, buffer-based measures of crime and NSES replaced the cluster-based measures. That step also systematically varied the buffer sizes used for measuring crime (Models 6-30) and NSES (Models 31-55) to determine the optimal buffer size for each of those predictors, using only a single MCMC chain for each model. The sixth GSM modeling step used the optimal buffer sizes for crime and NSES to estimate GSM models 56-58 (once again using three MCMC chains per model), which paralleled the models estimated in the third and fourth steps.

This model-building sequence provided the statistical output required to test the study hypotheses and answer the research questions. Details on how each hypothesis was tested are presented below.

Testing Hypothesis 1. Estimated variance components and measures of autocorrelation (i.e., ICC for HLM models and PSR for GSM models) from HLM Models 1 and 2 were compared to corresponding estimates from GSM Models 1 and 2 to test H1. For example, HLM Model 1 and GSM Model 1 were empty models that used only intercept and random effects terms, so their neighborhood-level variance components and estimates of autocorrelation were compared to test H1. Similarly, HLM Model 2 and GSM Model 2 were compared to examine whether adding individual-level predictors changed the results. Evidence that the neighborhood-level variance from the GSM was larger than the corresponding estimate from the HLM was interpreted as support for H1, as was evidence that the PSR was larger than the ICC by 5 percentage points.

In studies comparing OLS ANOVA models (which assume $ICC = 0$) to HLM models, ICCs as low as 0.05 seriously inflate the Type I error rate when the sample size per level 2 unit is around 20 as it is in this study (Barcikowski, 1981; Zucker, 1990). This suggests that if $(PSR - ICC) > 0.05$, then the significance levels of the contextual effects might differ substantially between HLM and GSM, with HLM producing artificially inflated t-statistics and therefore possibly increasing the risk of Type I errors.

Testing Hypothesis 2. Testing H2 involved comparing the estimated practical ranges of spatial autocorrelation from GSM models 1 and 2 to the distribution of the distances between the neighborhood clusters. This is a purely descriptive analysis that can be accomplished by plotting a frequency distribution of the distances between

clusters and adding a vertical reference line to the plot at the distance that would correspond to that range parameter. If that reference line bisects the body of the distribution, then would suggest that spatial autocorrelation could indeed spill over from one neighborhood cluster into another cluster.

Testing Hypothesis 3. The deviance information criterion (DIC) is a Bayesian measure of model fit (Chaix, et al., 2006; Spiegelhalter, et al., 2002) that can be used for model selection among either nested or non-nested models fit to the same data (Gelman & Hill, 2007; Gill, 2008; Spiegelhalter, et al., 2002). Roughly speaking, the DIC indicates how well a model might predict responses in a new dataset (lower DICs indicate better model fit).

Because using more parameters always improves a model's fit to the data, it is important to select models that minimize prediction error while remaining parsimonious (i.e., use as few parameters as possible). As a generalization of the Akaike's information criterion (AIC) (Ntzoufras, 2009), the DIC is not a pure measure of model fit. Instead, it is adjusted to penalize the fit statistic for model complexity (Spiegelhalter, et al., 2002). Thus, the DIC attempts to make a reasonable tradeoff between model fit and model complexity to avoid selecting models that only fit better because they use larger numbers of parameters. In this context, complexity is measured by a statistic called pD that represents the effective number of parameters in the model (Spiegelhalter, et al., 2002). Complex models have large pD values while simpler, more parsimonious models have small pD values. Given two models with equal fit (as measured by the deviance statistic), the more parsimonious model would have a lower DIC because a smaller pD value would lead to a smaller penalty for model complexity.

To test H3, the DIC was used to compare HLM Model 1 to GSM Model 1, and to compare HLM Model 2 to GSM Model 2. Differences of 3 or more points on the DICs between two models were interpreted as evidence that the model with the lower DIC fit the data better than the model with the higher DIC (Spiegelhalter, et al., 2002). Additional criteria for evaluating the difference in model fit included the deviance statistic (D); the proportional change in neighborhood-level variance (level 2 PCV), which is a level 2-specific R^2 (Raudenbush & Bryk, 2002); and the overall R^2 , which measures the proportion of variance explained by estimating the correlation between observed and predicted values (Roberts & Monaco, 2006).

Testing Hypothesis 4. To test H4, the level 1 and level 2 residuals from the HLM models estimated to test H1 were examined for evidence of spatial autocorrelation. An empty GSM model with the level 1 HLM residuals as the dependent variable was used to estimate the PSR, thereby quantifying the amount of spatial autocorrelation evident in those residuals. If HLM fully controlled for spatial autocorrelation in the raw data, it was expected that the 95% CI for the PSR would contain zero.

A different strategy was required to test the level 2 HLM residuals for spatial autocorrelation because they are tied to the neighborhood clusters rather than to the locations of the individual survey participants. They were essentially areal data rather than point-based data. One-sided, exact Moran's I tests for regression residuals were used to test for spatial autocorrelation in level 2 HLM residuals (Bailey & Gatrell, 1995; Chaix, Merlo, Subramanian, et al., 2005). This test depends on defining a spatial weights matrix that specifies which observations are considered neighbors (Bivand, Pebesma, & Gómez-Rubio, 2008). To maintain consistency with how spatial autocorrelation was

represented in the GSM models, weight matrix entries associated with pairs of clusters whose centroids were up to 4,622 m apart were assigned row-standardized, inverse-distance spatial weights. Weight matrix entries for clusters separated by more than 4622 m were assigned spatial weights of zero.

Testing Hypothesis 5. To test H5, individual-level residuals from the GSM models estimated to test H1 were examined for remaining hierarchical autocorrelation by using them as the dependent variables in empty HLM models. The 95% CI for the resulting ICC was then examined to see whether it contained zero. An ICC = 0, or an ICC that is trivially small, then H5 would be supported.

Testing Hypothesis 6. Testing H6 was very similar to testing H5; the procedure differed only in that now the neighborhood-level GSM residuals from the models estimated to test H1 were examined. In this case, support for H6 would consist of a significant LRT paired with an ICC estimated from the HLM run on the GSM residuals that is smaller than the PSR associated with the original GSM estimated when testing H1.

Testing Hypothesis 7. Testing H7 involved several steps. As a preliminary first step, two series of GSM models were used to determine the spatial scales on which crime and NSES operate. The buffer size for measuring crime was systematically varied in GSM Models 6-30 while controlling for all the individual-level predictors, then those models were compared to each other to determine the optimal buffer size. The optimal buffer sizes were chosen by selecting the model from each series that had (a) large absolute values for the regression coefficient and the corresponding t-statistic; (b) large values for the level 2 PCV and overall R^2 statistics; and (c) low values for PSR, the

practical range of spatial autocorrelation, and the DIC. The same process was used with NSES buffers in GSM Models 31-55.

There are several ways to calculate R^2 statistics that measure the overall proportion of variance explained by models like HLM and GSM (Edwards, Muller, Wolfinger, Qaqish, & Schabenberger, 2008; Gelman & Pardoe, 2006; Kramer, 2005; Orelieu & Edwards, 2008; Roberts & Monaco, 2006; Xu, 2003); some focus only on variance explained by the fixed effects (Edwards, et al., 2008) while others use both the fixed and random effects (Kramer, 2005), thereby enabling one to examine how much additional variance is explained by modeling correlated error structures. These variations on the R^2 statistic are useful supplements to measures such as the DIC for assessing model fit and comparing models (Kramer, 2005). This study used Roberts and Monaco's Equation 15 to calculate the overall R^2 , which is reproduced below as Equation 11. In that equation, σ'^2_{total} is the total sum of squares based on using only the grand mean as a predictor and σ'^2_{error} is a residual sum of squares based on predicted values that account for the effects of all the predictors included in the model. This method should be equally valid for HLM and GSM models because it measures the squared correlation between the observed and predicted values.

$$R^2 = 1 - (\sigma'^2_{\text{error}} / \sigma'^2_{\text{total}}). \quad (11)$$

In the second step, H7 was tested by conducting three-way model comparisons between (a) HLM models with cluster-based measures of crime and NSES, (b) GSM

models with cluster-based measures of crime and NSES, and (c) GSM models that used optimally-sized buffers to measure crime and NSES. Comparisons focused on models that had parallel structure, such as HLM Model 3 and GSM Models 3 and 56, which all contained individual-level predictors plus a neighborhood-level crime measure. Similarly, HLM Model 4 and GSM Models 4 and 57 were compared to each other to examine the effect of NSES, while HLM Models 5 and GSM Models 5 and 58 were compared to examine the combined effects of crime and NSES. These model comparisons all used multiple criteria: regression coefficients, t-statistics, neighborhood-level variance components, level 2 PCV values, measures of residual autocorrelation (ICC and PSR), overall R^2 values, and the DIC. It was expected that support for H7 would be evident if the two GSM models both had larger crime and NSES effects and better fit than the HLM model, but GSM models based on using buffers performed better and had stronger crime and NSES than the GSM models based on using cluster boundaries.

In the third step, H7 was further tested by comparing HLM Model 6 to GSM Models 5 and 58 to see whether enhancing HLM with a CAR structure for the level 2 residuals changed any of the conclusions from the previous tests of H7. Again, support for H7 was expected to take the form of HLM performing worse than the GSM models.

Testing Hypothesis 8. Testing H8 involved comparing the optimal buffer size for measuring crime (as determined from GSM Models 6-30) to the optimal buffer size for measuring NSES (as determined from GSM Models 31-55). If those two buffers are of different sizes, then H8 would be supported. Similarly, if those two buffers each differ from the average size of the neighborhood clusters, that would be additional support for H8. This is a purely descriptive analysis that can be accomplished by examining a graph

that compares the sizes (in km²) of the optimal crime and NSES buffers and also shows how they relate to the distribution of cluster sizes.

Criteria for evaluating and comparing models. Table 4 summarizes the set of criteria that were used to evaluate model quality and to compare alternative models. Each entry in the table links a criterion to the hypotheses to which it is relevant and describes what evidence would constitute support for the hypothesis.

Table 4: Criteria for evaluating and comparing models

Criterion	Comments
H1: Size of the neighborhood-level variance components estimated by HLM and GSM	Comparing the neighborhood-level variance components from HLM and GSM models speaks to which method detects more neighborhood-level variability in outcomes (i.e., how much neighborhoods matter). Finding that the GSM estimate is larger than the HLM estimate would support H1. The strength of the evidence for or against H1 using this criterion can be quantified by the percent overlap in the Bayesian credible intervals ⁹ (BCIs) for these variance components (smaller overlap indicates more evidence for a difference between the estimates).
H1: Difference between ICC and PSR measures of autocorrelation	The magnitude and sign of the difference between the ICC and the PSR indicates which method (HLM or GSM) detects more autocorrelation. Finding PSR > ICC by 5 or more percentage points would support H1. The strength of the evidence for or against H1 using this criterion can also be quantified by the percent overlap in the BCIs for the ICC and PSR: the smaller the overlap, the stronger the evidence for a difference.
H2: Practical range of GSM variograms	Observing that the estimated practical range of spatial autocorrelation from a GSM model is larger than a substantial portion of the pairwise distances between the neighborhood clusters would support H2.
H3 & H7: Deviance information criterion (DIC)	The DIC is a Bayesian measure of model fit that adjusts for model complexity (Spiegelhalter, et al., 2002); it is used for model selection among either nested or non-nested models fit to the same dataset (Gelman & Hill, 2007; Gill, 2008; Spiegelhalter, et al., 2002) and for comparing alternative GSM models (Chaix, et al., 2006; Finley, et al., 2007). Lower DIC values indicate better fit, with a difference of 3 or more points between two models indicating a considerable difference in model fit (Spiegelhalter, et al., 2002).
H4: Moran's <i>I</i> estimated from neighborhood-level HLM residuals	Because HLM assumes that the neighborhood-level residuals are independent, violation of this assumption is an indication of poor model fit. Moran's <i>I</i> was used to test for spatial autocorrelation in these residuals (Bailey & Gatrell, 1995; Chaix, Merlo, Subramanian, et al., 2005); a significant Moran's <i>I</i> will indicate that this independence assumption has been violated and support H4.
H4: PSR estimated from individual-level HLM residuals	Because HLM assumes that the individual-level residuals are independent, violation of this assumption is an indication of poor model fit. The PSR was used to quantify the spatial autocorrelation remaining in those residuals. A BCI for this PSR that does not contain PSR = 0 will indicate that this independence assumption has been violated and support H4.

⁹ Bayesian credible intervals are analogous to the confidence intervals associated with the traditional Frequentist approach to statistics (Gill, 2008).

Table 4 (cont'd)

Criterion	Comments
H5: ICC estimated from individual-level GSM residuals	Because GSM assumes that the individual-level residuals are independent, violation of this assumption is an indication of poor model fit. The ICC was used to quantify the hierarchical autocorrelation remaining in those residuals. A BCI for this ICC that contains ICC = 0 will indicate that this independence assumption has not been violated and support H5.
H6: ICC estimated from neighborhood-level GSM residuals	Because the spatial autocorrelation in GSM is captured in the neighborhood-level residuals, some of that spatial autocorrelation might be detected by fitting an HLM using those residuals as the outcome. The ICC will be used to quantify the hierarchical autocorrelation remaining in those residuals. A BCI for this ICC that does not contain ICC = 0 will provide support for H6, as will finding that the upper bound for the BCI on this ICC is lower than the PSR for the GSM from which the residuals were extracted.
H7: Amount of residual neighborhood-level variance	The size of the residual neighborhood-level variance components can be used to compare HLM and GSM models containing both individual and contextual predictors (Chaix, Merlo, & Chauvin, 2005). Support for H7 would consist of evidence that GSM produces models with smaller residual neighborhood-level variance components than HLM, after all predictors have been added to the models.
H7: Overall R^2 and R^2 at each level of analysis	This study used Roberts and Monaco's (Roberts & Monaco, 2006) Equation 15 to calculate overall R^2 statistics that measure the proportion of variance explained by the HLM and GSM models. These R^2 statistics are useful supplements to the DIC for assessing model fit and comparing models (Kramer, 2005). Support for H7 would consist of larger overall R^2 values for GSM than for HLM models. Separate R^2 values can be calculated at each level of analysis in an HLM (Gelman & Pardoe, 2006; Merlo, 2003; Merlo, Chaix, et al., 2005a); this can also be done with GSM. The ability to explain more neighborhood-level variability may be an important indicator of which model performs better. If the quantity $(\text{GSM } R^2 - \text{HLM } R^2) > 0.05$, then it should be reasonable to conclude that GSM has noticeably better predictive capability than HLM, providing support for H7. This cutoff was applied to both the overall and the level-specific R^2 measures.
H7: Raw coefficients for contextual predictors	Because the units of measurement for crime and NSES are the same across methods, the magnitude and sign of the raw crime and NSES coefficients from the HLM and GSM models were directly compared. Observing that GSM analyses produce larger coefficients than the HLM analyses would support H7. The strength of the evidence for or against H7 using this criterion can also be quantified by the percent overlap in the BCIs for the coefficients: the smaller the overlap, the stronger the evidence for a difference.
H7: Differences in the t-statistics for the contextual predictors	The t -statistics for the contextual predictors can be used to quantify and compare the statistical significance of the contextual predictors in both HLM and GSM analyses. Previous research found that HLM consistently produced t -statistics 1 to 2 points larger than those associated with GSM models that used buffers to measure the contextual predictors (Chaix, Merlo, & Chauvin, 2005). T -statistics of around ± 2 are typically significant at the conventional $\alpha = .05$ level, so a 1 point difference in a t -statistic translates into relatively large differences in the significance level. A difference of 1 point between the HLM and GSM t -ratios was considered a substantial difference in the present study.

Table 4 (cont'd)

Criterion	Comments
H7: Source of the differences between HLM and GSM estimates of contextual effects	A three-way comparison between HLM, cluster-based GSM, and buffer-based GSM models should reveal whether differences between the contextual effects are primarily driven by differences in how autocorrelation is modeled or by that plus changing how neighborhood boundaries are defined for measurement purposes. Support for the latter possibility would be a more important substantive contribution to the literature on neighborhood effects than support for the former.
H7: Predictions about how much change in a contextual predictor would be required to improve outcomes	To translate the statistical relationships between contextual predictors (crime and NSES) and residents' perceptions from the HLM and GSM models into useful information, the study calculated how much change in each contextual predictor would be required to achieve a 0.5 standard deviation change in outcomes. This clarified how much would an intervention need to reduce crime or increase NSES in order to produce a decrease of that magnitude in residents' mean level of perceived problems. The aim here was to discover whether the two models made substantially different predictions about how much change would be required. Such information might be useful to change agents who want to implement interventions.
H1 to H8: Differences in the pattern of scientific inferences about the phenomena under study	HLM and GSM analyses may yield different patterns of scientific inferences about the phenomena under study. For example, the two methods might lead to different conclusions how much neighborhoods matter in shaping residents' perceptions of neighborhood problems or about which contextual predictors are significantly related to those perceptions. To the extent that using GSM leads to a richer and more nuanced understanding of the relationships between crime and NSES and residents' perceptions of neighborhood problems than HLM, it may offer valuable new substantive knowledge to neighborhood researchers.

Software. The analyses were conducted using R version 2.9.2 (Ihaka & Gentleman, 1996; R Development Core Team, 2009) and WinBUGS version 1.4.3 (Lunn, et al., 2000; Spiegelhalter, et al., 2007), which are both free, open-source statistical computing software packages. The mix package for R (Schafer, 2009) was used to impute missing data using procedures described by Schafer (1997). The spBayes package for R (Finley, et al., 2007; Finley, Banerjee, & Carlin, 2009) was used to run the GSM analyses, while the R2WinBUGS package for R (Sturtz, Ligges, & Gelman, 2005) was used to export data to WinBUGS, run the HLM analyses, and import the results back into R. Summaries of the Bayesian posterior distributions for HLM and GSM model parameters were computed with the coda package for R (Plummer, Best, Cowles, & Vines, 2006, 2009), as were various diagnostics used to assess model convergence.

RESULTS

This section begins with descriptive statistics and exploratory analyses that provide insight into the distributions of the outcome and the neighborhood-level predictors. After that, it summarizes the spatial relationships between clusters and between survey participants' locations before moving on to present the HLM and GSM models. Where possible, the results are presented graphically to enhance clarity, highlight the most important comparisons and patterns in the findings (Kastellac & Leoni, 2007), and depict the degree of overlap (or lack thereof) between the confidence intervals around the point estimates (Cumming, 2009; Cumming & Finch, 2005).

Exploratory and Descriptive Analyses

Participants' imputed scores on neighborhood problems spanned the full range of values available on the scale, from 1 to 6 ($M = 3.77$, $SD = 1.48$). They were not normally distributed: instead they were mildly negatively skewed and rather platykurtic (skewness = -0.14, kurtosis = -1.08, see Figure 5). However, what that summary conceals are the varying shapes of the distributions within subsets of the data representing smaller geographic areas. The boxplots in Figure 6 show that while the scores in many clusters do still span the entire range of possible values, there are some clusters that have much narrower ranges of values (e.g., clusters 8, 26, and 40). More importantly, examining the central parts of the clusters' distributions (the boxes enclose the middle 50% of the scores for each cluster) in Figure 6 reveals that some clusters mostly contain residents who reported low levels of neighborhood problems (clusters on the left), others mostly contain residents who reported moderate levels of problems (clusters in the center), and yet others contain mostly residents who reported high levels of problems (clusters on the right).

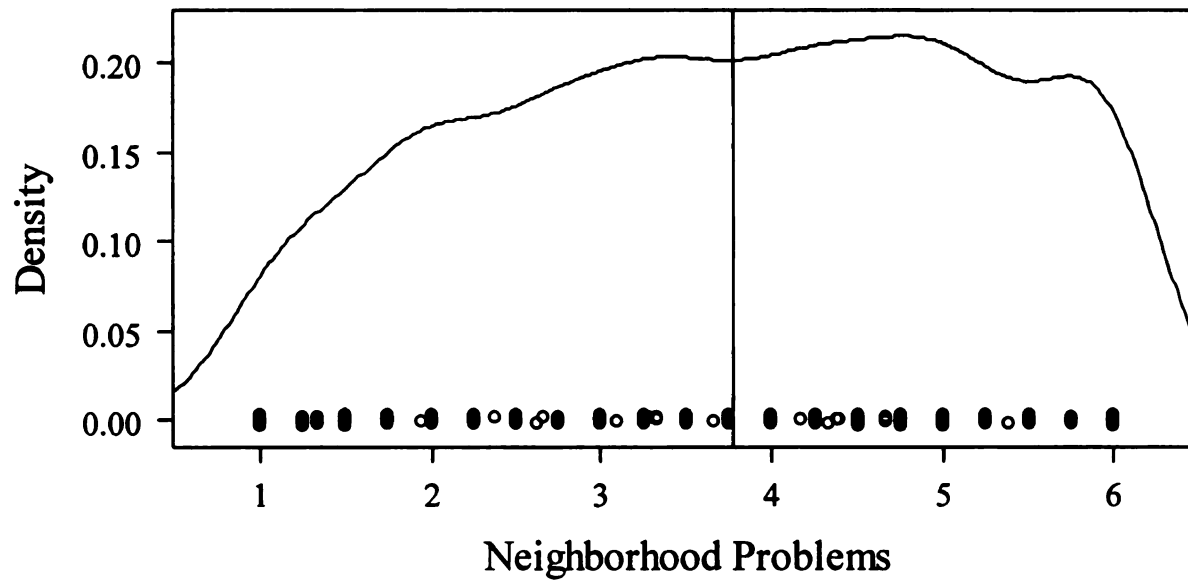


Figure 5: Distribution of the imputed neighborhood problems scores ($N = 1049$). The vertical line is located at the mean ($M = 3.77$, $SD = 1.48$); the circles at the bottom show individual data points (vertically jittered to reduce overlap).

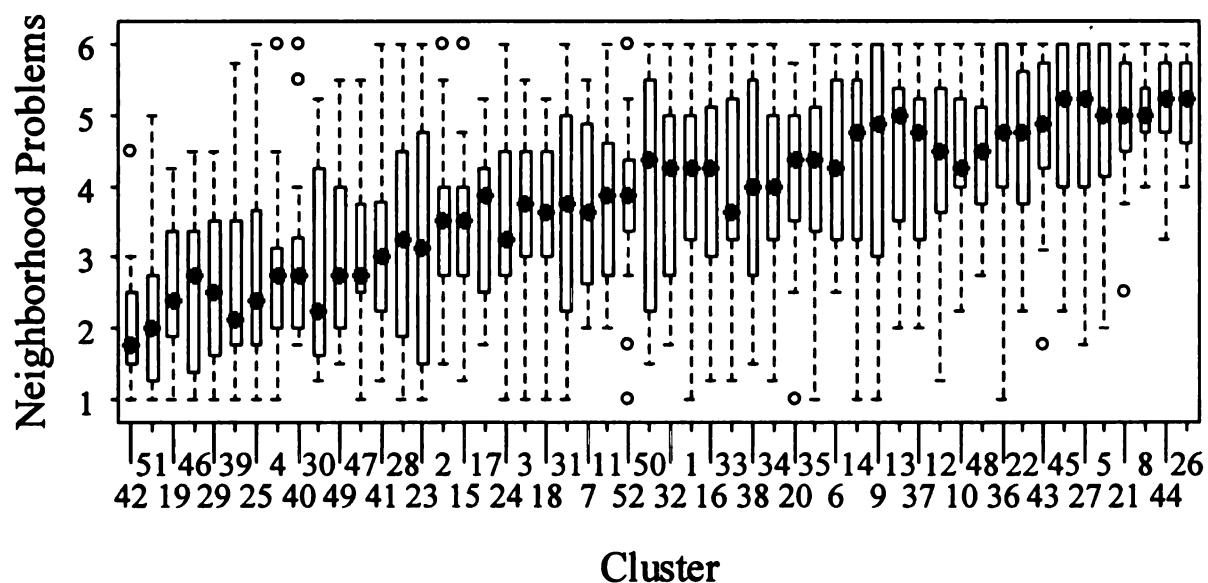


Figure 6: Boxplots of neighborhood problems scores for each cluster, in ascending order by cluster-level mean score. The dots show the clusters' medians.

Of course, the sort of between-neighborhood variability in outcomes illustrated in Figure 6 is what often motivates researchers to use HLM in neighborhood research. Constructing an exact analogue to Figure 6 that highlights spatial rather than hierarchical variation in the distribution of the outcome is difficult. Instead, GSM methods typically

approach exploratory data visualization with maps like the one in Figure 7, which uses color-coded dots to represent the survey responses. Darker dots indicate higher levels of perceived problems, so groups of many dark dots located close together suggest areas where perceived neighborhood problems tend to be high; groups of lighter dots indicate areas where they tend to be low.

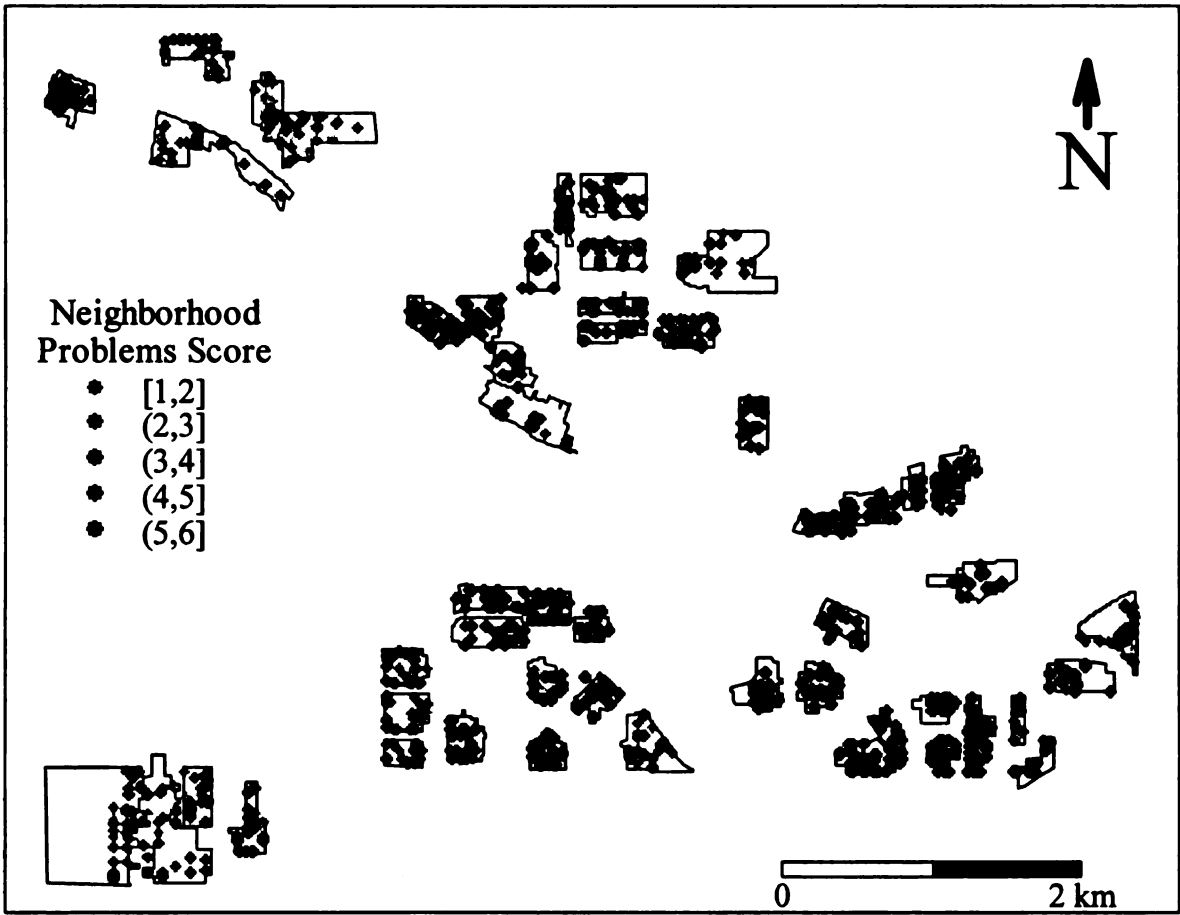


Figure 7: Map of the imputed neighborhood problems scores (N = 1,049).

Descriptive statistics on the individual-level predictors were presented above in Table 2, so descriptive statistics on the neighborhood-level predictors are presented next. For brevity, statistics for buffer-based measures are shown only for the optimally sized buffers (1.1 km radius for crime, 0.2 km radius for NSES, see below for the evidence supporting the selection of these optimal buffer sizes).

When measured within neighborhood cluster boundaries, crime density varied tremendously, from 0.00 to 654.46 crimes/km² ($M = 148.85$, $SD = 147.51$, $Mdn = 113.08$, $N = 52$). Due to the smoothing associated with aggregating crime data within large buffers (the optimal radius was 1.1 km, see below), crime density was less variable when measured within buffers instead of clusters: It ranged from 6.84 to 106.90 crimes/km² ($M = 57.48$, $SD = 24.41$, $Mdn = 60.24$, $N = 1,049$).

NSES, as measured by median housing value within cluster boundaries, ranged from \$22,380 to \$106,100 ($M = \$50,050$; $SD = \$18,063$; $Mdn = \$43,240$; $N = 52$). Unlike crime, NSES was ultimately measured within rather small buffers (0.2 km radius, see below) that were quite comparable in size to the clusters. As a result, buffer-based descriptive statistics for NSES were more similar to the values observed when measuring within clusters: Within the optimal buffers, it ranged from \$22,200 to \$136,000 ($M = \$50,230$; $SD = \$17,382$; $Mdn = \$43,860$; $N = 1,049$).

Figure 8 shows the distributions of the pairwise distances between the centroids of different clusters and between the locations of different survey participants. Due to the size of the study area and the spatial arrangement of the clusters, these distances were quite large in some cases, reaching up to 7.85 km for clusters and 8.17 km for survey participants. The minimum distance between cluster centroids was 0.18 km, while the minimum distance between survey participants' locations (after eliminating exact overlap, i.e., distance = 0) was a mere 0.55 m.

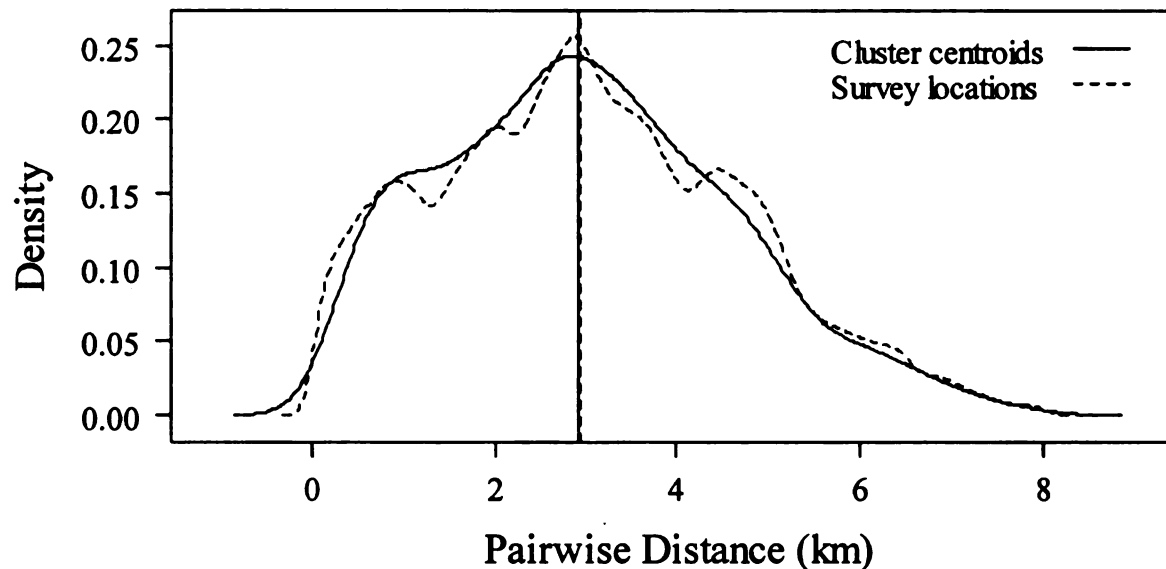


Figure 8: Distributions of pairwise distances between cluster centroids ($N = 52$) and between survey locations ($N = 1,049$). The vertical lines mark the medians, which are almost identical ($Mdn = 2.91$ and $Mdn = 2.95$, respectively).

HLM and GSM Analyses

Most of the hypotheses were tested by comparing estimates of key parameters from alternative models to one another, or by comparing the overall fit of those models. Accordingly, parameter estimates and model fit statistics for HLM models 1-6 are shown in Table 5; corresponding results for GSM Models 1-5 and 56-58 are shown in Table 6. The sections below are organized to present and interpret the results pertinent to the research questions and hypotheses that could be addressed at different stages of the modeling process as the models increased in complexity from empty models to full models containing all of the predictors (refer back to Table 3). Each section extracts relevant information from Tables 5 and 6, interprets the evidence and supplements the tabular presentation with statistical graphics that more directly and intuitively convey the essential patterns in the findings.

Research Question 1. The first research question was: how do GSM estimates of neighborhood-level variance and autocorrelation compare to HLM estimates? Two

hypotheses were relevant to this research question: H1, which predicted that neighborhood-level variance and autocorrelation would be higher in GSM than in HLM, both before and after controlling for individual-level predictors; and H2, which predicted that the range of spatial autocorrelation in the GSM model would be long enough to reach across the borders between clusters.

Hypothesis 1. Figure 9 shows the posterior means and 95% credible intervals for the variance components and the measures of autocorrelation (ICC and PSR) from HLM and GSM Models 1 and 2. In absolute terms, GSM Model 1 attributed more variance to the neighborhood-level ($\sigma^2 = 0.622$) and less to the individual-level ($\tau^2 = 1.513$) than HLM Model 1 ($\tau_{00} = 0.601$, $\sigma^2 = 1.567$), so autocorrelation was slightly higher in the empty GSM model than in the empty HLM model (PSR = .288 vs. ICC = .275). The same pattern was evident with GSM Model 2 ($\sigma^2 = 0.605$, $\tau^2 = 1.505$, PSR = .283) and HLM Model 2 ($\tau_{00} = 0.566$, $\sigma^2 = 1.560$, ICC = .264). However, as Figure 9 illustrates, these were small differences in the posterior means and the GSM credible intervals around these statistics for the neighborhood-level variances and levels of autocorrelation fully enclosed the corresponding HLM credible intervals (i.e., there was 100% overlap) for the empty models.

It is also clear from Figure 9 that controlling for individual-level predictors did not substantially change the neighborhood-level variances or levels of autocorrelation. Neither HLM nor GSM detected much change in those values when one compares the Model 2 results for each method back to the values observed in the empty models. There was still more than 97% overlap in the HLM and GSM credible intervals for the level of

Table 5: Parameter estimates and model fit statistics for HLM Models 1-6.

Parameter	Model 1			Model 2		
	P. Mean	95% CI	t	P. Mean	95% CI	t
L2 fixed effects						
Intercept	3.802	[3.577, 4.024]	33.26	3.804	[3.583, 4.022]	34.07
Crime (cluster)						
NSES (cluster)						
L1 fixed effects						
Age(years)						
36-55				0.131	[-0.061, 0.326]	1.33
≥ 56				0.117	[-0.144, 0.379]	0.88
Female				0.113	[-0.069, 0.296]	1.22
Race						
Black				-0.251	[-0.466, -0.033]	-2.26
Hispanic				-0.033	[-0.428, 0.362]	-0.16
Other				-0.108	[-0.687, 0.474]	-0.36
Marital status						
Married				0.011	[-0.206, 0.230]	0.10
Divorced				0.024	[-0.225, 0.271]	0.19
Widowed				-0.160	[-0.497, 0.178]	-0.93
Education						
< High school				-0.009	[-0.228, 0.211]	-0.08
Undergraduate				0.107	[-0.110, 0.327]	0.96
Postgraduate				0.288	[-0.274, 0.852]	1.00
Employed				-0.053	[-0.237, 0.130]	-0.57
Income (\$1,000s)						
< 15				0.222	[-0.078, 0.529]	1.43
15-25				0.014	[-0.273, 0.304]	0.09
25-45				0.002	[-0.266, 0.271]	0.01
Home owner				-0.188	[-0.386, 0.011]	-1.86
Children present				0.162	[-0.032, 0.355]	1.64
Random effects						
L2 intercept	0.601	[0.390, 0.909]	.000	0.566	[0.361, 0.866]	.059
L2 CAR						
L1 residuals	1.567	[1.436, 1.709]	.000	1.560	[1.428, 1.704]	.004
ICC	.275	[.197, .370]		.264	[.186, .359]	
Model fit index						
	DIC	Deviance	R^2	DIC	Deviance	R^2
Statistic	3496.20	3449.52	.000	3509.20	3445.00	.029
pD	46.70			64.20		

Note: Estimates obtained with Bayesian Markov chain Monte Carlo estimation via Gibbs sampling. CAR = conditional autoregressive spatial random effect on neighborhood-level intercept; 95% CI = central 95% credible interval; DIC = deviance information criterion; ICC = intraclass correlation; L1 = level 1 (individual); L2 = level 2 (neighborhood); P. Mean = posterior mean; PCV = proportional change in variance from Model 1 (level-specific); pD = effective number of parameters; R^2 = overall proportion of variance explained; Spatial lag = Effect of predicted weighted average level of neighborhood problems in surrounding neighborhoods.

Table 5 (cont'd)

Parameter	P. Mean	Model 3			Model 4		
		95% CI	t		95% CI	t	
L2 fixed effects							
Intercept	3.805	[3.613, 3.998]	38.85	3.807	[3.627, 3.986]	41.97	
Crime (cluster)	0.028	[0.014, 0.041]	4.11				
NSES (cluster)				-0.028	[-0.038, -0.018]	-5.38	
L1 fixed effects							
Age(years)							
36-55	0.135	[-0.059, 0.325]	-1.66	0.128	[-0.063, 0.320]	1.30	
≥ 56	0.126	[-0.135, 0.384]	-0.98	0.120	[-0.140, 0.377]	0.90	
Female	0.113	[-0.072, 0.294]	-1.30	0.121	[-0.062, 0.305]	1.30	
Race							
Black	-0.239	[-0.454, -0.024]	-0.24	-0.264	[-0.479, -0.050]	-2.41	
Hispanic	-0.033	[-0.430, 0.359]	-0.47	-0.060	[-0.456, 0.335]	-0.30	
Other	-0.133	[-0.716, 0.454]	-0.42	-0.118	[-0.695, 0.461]	-0.40	
Marital status							
Married	0.010	[-0.209, 0.228]	-0.53	0.020	[-0.199, 0.237]	0.18	
Divorced	0.018	[-0.231, 0.265]	-0.55	0.025	[-0.219, 0.274]	0.20	
Widowed	-0.174	[-0.510, 0.165]	-0.34	-0.156	[-0.494, 0.180]	-0.91	
Education							
< High school	-0.014	[-0.232, 0.203]	-0.48	-0.027	[-0.246, 0.193]	-0.24	
Undergraduate	0.104	[-0.115, 0.322]	-0.97	0.115	[-0.104, 0.334]	1.03	
Postgraduate	0.291	[-0.270, 0.854]	-1.07	0.278	[-0.289, 0.838]	0.97	
Employed	-0.052	[-0.234, 0.130]	-0.40	-0.056	[-0.239, 0.128]	-0.60	
Income (\$1,000s)							
< 15	0.216	[-0.086, 0.517]	-1.79	0.176	[-0.125, 0.476]	1.15	
15-25	0.008	[-0.279, 0.295]	-0.52	-0.017	[-0.301, 0.271]	-0.11	
25-45	-0.011	[-0.277, 0.254]	-0.49	-0.029	[-0.295, 0.237]	-0.22	
Home owner	-0.178	[-0.375, 0.022]	-0.27	-0.176	[-0.373, 0.024]	-1.74	
Children present	0.153	[-0.042, 0.348]		0.154	[-0.044, 0.347]	1.55	
Random effects							
	P. Mean	95% CI	PCV	P. Mean	95% CI	PCV	
L2 intercept	0.414	[0.256, 0.646]	.311	0.345	[0.213, 0.538]	.425	
L2 CAR							
L1 residuals	1.560	[1.429, 1.704]	.004	1.559	[1.427, 1.703]	.005	
ICC	.208	[0.139, 0.295]		0.180	[0.119, 0.258]		
Model fit index							
	DIC	Deviance	R ²	DIC	Deviance	R ²	
Statistic	3,508.30	3,445.86	.132	3,505.70	3,444.48	.177	
pD	62.50			61.20			

Note: Estimates obtained with Bayesian Markov chain Monte Carlo estimation via Gibbs sampling. CAR = conditional autoregressive spatial random effect on neighborhood-level intercept; 95% CI = central 95% credible interval; DIC = deviance information criterion; ICC = intraclass correlation; L1 = level 1 (individual); L2 = level 2 (neighborhood); P. Mean = posterior mean; PCV = proportional change in variance from Model 1 (level-specific); pD = effective number of parameters; R^2 = overall proportion of variance explained; Spatial lag = Effect of predicted weighted average level of neighborhood problems in surrounding neighborhoods.

Table 5 (cont'd)

Parameter	P. Mean	Model 5			Model 6 (CAR HLM)		
		95% CI	t	P. Mean	95% CI	t	
L2 fixed effects							
Intercept	3.807	[3.638, 3.975]	44.60	3.811	[3.657, 3.966]	48.81	
Crime (cluster)	0.018	[0.005, 0.030]	2.80	0.014	[0.001, 0.026]	2.08	
NSES (cluster)	-0.022	[-0.032, -0.012]	-4.34	-0.018	[-0.030, -0.007]	-3.16	
L1 fixed effects							
Age(years)							
36-55	0.130	[-0.060, 0.322]	1.34	0.131	[-0.061, 0.322]	1.34	
≥ 56	0.127	[-0.132, 0.387]	0.96	0.130	[-0.127, 0.390]	0.99	
Female	0.120	[-0.063, 0.303]	1.28	0.118	[-0.064, 0.302]	1.26	
Race							
Black	-0.254	[-0.467, -0.042]	-2.34	-0.262	[-0.475, -0.048]	-2.39	
Hispanic	-0.056	[-0.450, 0.338]	-0.28	-0.058	[-0.454, 0.338]	-0.29	
Other	-0.137	[-0.713, 0.444]	-0.46	-0.141	[-0.721, 0.438]	-0.48	
Marital status							
Married	0.018	[-0.201, 0.236]	0.16	0.018	[-0.199, 0.236]	0.16	
Divorced	0.022	[-0.226, 0.267]	0.17	0.024	[-0.225, 0.272]	0.19	
Widowed	-0.167	[-0.505, 0.172]	-0.97	-0.162	[-0.500, 0.178]	-0.94	
Education							
< High school	-0.028	[-0.245, 0.192]	-0.25	-0.031	[-0.247, 0.188]	-0.28	
Undergraduate	0.114	[-0.107, 0.331]	1.02	0.112	[-0.106, 0.331]	1.01	
Postgraduate	0.285	[-0.275, 0.855]	0.99	0.288	[-0.275, 0.850]	1.00	
Employed	-0.054	[-0.237, 0.128]	-0.58	-0.057	[-0.240, 0.124]	-0.61	
Income (\$1,000s)							
< 15	0.175	[-0.130, 0.475]	1.13	0.169	[-0.133, 0.470]	1.10	
15-25	-0.021	[-0.310, 0.267]	-0.14	-0.021	[-0.310, 0.266]	-0.14	
25-45	-0.036	[-0.305, 0.229]	-0.27	-0.038	[-0.303, 0.223]	-0.28	
Home owner	-0.168	[-0.364, 0.028]	-1.68	-0.171	[-0.368, 0.024]	-1.70	
Children present	0.149	[-0.043, 0.341]	1.52	0.141	[-0.053, 0.333]	1.43	
Random effects	P. Mean	95% CI	PCV	P. Mean	95% CI	PCV	
L2 intercept	0.293	[0.176, 0.464]	.512	0.233	[0.129, 0.390]	.610	
L2 CAR				0.001	[0.000, 0.002]	.000	
L1 residuals	1.560	[1.425, 1.703]	.004	1.558	[1.426, 1.702]	.005	
ICC	0.157	[0.100, 0.232]		0.129	[0.076, 0.202]		
Model fit index	DIC	Deviance	R^2	DIC	Deviance	R^2	
Statistic	3,505.20	3,445.20	.207	3503.70	3443.72	.239	
pD	60.00			60.00			

Note: Estimates obtained with Bayesian Markov chain Monte Carlo estimation via Gibbs sampling. CAR = conditional autoregressive spatial random effect on neighborhood-level intercept; 95% CI = central 95% credible interval; DIC = deviance information criterion; ICC = intraclass correlation; L1 = level 1 (individual); L2 = level 2 (neighborhood); P. Mean = posterior mean; PCV = proportional change in variance from Model 1 (level-specific); pD = effective number of parameters; R^2 = overall proportion of variance explained.

Table 6: Parameter estimates and model fit statistics for GSM Models 1-5 and 56-58.

Parameter	Model 1			Model 2		
	P. Mean	95% CI	t	P. Mean	95% CI	t
L2 fixed effects						
Intercept	3.381	[2.765, 3.927]	11.53	3.390	[2.767, 3.949]	11.39
Crime (cluster)						
NSES (cluster)						
Crime (1.1 km)						
NSES (0.2 km)						
L1 fixed effects						
Age(years)						
36-55				0.148	[-0.043, 0.338]	1.53
≥ 56				0.163	[-0.096, 0.418]	1.24
Female				0.127	[-0.056, 0.310]	1.36
Race						
Black				-0.270	[-0.492, -0.047]	-2.39
Hispanic				-0.039	[-0.432, 0.355]	-0.19
Other				-0.107	[-0.683, 0.470]	-0.36
Marital status						
Married				0.034	[-0.184, 0.253]	0.31
Divorced				0.017	[-0.228, 0.262]	0.14
Widowed				-0.161	[-0.498, 0.177]	-0.94
Education						
< High school				-0.030	[-0.248, 0.190]	-0.27
Undergraduate				0.108	[-0.110, 0.328]	0.97
Postgraduate				0.309	[-0.260, 0.874]	1.07
Employed				-0.076	[-0.258, 0.104]	-0.83
Income (\$1,000s)						
< 15				0.168	[-0.133, 0.468]	1.10
15-25				-0.002	[-0.287, 0.284]	-0.01
25-45				-0.015	[-0.280, 0.248]	-0.11
Home owner				-0.232	[-0.428, -0.036]	-2.32
Children present				0.163	[-0.030, 0.355]	1.66
Random effects						
L2 intercept	P. Mean	95% CI	PCV	P. Mean	95% CI	PCV
L2 intercept	0.622	[0.371, 0.998]	.000	0.605	[0.358, 0.975]	0.027
L1 residuals	1.513	[1.378, 1.661]	.000	1.505	[1.371, 1.649]	0.006
PSR	0.288	[0.194, 0.404]		0.283	[0.190, 0.399]	
Spatial parameter						
Phi (φ) x 1000	P. Mean	95% CI		P. Mean	95% CI	
Phi (φ) x 1000	1.011	[0.675, 1.748]		0.980	[0.659, 1.740]	
Range (km)	2.962	[1.714, 4.440]		3.058	[1.722, 4.544]	
Model fit index						
	DIC	Deviance	R^2	DIC	Deviance	R^2
Statistic	1,560.66	1,485.19	.000	1,569.66	1,479.32	.033
pD	75.47			90.35		

Note: Estimates obtained with Bayesian Markov chain Monte Carlo estimation via Gibbs sampling. 95% CI = central 95% credible interval; DIC = deviance information criterion; L1 = level 1 (individual); L2 = level 2 (neighborhood); P. Mean = posterior mean; PCV = proportional change in variance from Model 1 (level-specific); pD = effective number of parameters; Phi = rate of decrease in autocorrelation (multiplied by 1,000); PSR = partial sill ratio; R^2 = overall proportion of variance explained; Range = practical range of variogram.

Table 6 (cont'd)

Parameter	Model 3			Model 4		
	P. Mean	95% CI	t	P. Mean	95% CI	t
L2 fixed effects						
Intercept	3.469	[2.918, 3.940]	13.52	3.490	[2.932, 3.956]	13.55
Crime (cluster)	0.013	[0.004, 0.023]	2.68			
NSES (cluster)				-0.008	[-0.019, 0.004]	-1.33
Crime (1.1 km)						
NSES (0.2 km)						
L1 fixed effects						
Age(years)						
36-55	0.152	[-0.038, 0.342]	1.57	0.145	[-0.048, 0.335]	1.48
≥ 56	0.175	[-0.082, 0.432]	1.34	0.160	[-0.096, 0.416]	1.23
Female	0.126	[-0.057, 0.309]	1.35	0.130	[-0.054, 0.312]	1.38
Race						
Black	-0.271	[-0.492, -0.049]	-2.39	-0.260	[-0.483, -0.036]	-2.29
Hispanic	-0.051	[-0.444, 0.345]	-0.25	-0.037	[-0.432, 0.358]	-0.18
Other	-0.135	[-0.708, 0.442]	-0.46	-0.106	[-0.683, 0.472]	-0.36
Marital status						
Married	0.029	[-0.189, 0.245]	0.26	0.036	[-0.183, 0.253]	0.32
Divorced	0.008	[-0.238, 0.255]	0.07	0.018	[-0.229, 0.265]	0.14
Widowed	-0.181	[-0.516, 0.151]	-1.06	-0.160	[-0.496, 0.176]	-0.93
Education						
< High school	-0.028	[-0.247, 0.190]	-0.25	-0.031	[-0.250, 0.186]	-0.28
Undergraduate	0.106	[-0.112, 0.325]	0.95	0.114	[-0.104, 0.333]	1.02
Postgraduate	0.301	[-0.262, 0.865]	1.05	0.305	[-0.256, 0.871]	1.06
Employed	-0.068	[-0.249, 0.112]	-0.74	-0.078	[-0.261, 0.103]	-0.85
Income (\$1,000s)						
< 15	0.162	[-0.139, 0.463]	1.06	0.153	[-0.145, 0.453]	1.00
15-25	-0.019	[-0.303, 0.267]	-0.13	-0.011	[-0.296, 0.276]	-0.07
25-45	-0.031	[-0.296, 0.235]	-0.23	-0.025	[-0.288, 0.241]	-0.18
Home owner	-0.222	[-0.418, -0.027]	-2.22	-0.227	[-0.425, -0.030]	-2.25
Children present	0.161	[-0.029, 0.353]	1.65	0.163	[-0.027, 0.355]	1.67
Random effects						
L2 intercept	P. Mean	95% CI	PCV	P. Mean	95% CI	PCV
L2 intercept	0.500	[0.288, 0.824]	0.196	0.505	[0.285, 0.846]	0.188
L1 residuals	1.507	[1.373, 1.653]	0.004	1.511	[1.373, 1.660]	0.002
PSR	0.246	[0.157, 0.357]		0.247	[0.155, 0.364]	
Spatial parameter						
Phi (ϕ) x 1000	P. Mean	95% CI		P. Mean	95% CI	
Phi (ϕ) x 1000	1.132	[0.660, 2.098]		1.222	[0.663, 2.499]	
Range (km)	2.647	[1.428, 4.537]		2.451	[1.199, 4.519]	
Model fit index						
	DIC	Deviance	R^2	DIC	Deviance	R^2
Statistic	1,569.22	1,481.05	0.062	1,573.47	1,482.75	0.048
pD	88.17			90.72		

Note: Estimates obtained with Bayesian Markov chain Monte Carlo estimation via Gibbs sampling. 95% CI = central 95% credible interval; DIC = deviance information criterion; L1 = level 1 (individual); L2 = level 2 (neighborhood); P. Mean = posterior mean; PCV = proportional change in variance from Model 1 (level-specific); pD = effective number of parameters; Phi = rate of decrease in autocorrelation

(multiplied by 1,000); PSR = partial sill ratio; R^2 = overall proportion of variance explained; Range = practical range of variogram.

Table 6 (cont'd)

Parameter	Model 5			Model 56		
	P. Mean	95% CI	t	P. Mean	95% CI	t
L2 fixed effects						
Intercept	3.554	[3.054, 3.941]	15.91	3.753	[3.513, 3.963]	32.91
Crime (cluster)	0.013	[0.003, 0.023]	2.60			
NSES (cluster)	-0.007	[-0.018, 0.005]	-1.18			
Crime (1.1 km)				0.263	[0.186, 0.338]	6.88
NSES (0.2 km)						
L1 fixed effects						
Age(years)						
36-55	0.149	[-0.042, 0.339]	1.53	0.144	[-0.046, 0.334]	1.48
≥ 56	0.172	[-0.084, 0.430]	1.31	0.175	[-0.083, 0.433]	1.34
Female	0.129	[-0.054, 0.313]	1.38	0.122	[-0.062, 0.305]	1.31
Race						
Black	-0.261	[-0.483, -0.040]	-2.31	-0.254	[-0.471, -0.037]	-2.31
Hispanic	-0.050	[-0.448, 0.347]	-0.25	-0.041	[-0.433, 0.354]	-0.20
Other	-0.135	[-0.717, 0.442]	-0.46	-0.078	[-0.649, 0.492]	-0.26
Marital status						
Married	0.033	[-0.186, 0.251]	0.29	0.033	[-0.185, 0.252]	0.30
Divorced	0.010	[-0.236, 0.257]	0.08	0.020	[-0.228, 0.268]	0.16
Widowed	-0.178	[-0.515, 0.161]	-1.03	-0.162	[-0.497, 0.174]	-0.95
Education						
< High school	-0.029	[-0.248, 0.190]	-0.26	-0.014	[-0.232, 0.203]	-0.13
Undergraduate	0.111	[-0.109, 0.328]	0.99	0.115	[-0.103, 0.333]	1.03
Postgraduate	0.301	[-0.268, 0.864]	1.05	0.284	[-0.279, 0.846]	0.99
Employed	-0.069	[-0.251, 0.111]	-0.75	-0.082	[-0.263, 0.099]	-0.89
Income (\$1,000s)						
< 15	0.151	[-0.150, 0.452]	0.99	0.162	[-0.137, 0.457]	1.06
15-25	-0.024	[-0.313, 0.262]	-0.17	-0.008	[-0.294, 0.275]	-0.06
25-45	-0.039	[-0.305, 0.226]	-0.29	-0.009	[-0.275, 0.252]	-0.07
Home owner	-0.216	[-0.412, -0.019]	-2.14	-0.234	[-0.429, -0.040]	-2.36
Children present	0.160	[-0.033, 0.352]	1.63	0.162	[-0.030, 0.354]	1.65
Random effects						
L2 intercept	0.426	[0.232, 0.752]	0.315	0.266	[0.153, 0.455]	0.573
L1 residuals	1.510	[1.370, 1.660]	0.002	1.485	[1.344, 1.637]	0.019
PSR	0.217	[0.131, 0.338]		0.151	[0.092, 0.238]	
Spatial parameter						
Phi (ϕ) x 1000	1.525	[0.705, 3.595]		3.914	[0.829, 8.641]	
Range (km)	1.964	[0.833, 4.248]		0.765	[0.347, 3.613]	
Model fit index						
	DIC	Deviance	R^2	DIC	Deviance	R^2
Statistic	1,573.32	1,482.03	0.085	1,567.85	1,465.06	0.265
pD	91.29			102.78		

Note: Estimates obtained with Bayesian Markov chain Monte Carlo estimation via Gibbs sampling. 95% CI = central 95% credible interval; DIC = deviance information criterion; L1 = level 1 (individual); L2 = level 2 (neighborhood); P. Mean = posterior mean; PCV = proportional change in variance from Model 1 (level-specific); pD = effective number of parameters; Phi = rate of decrease in autocorrelation (multiplied by 1,000); PSR = partial sill ratio; R^2 = overall proportion of variance explained; Range = practical range of variogram.

Table 6 (cont'd)

Parameter	P. Mean	Model 57			P. Mean	Model 58		t
		95% CI	t	95% CI				
L2 fixed effects								
Intercept	3.600	[3.134, 3.968]	17.32	3.760	[3.594, 3.919]	45.81		
Crime (cluster)								
NSES (cluster)								
Crime (1.1 km)				0.200	[0.125, 0.275]	5.26		
NSES (0.2 km)	-0.021	[-0.031, -0.010]	-3.90	-0.014	[-0.023, -0.005]	-3.11		
L1 fixed effects								
Age(years)								
36-55	0.149	[-0.042, 0.339]	1.53	0.146	[-0.045, 0.336]	1.50		
≥ 56	0.184	[-0.074, 0.440]	1.41	0.190	[-0.065, 0.446]	1.46		
Female	0.124	[-0.059, 0.305]	1.33	0.124	[-0.057, 0.307]	1.34		
Race								
Black	-0.261	[-0.481, -0.041]	-2.33	-0.251	[-0.461, -0.041]	-2.34		
Hispanic	-0.045	[-0.440, 0.347]	-0.22	-0.055	[-0.451, 0.341]	-0.27		
Other	-0.114	[-0.691, 0.463]	-0.39	-0.084	[-0.661, 0.493]	-0.29		
Marital status								
Married	0.057	[-0.161, 0.274]	0.51	0.055	[-0.161, 0.274]	0.50		
Divorced	0.031	[-0.217, 0.276]	0.24	0.032	[-0.215, 0.280]	0.25		
Widowed	-0.160	[-0.495, 0.176]	-0.93	-0.160	[-0.493, 0.176]	-0.93		
Education								
< High school	-0.037	[-0.254, 0.181]	-0.33	-0.023	[-0.240, 0.194]	-0.21		
Undergraduate	0.127	[-0.093, 0.344]	1.15	0.132	[-0.089, 0.349]	1.18		
Postgraduate	0.299	[-0.263, 0.864]	1.04	0.275	[-0.289, 0.836]	0.96		
Employed	-0.083	[-0.264, 0.098]	-0.89	-0.085	[-0.266, 0.096]	-0.92		
Income (\$1,000s)								
< 15	0.128	[-0.173, 0.429]	0.84	0.126	[-0.173, 0.424]	0.83		
15-25	-0.003	[-0.289, 0.282]	-0.02	-0.019	[-0.304, 0.266]	-0.13		
25-45	-0.033	[-0.296, 0.231]	-0.24	-0.028	[-0.291, 0.236]	-0.21		
Home owner	-0.232	[-0.427, -0.037]	-2.33	-0.236	[-0.431, -0.040]	-2.37		
Children present	0.164	[-0.028, 0.355]	1.67	0.159	[-0.033, 0.350]	1.62		
Random effects	P. Mean	95% CI	PCV	P. Mean	95% CI	PCV		
L2 intercept	0.399	[0.225, 0.697]	0.358	0.235	[0.138, 0.373]	0.622		
L1 residuals	1.500	[1.362, 1.648]	0.009	1.467	[1.312, 1.624]	0.031		
PSR	0.208	[0.128, 0.319]		0.138	[0.083, 0.212]			
Spatial parameter	P. Mean	95% CI		P. Mean	95% CI			
Phi (φ) x 1000	1.628	[0.691, 3.651]		5.929	[1.809, 15.170]			
Range (km)	1.840	[0.820, 4.335]		0.505	[0.197, 1.656]			
Model fit index	DIC	Deviance	R ²	DIC	Deviance	R ²		
Statistic	1,566.95	1,476.10	0.116	1,564.06	1,452.12	0.282		
pD	90.85			111.94				

Note: Estimates obtained with Bayesian Markov chain Monte Carlo estimation via Gibbs sampling.
 95% CI = central 95% credible interval; DIC = deviance information criterion; L1 = level 1 (individual);
 L2 = level 2 (neighborhood); P. Mean = posterior mean; PCV = proportional change in variance from
 Model 1 (level-specific); pD = effective number of parameters; Phi = rate of decrease in autocorrelation
 (multiplied by 1,000 for display); PSR = partial sill ratio; R^2 = overall proportion of variance explained;
 Range = practical range of variogram.

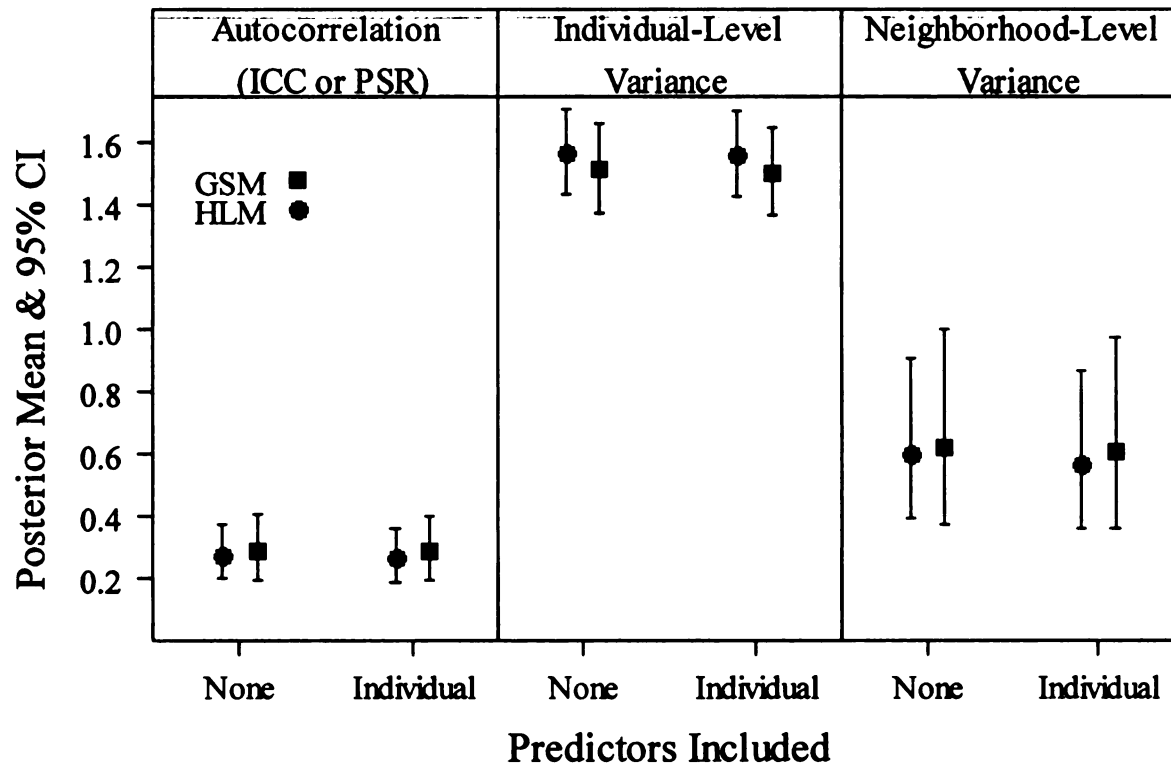


Figure 9: Estimated variance components and levels of autocorrelation from HLM and GSM Models 1 and 2. The plot shows the posterior means (symbols) plus the central 95% credible intervals (whiskers) around those estimates. Model 1 for each method was an empty model that included no predictors, while Model 2 included all individual-level predictors.

autocorrelation after adding the individual-level predictors. Thus, the evidence fails to provide strong support H1 both before and after controlling for the set of resident characteristics considered in this study.

Hypothesis 2. Figure 10 shows the variograms and autocorrelation functions associated with GSM Models 1 and 2. The two sets of curves are very similar (almost exactly overlapping), indicating that controlling for individual-level variables did not explain much of the neighborhood-level variance. Both models have practical ranges of approximately 3 km, indicating the spatial autocorrelation persists over long distances in these data. Indeed, these ranges exceed both of the median pairwise distances (between cluster centroids and between survey locations) shown in Figure 8, which provides strong

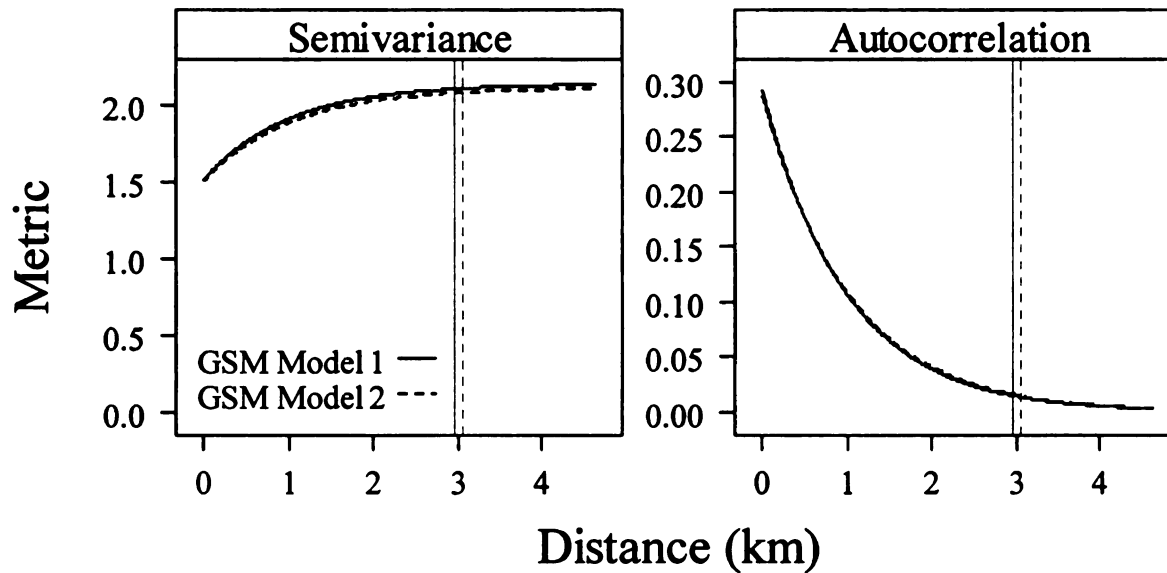


Figure 10: Exponential variograms and correlation functions for GSM Models 1 and 2. The maximum value for the autocorrelation is the PSR for each model (.288 and .283, respectively). Vertical lines mark the practical ranges of spatial autocorrelation for the two models (2,962 m and 3,058 m, respectively), which occur at the distances where the neighborhood-level covariances have decreased to 5% of their initial sizes, leaving little residual autocorrelation.

support for H2 by indicating that spatial autocorrelation can easily reach across the borders between clusters.

Research Question 2. The second research question asked which method is more effective at modeling the autocorrelation actually observed in these data. Several hypotheses were associated with that question. The prediction in H3 was that empty GSM models would fit better than empty HLM models, and that controlling for individual-level predictors would not change that result. Meanwhile, H4-H6 made predictions pertaining to properties of the residuals of the empty models and the models with individual-level predictors.

Hypothesis 3. The deviance and DIC values for GSM Model 1 ($D = 1485.19$, $DIC = 1,560.66$, $pD = 75.47$) were far smaller than those associated with HLM Model 1 ($D = 3449.52$, $DIC = 3496.20$, $pD = 46.70$), indicating that the GSM approach provided a

better fit to the data. The pD statistics show that the GSM model did have more effective parameters than the HLM model, but that difference was small compared to the size of the DIC statistics for these models. Because the empty models are the baseline to which other models are compared to calculate level 2 PCV statistics, those statistics could not be calculated for these models. The overall R^2 values are zero for both empty models.

The results were similar after controlling for individual-level predictors in Model 2 for each method: The deviance and DIC values for the GSM model ($D = 1479.32$, $DIC = 1,569.66$, $pD = 90.35$) remained far smaller than those of the HLM model ($D = 3445.00$, $DIC = 3509.20$, $pD = 64.20$). While HLM Model 2 explained slightly more of the neighborhood-level variance (level 2 PCV = .057) and had fewer effective parameters, it had a slightly smaller overall R^2 (.029) than GSM Model 2 (level 2 PCV = .027, $R^2 = .033$).

Finally, while the deviances decreased for both methods as one might expect when adding predictors, the DICs actually increased from Model 1 to Model 2. The DIC went up by 9 points for GSM and by 13 points for HLM, both of which would indicate that the increase in model fit doesn't necessarily warrant the increase in model complexity. Few of the individual-level predictors were significant and most had coefficients that were very similar between the HLM and GSM models, so there were few noteworthy differences in the inferences about their effects across methods. Home ownership was the exception: It had a significant effect in GSM Model 2 ($\beta = -0.232$, 95% CI = [-0.428, -0.036]) but not in HLM Model 2 ($\beta = -0.188$, 95% CI = [-0.386, 0.011]), but even this was a matter of degree rather than a difference in sign and

substantive meaning. Indeed, there is 89% overlap between these two credible intervals. Controlling for neighborhood composition was more important than making the models more parsimonious, so all individual-level predictors were retained in remaining models.

Hypothesis 4. To test whether level 1 residuals from HLM Models 1 and 2 contained residual spatial autocorrelation, they were saved to new datasets and then used as the dependent variables in a pair of empty GSM models (see Table 7, plus Figures 11-12). These residuals from HLM Model 1 still contain a substantial amount of spatial autocorrelation (PSR = .441, 95% CI = [.126, .840]), as do the level 1 residuals from HLM Model 2 (PSR = .425, 95% CI = [.098, .833]).

Figure 11 shows that the posterior means for the practical range of spatial autocorrelation remaining in the level 1 HLM residuals are about 4.4 m in Model 1 and 4.5 m in Model 2, though the credible intervals (95% CIs = [3.040, 14.326] and [3.016,

Table 7: Parameter estimates and model fit statistics for empty GSM models fit to individual-level residuals from HLM Models 1 and 2.

Parameter	HLM Model 1 L1 Residuals			HLM Model 2 L1 Residuals		
	P. Mean	95% CI	t	P. Mean	95% CI	t
L2 fixed effects Intercept	0.000	[-0.075, 0.074]	-0.01	0.000	[-0.077, 0.075]	-0.01
Random effects	P. Mean	95% CI		P. Mean	95% CI	
L2 intercept	0.664	[0.189, 1.273]		0.625	[0.147, 1.235]	
L1 residuals	0.840	[0.240, 1.338]		0.848	[0.247, 1.372]	
PSR	.441	[.126, .840]		.425	[.098, .833]	
Spatial parameter	P. Mean	95% CI		P. Mean	95% CI	
Phi (ϕ) x 1000	681.518	[209.107, 985.566]		664.777	[154.580, 993.190]	
Range (m)	4.396	[3.040, 14.326]		4.506	[3.016, 19.380]	
Model fit index	DIC	Deviance		DIC	Deviance	
Statistic	1144.65	774.09		1134.65	782.81	
pD	370.56			351.84		

Note: Estimates obtained with Bayesian Markov chain Monte Carlo estimation via Gibbs sampling (3 chains; 32,000 iterations/chain; 7,000 iteration burn-in periods). 95% CI = central 95% credible interval; DIC = deviance information criterion; L1 = level 1 (individual); L2 = level 2 (neighborhood); P. Mean = posterior mean; pD = effective number of parameters; Phi = rate of decrease in autocorrelation (multiplied by 1,000); PSR = partial sill ratio; Range = practical range of variogram.

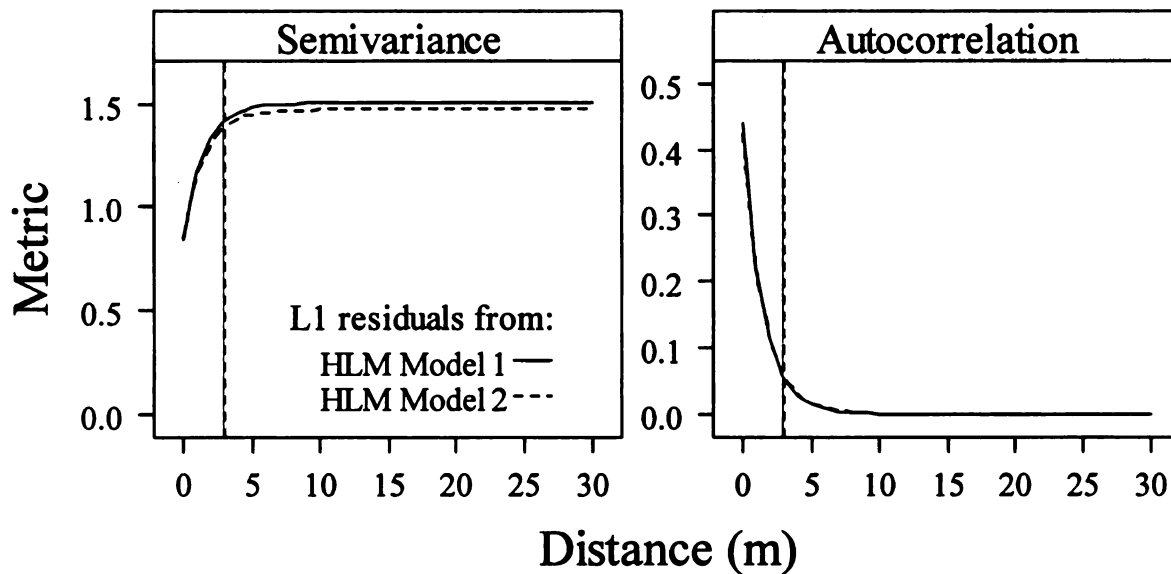


Figure 11: Exponential variograms and correlation functions for empty GSM models fit to the individual-level residuals for HLM Models 1 and 2. The maximum value for the autocorrelation is the PSR for each model (.441 and .425, respectively). Vertical lines mark the practical ranges of spatial autocorrelation for the two models (4.4 m and 4.5 m, respectively), which occur at the distances where the neighborhood-level covariances have decreased to 5% of their initial sizes, leaving little residual autocorrelation.

19.380], respectively) suggest that these ranges might be as high as 14.3 and 19.4 m.

This is a very small spatial scale, but that is not surprising because level 1 residuals represent variability within clusters, which are all quite small geographic areas. Spatial autocorrelation persisting over longer distances would be more likely to show up in the level 2 HLM residuals.

The level 2 residuals from HLM Models 1 and 2 were also tested for residual spatial autocorrelation, though a method better suited to testing autocorrelation in data associated with areal units such as the clusters was required. An exact Moran's I test for regression residuals was applied to test this part of H4. Both Model 1 (Moran's $I = .27$, $z = 5.30$, $p < .001$) and Model 2 residuals (Moran's $I = .26$, $z = 5.18$, $p < .001$) contained strong evidence of spatial autocorrelation that decayed as a function of distance.

Hypothesis 5. To test whether level 1 residuals from GSM Models 1 and 2 contained residual hierarchical autocorrelation, they were saved to new datasets and then used as the dependent variables in a pair of empty HLM models (see Table 8, plus Figure 12). HLM was able to detect only small amounts of hierarchical autocorrelation remaining in the residuals from both GSM Model 1 (ICC = .056, 95% CI = [.036, .084]) and GSM Model 2 (ICC = .057, 95% CI = [.036, .086]).

There was asymmetry in how much remaining autocorrelation the two methods could each detect in the level 1 residuals produced by the other method. While GSM detected substantial amounts of spatial autocorrelation remaining in the HLM residuals, HLM detected far less hierarchical autocorrelation in the GSM residuals (see Figure 12). This suggests that GSM, not HLM, is better at modeling the autocorrelation in these data.

Hypothesis 6. The last hypothesis associated with the second research question (H6) predicted that applying an empty HLM to the neighborhood-level residuals from GSM Models 1 and 2 would detect hierarchical autocorrelation, but that the ICC would

Table 8: Parameter estimates and model fit statistics for empty HLM models fit to individual-level residuals from GSM Models 1 and 2.

Parameter	GSM Model 1 L1 Residuals			GSM Model 2 L1 Residuals		
	P. Mean	95% CI	t	P. Mean	95% CI	t
L2 fixed effects Intercept	-0.003	[-0.110, 0.103]	-0.06	-0.003	[-0.110, 0.103]	-0.06
Random effects	P. Mean	95% CI		P. Mean	95% CI	
L2 intercept	0.084	[0.053, 0.129]		0.083	[0.052, 0.129]	
L1 residuals	1.418	[1.302, 1.547]		1.390	[1.276, 1.515]	
ICC	.056	[.036, .084]		.057	[.036, .086]	
Model fit index	DIC	Deviance		DIC	Deviance	
Statistic	3374.30	3345.41		3352.40	3323.28	
pD	28.90			29.10		

Note: Estimates obtained with Bayesian Markov chain Monte Carlo estimation via Gibbs sampling (3 chains; 16,000 iterations/chain; 1,000 iteration burn-in periods). 95% CI = central 95% credible interval; DIC = deviance information criterion; ICC = intra-class correlation; L1 = level 1 (individual); L2 = level 2 (neighborhood); P. Mean = posterior mean; pD = effective number of parameters.

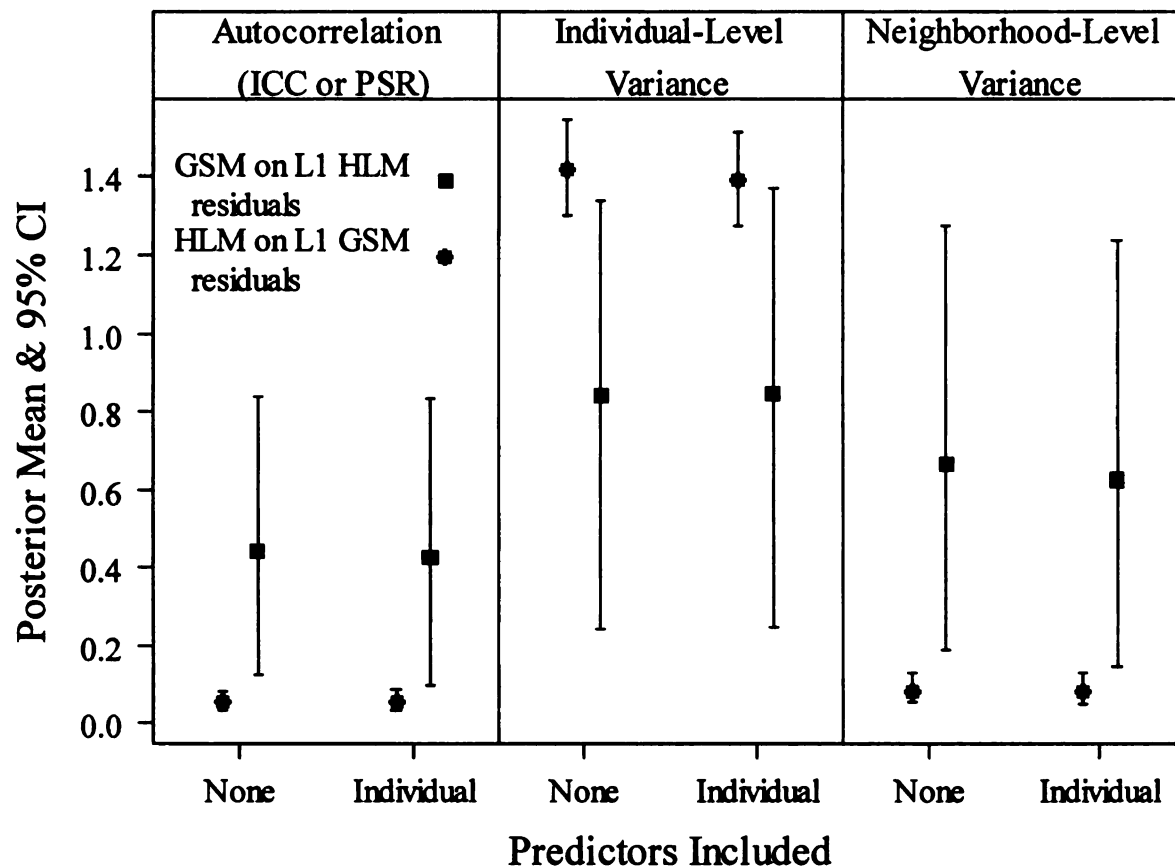


Figure 12: Estimated variance components and levels of autocorrelation from empty HLM and GSM models fit to the individual-level residuals from GSM and HLM Models 1 and 2. The plot shows the posterior means (symbols) plus the central 95% credible intervals (whiskers) around those estimates. Predictors included refers here to the predictors included in the model that generated the residuals being analyzed: Model 1 for each method was an empty model that included no predictors, while Model 2 included all individual-level predictors. ICC = intra-class correlation; L1 = level 1 (individual); PSR = partial sill ratio.

be lower than the PSRs in those original models. Table 9 shows the results of testing H6, which was partly supported—the ICCs were indeed significant—and partly unsupported because the ICCs were far larger than the original PSRs. HLM detected an extremely high level of hierarchical autocorrelation (ICC = .96, 95% CI = [.944, .974]) in the GSM Model 1 neighborhood-level residuals. The result was nearly identical for the GSM Model 2 neighborhood-level residuals (ICC = .96, 95% CI = [.945, .974]).

Table 9: Parameter estimates and model fit statistics for empty HLM models fit to neighborhood-level residuals from GSM Models 1 and 2.

Parameter	GSM Model 1 L1 Residuals			GSM Model 2 L1 Residuals		
	P. Mean	95% CI	t	P. Mean	95% CI	t
L2 fixed effects Intercept	0.436	[0.235, 0.635]	4.31	0.428	[0.235, 0.622]	4.35
Random effects	P. Mean	95% CI		P. Mean	95% CI	
L2 intercept	0.529	[0.361, 0.770]		0.503	[0.345, 0.733]	
L1 residuals	0.021	[0.019, 0.023]		0.020	[0.018, 0.022]	
ICC	.960	[.944, .974]		.960	[.945, .974]	
Model fit index	DIC	Deviance		DIC	Deviance	
Statistic	-1115.60	-1168.55		-1172.10	-1225.09	
pD	52.90			53.00		

Note: Estimates obtained with Bayesian Markov chain Monte Carlo estimation via Gibbs sampling (3 chains; 16,000 iterations/chain; 1,000 iteration burn-in periods). 95% CI = central 95% credible interval; DIC = deviance information criterion; ICC = intra-class correlation; L1 = level 1 (individual); L2 = level 2 (neighborhood); P. Mean = posterior mean; pD = effective number of parameters.

These results can be explained by the fact that the neighborhood-level GSM residuals were almost completely purged of individual-level variance (note the extremely small level 1 residual variances in Table 9). Thus, HLM naturally attributed nearly all the variance in these residuals to differences between clusters. In hindsight, this should have been the prediction in H6. The reason is simple: If the data really are more consistent with spatially rather than hierarchically structured autocorrelation, the fact that all the observations are separated from one another by at least a few meters and observations in different clusters are generally (though not always) separated by longer distances means that the variograms in the original GSM models would virtually guarantee this result.

It is interesting to note that the neighborhood-level variances detected by HLM (in Table 9) are systematically smaller than the neighborhood-level variances detected in the GSM models that produced the residuals in the first place (in Table 6). This implies that the core concept underlying H6 was still correct: There was still more evidence in the data for spatially structured variance than for hierarchically structured variance.

Research Question 3. The third research question asked how GSM estimates of contextual effects and model fit compare to HLM estimates. The associated hypothesis (H7) predicted that (a) GSM models would fit better and have stronger contextual effects of crime and NSES on perceived neighborhood problems than HLM models when the contextual measures are calculated within appropriately sized buffers, and (b) GSM models that use cluster boundaries to measure crime and NSES in GSM models would outperform HLM models, but not by as much as when the GSM models use the buffers instead. The first step in testing H7 was determining the optimal spatial scales for measuring crime and NSES. The second step was comparing HLM and GSM parameter estimates and fit indices. The third step was comparing CAR HLM results to traditional HLM results, and then comparing CAR HLM results to the GSM results.

Optimal buffer size for crime. GSM Models 6-30 varied the buffer radius used to measure crime. Complete tabular output for these models was omitted for several reasons: (a) only selected parameter estimates and model fit indices were relevant to selecting the optimal buffer size, (b) the resulting table would be too large, making it hard to find the key pieces of information, and (c) Figure 13 more concisely and effectively communicates the overall patterns evident across the criteria of interest. However, a table with complete output for the model believed to have the optimal buffer size, plus the models with buffers one step smaller and one step larger, is provided in the Appendix. All parameter estimates from Models 6-30 are available from the author upon request.

Figure 13 shows that using a 1.1 km radius buffer in Model 16 yielded the lowest DIC value; it also produced a large regression coefficient and the largest values for the t statistic, the level 2 PCV, and the overall R^2 among this set of models. It also yielded low

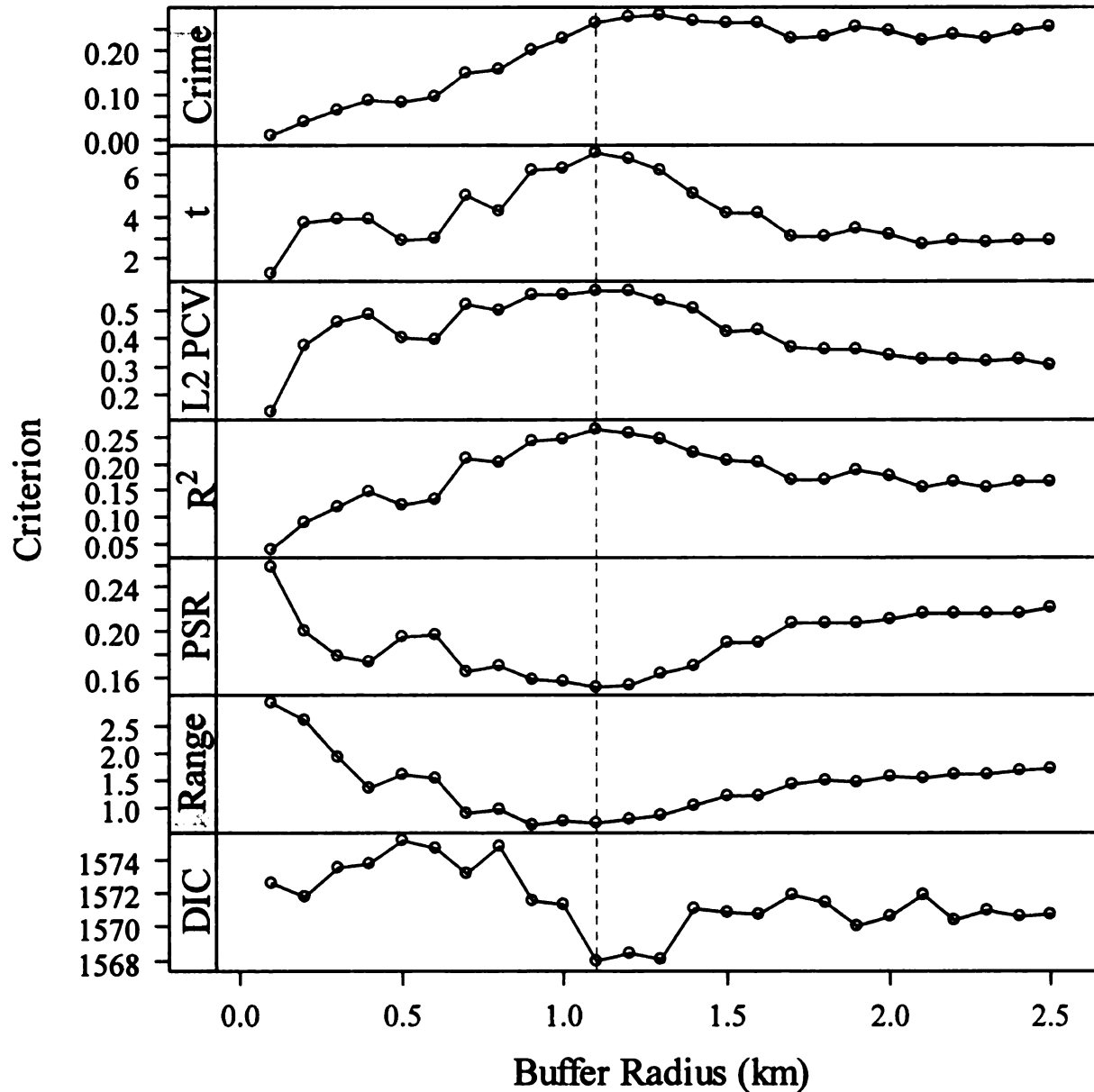


Figure 13: Parameter estimates and model fit criteria for GSM Models 6-30, shown as a function of buffer radius. The dashed, vertical line marks the optimal buffer radius (1.1 km) for measuring crime. Crime = coefficient for the buffer-based crime measure; DIC = deviance information criterion; L2 PCV = level 2 proportional change in variance relative to GSM Model 1 (level-specific R^2); PSR = partial sill ratio; R^2 = overall proportion of variance explained; Range = practical range (in km) of variogram; t = t-statistic for the crime coefficient. These models also included all individual-level predictors.

values for the practical range of remaining spatial autocorrelation and for the PSR. While Models 17 and 18 also had DIC values lower than most of the other models in this series,

they had slightly worse performance with respect to some of the other criteria shown in Figure 13. As a result, 1.1 km was deemed the optimal buffer radius for measuring crime.

Optimal buffer size for NSES. GSM Models 31-55 varied the buffer radius used to measure NSES. As with crime, the results of the analyses used to select the optimal buffer size are summarized graphically for the whole series of models (see Figure 14) and a table with complete details on the model believed to have the optimal buffer size, plus the models with buffers one step smaller and one step larger, is provided in the Appendix. Complete details on all these models are available from the author upon request.

Figure 14 shows that the lowest DIC value was associated with Model 32, which used a 0.2 km buffer to measure NSES. This model also had a large regression coefficient for NSES, the largest t statistic, and large values for the level 2 PCV and the overall R^2 statistic paired with low values for the PSR and the practical range of remaining spatial autocorrelation. While Model 33 had slightly better values on the level 2 PCV, R^2 , PSR, and practical range than Model 32, the DIC, t , and regression coefficient values favored Model 32 instead. Because model selection is one of the intended uses of the DIC, and the difference on the DIC was more noteworthy than the differences on the other measures, the 0.2 km radius buffer was deemed the optimal one for measuring NSES.

Hypothesis 7. GSM Models 3-5 measured crime and NSES within the same cluster boundaries used in HLM Models 3-5, while GSM Models 56-58 rely on measuring crime and NSES within optimally-sized buffers (1.1 km and 0.2 km, respectively). Comparing these three series of models provides the critical test of H7. As shown in Table 3, the models in these series expand Model 2 in each method to include

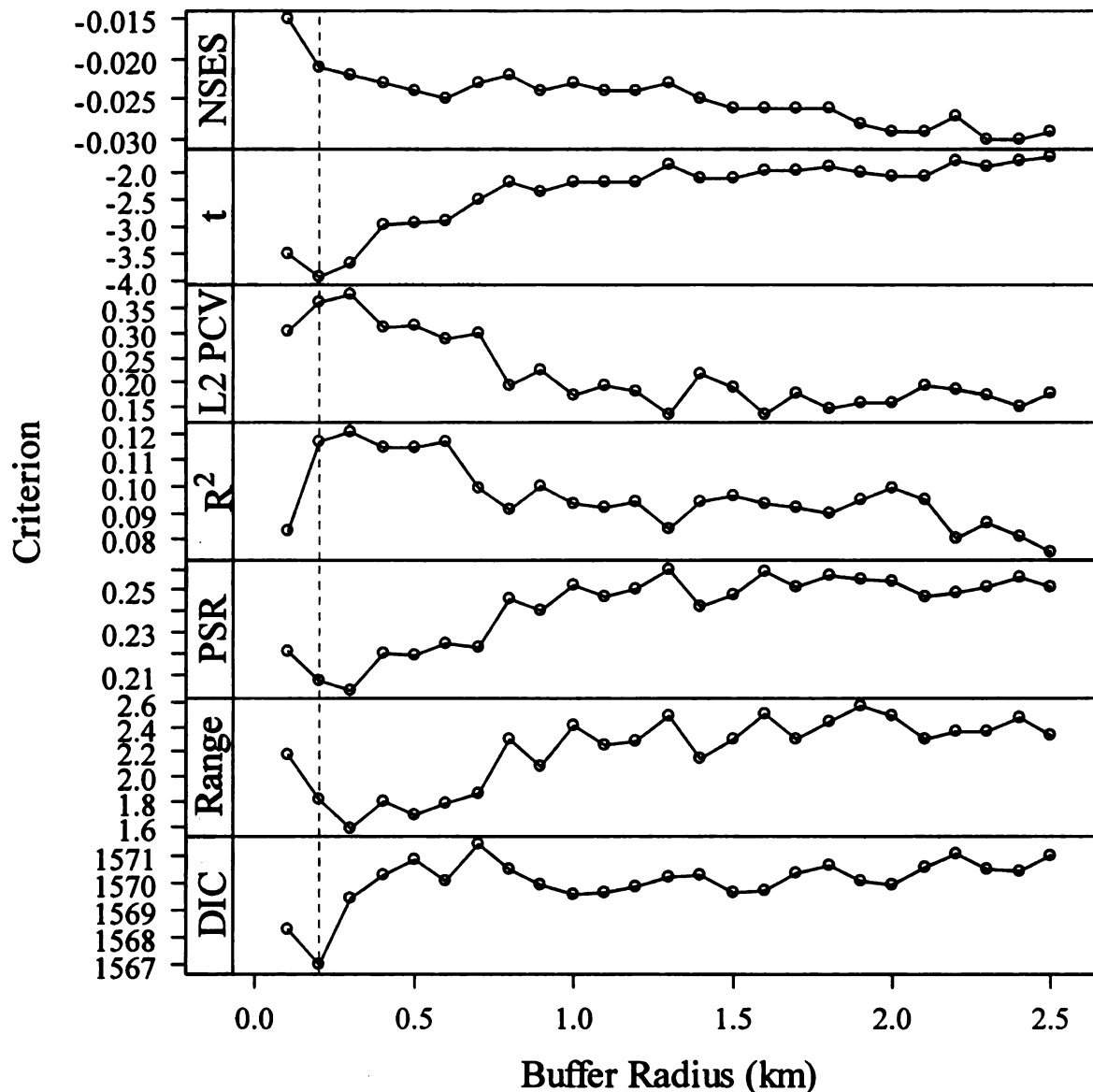


Figure 14: Parameter estimates and model fit criteria for GSM Models 31-55, shown as a function of buffer radius. The dashed, vertical line marks the optimal buffer radius (0.2 km) for measuring NSES. NSES = coefficient for the buffer-based measure of neighborhood socioeconomic status; DIC = deviance information criterion; L2 PCV = level 2 proportional change in variance relative to GSM Model 1 (level-specific R^2); PSR = partial sill ratio; R^2 = overall proportion of variance explained; Range = practical range (in km) of variogram; t = t-statistic for the NSES coefficient. These models also included all individual-level predictors.

neighborhood-level predictors. The first model in each series added only crime, the

second model added only NSES, and the third added both crime and NSES

simultaneously. HLM Model 6 expanded on HLM Model 5 by adding the CAR structure

as an additional spatial random effect in the level 2 model. Enhancing the model that way was meant to test whether incorporating more spatial information into HLM would substantially affect how it compares to the GSM Models 5 and 58.

Tables 5 and 6 contain all the parameter estimates and model fit indices for these models. Three criteria pertinent to model fit (DIC, level 2 PCV, and R^2) for these models are displayed in Figure 15, while estimates of the variance components and levels of residual autocorrelation (ICC and PSR) are displayed in Figure 16. The variograms and autocorrelation functions for GSM Models 3-5 and 56-58 are shown in Figure 17. The parameter estimates and t-statistics for the crime effect in these models are shown in Figure 18, while Figure 19 shows them for the NSES effect. The narrative results related to these figures are presented in separate subsections below that focus on what happened when (a) crime was added by itself, (b) NSES was added by itself, (c) both crime and NSES were added, and (d) the CAR structure was added in HLM Model 6.

Adding crime alone. While adding crime as a neighborhood-level predictor of perceived neighborhood problems improved the HLM and GSM models relative to the corresponding models containing only individual-level predictors, comparing how the updated HLM and GSM models compare to each other was more important to H7. Accordingly, the improvement over the simpler models estimated with the same method is addressed here only to inform that comparison.

The lower panel of Figure 15 shows that the DIC values for the GSM models (regardless of the boundaries used for measuring crime and NSES) were far smaller than the DIC values for the corresponding HLM model, just as they were in the models without any neighborhood-level predictors. Indeed, GSM Models 3 and 56 have DICs of

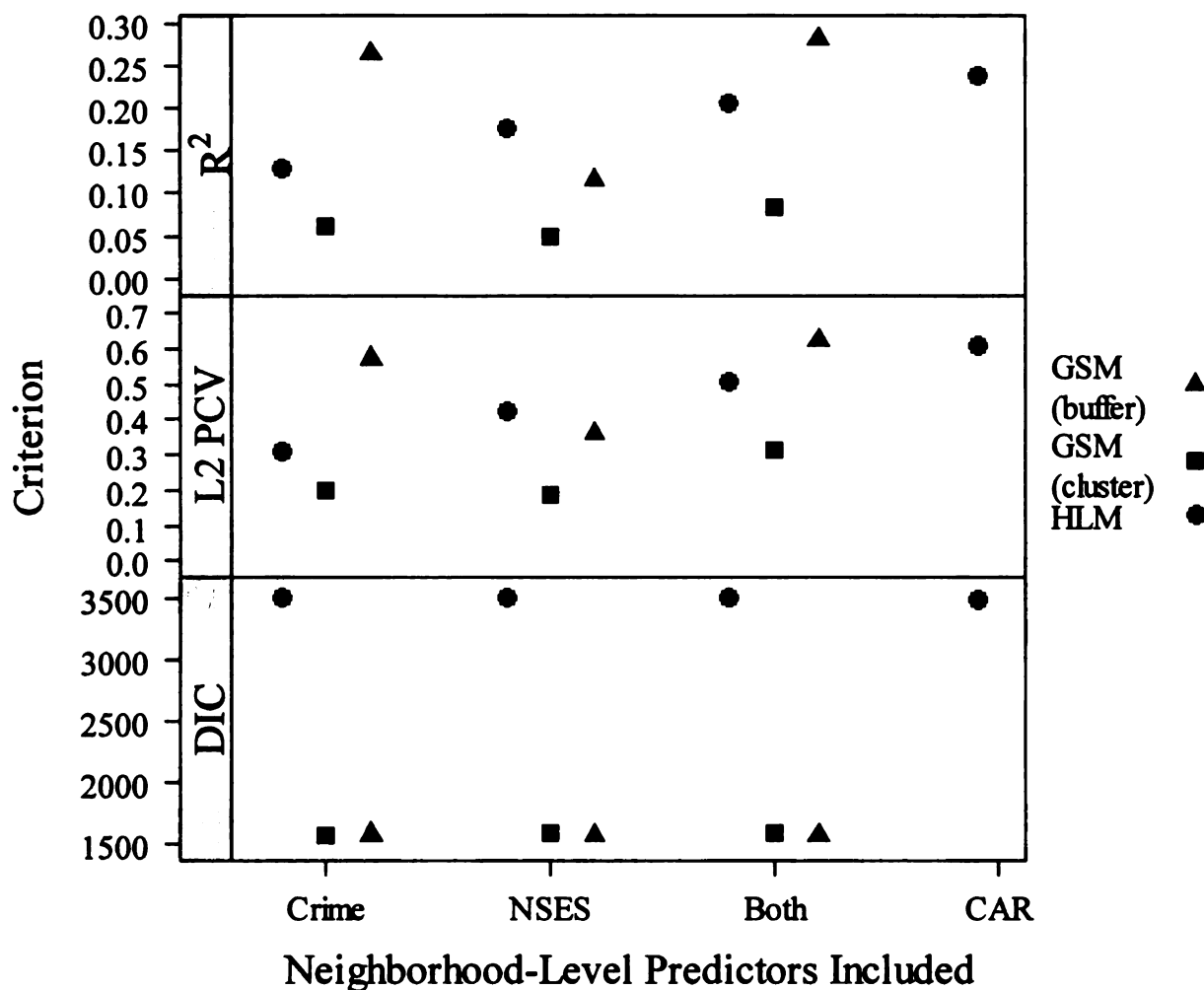


Figure 15: Model fit criteria for HLM Models 3-6 and GSM Models 3-5 and GSM Models 56-58. Both = both crime and NSES; CAR = conditional autoregressive model at level 2 with both crime and NSES as predictors; DIC = deviance information criterion; L2 PCV = level 2 proportional change in variance relative to GSM Model 1 (level-specific R^2); R^2 = overall proportion of variance explained. These models also included all individual-level predictors.

1569.22 and 1567.85 respectively, while HLM Model 3 has a DIC of 3508.30. Finding DICs that favor the two GSM models over the corresponding HLM model by more than 1900 points supports H7 and suggests that GSM outperformed HLM. At first glance, the small difference in DICs between these two GSM models suggests that improvement in model fit compared to HLM may primarily reflect how the autocorrelation was modeled, rather than how neighborhoods were defined for measuring crime.

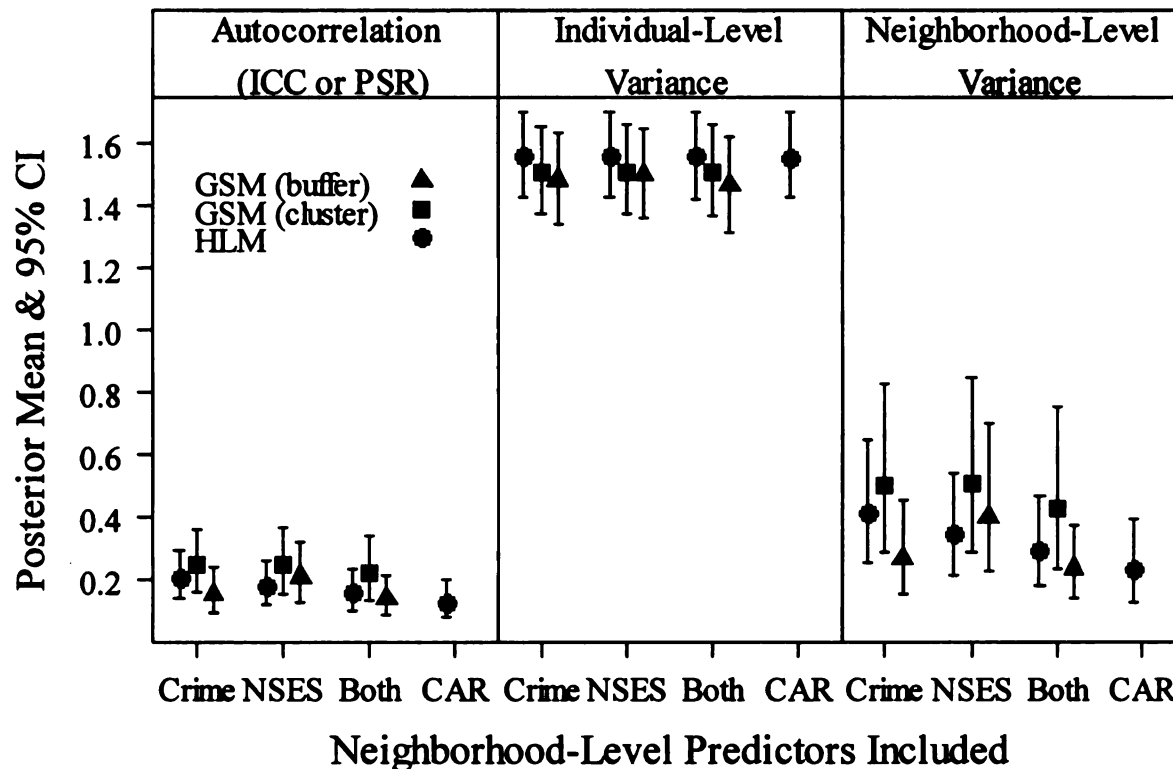


Figure 16: Estimated variance components and levels of autocorrelation from HLM Models 3-6, and GSM Models 3-5 and 56-58. The plot shows the posterior means (symbols) plus the central 95% credible intervals (whiskers) around those estimates. Both = both crime and NSES; CAR = conditional autoregressive model at level 2 with both crime and NSES as predictors. These models also included all individual-level predictors.

However, examining other model fit criteria and parameter estimates revealed a more complex pattern of results that cautions against over-interpreting the differences in DIC values. For example, adding crime in HLM Model 3 decreased the residual autocorrelation in the data by 5.4% as compared to HLM Model 2 (cf. Figures 9 and 16). In comparison, adding the cluster-based crime measure in GSM Model 3 decreased the PSR by only 3.7% compared to GSM Model 2, but adding the buffer-based crime measure in GSM Model 56 decreased the PSR by 13.2% (cf. Figures 9 and 16). It should also be noted that adding crime as a predictor reduced the spatial range over which

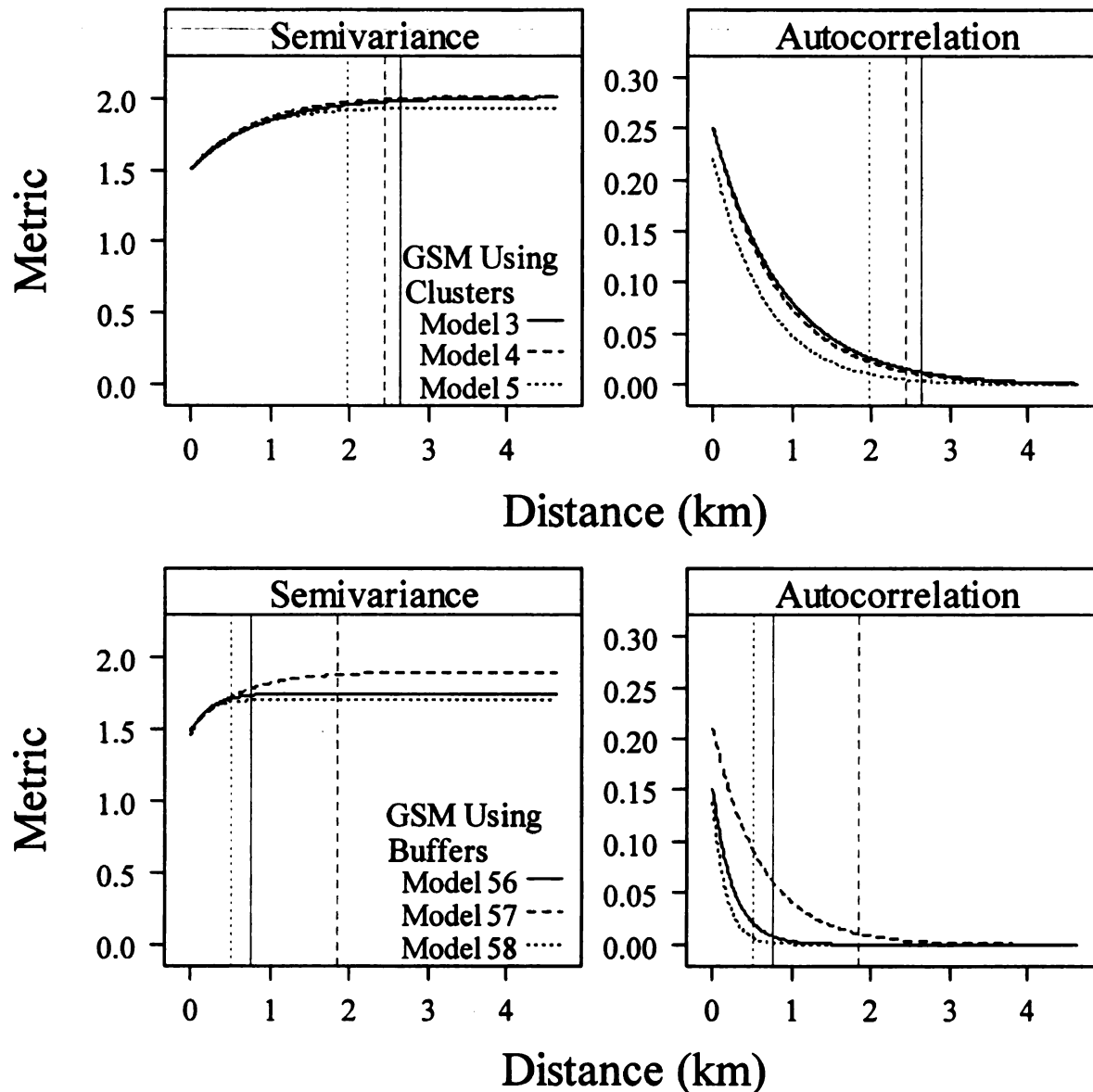


Figure 17: Exponential variograms and correlation functions for GSM Models 3-5 (top) and GSM Models 56-58 (bottom). The maximum value for the autocorrelation is the PSR for each model (for Models 3-5, PSR = .246, .247, and .217 respectively; for Models 56-58, PSR = .151, .208, and .138, respectively). Vertical lines mark the practical ranges of spatial autocorrelation for the models, which occur at the distances where the neighborhood-level covariances have decreased to 5% of their initial sizes, leaving little residual autocorrelation. For Models 3-5, range = 2,647 m, 2,451 m, and 1,964 m respectively; for Models 56-58, range = 765 m, 1840 m, and 505 m, respectively.

autocorrelation persisted in the GSM models from 3,058 m in Model 2 to 2,647 m in Model 3 and 765 m in Model 56 (cf. Figures 10 and 17).

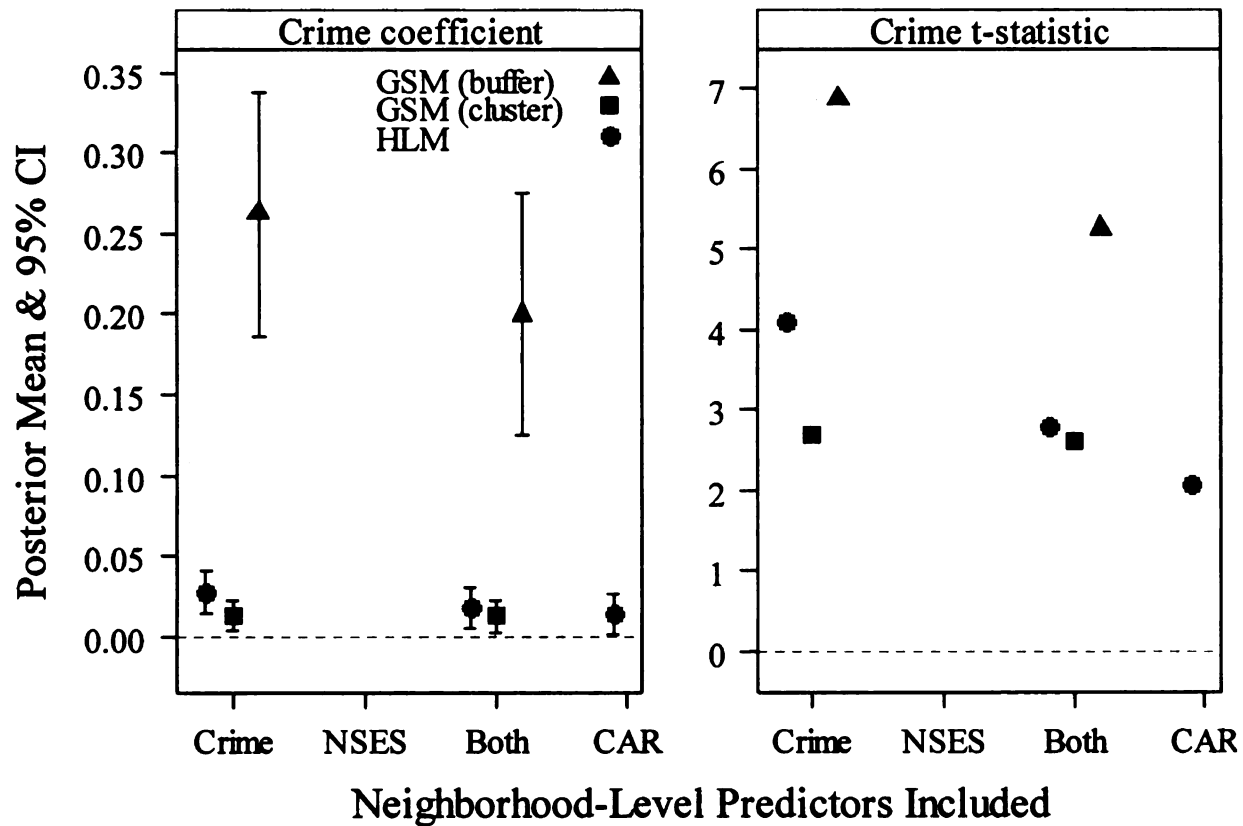


Figure 18: Estimated crime coefficients and t-statistics for HLM Models 3, 5, and 6 and GSM Models 3, 5, 56, and 58. The plot shows the posterior means (symbols) plus the central 95% credible intervals (whiskers) around the estimated coefficients. Credible intervals intersected by the dashed reference line at zero indicate non-significant effects. Both = both crime and NSES were included; CAR = conditional autoregressive model at level 2 with both crime and NSES as predictors. These models included all individual-level predictors.

The point here is that the relative changes in residual autocorrelation (observed within each method by comparing models with and without crime as a predictor) suggest that HLM did slightly, but not significantly, better at reducing the residual autocorrelation than GSM based on using the clusters, but significantly worse than GSM based on using buffers. That conflicts with the information gleaned from the DIC values because while it supports the first part of H7, it fails to support the latter part of that hypothesis.

Because the individual-level variance components remained quite stable across all the models (cf. the level 1 PCV values shown in Tables 5-6, plus the variance estimates in Figures 9 and 16), the relative decreases in residual autocorrelation are mostly

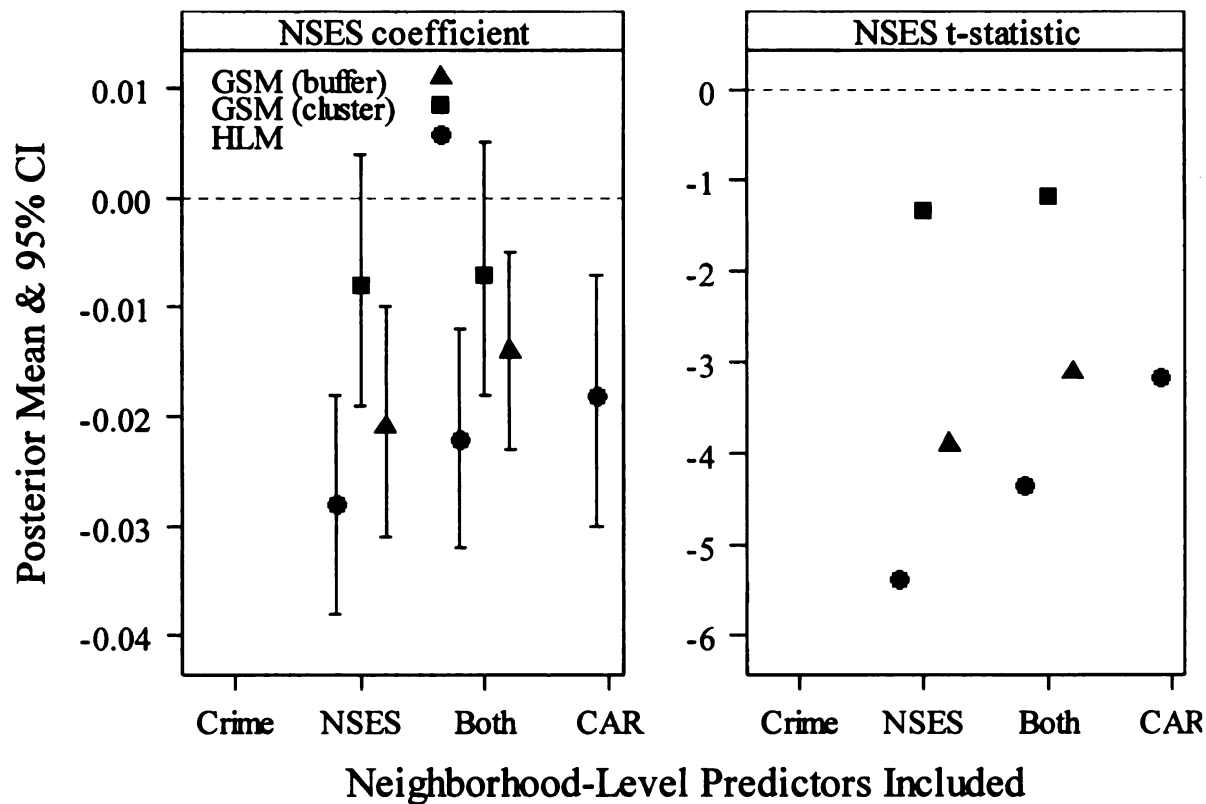


Figure 19: Estimated NSES coefficients and t-statistics for HLM Models 4, 5, and 6 and GSM Models 4, 5, 57, and 58. The plot shows the posterior means (symbols) plus the central 95% credible intervals (whiskers) around the estimated coefficients. Credible intervals intersected by the dashed reference line at zero indicate non-significant effects. Both = both crime and NSES were included; CAR = conditional autoregressive model at level 2 with both crime and NSES as predictors. These models included all individual-level predictors.

attributable to changes in the neighborhood-level variance components. Another way to examine the performance of the two methods is to compare how much variance they explain and directly compare the actual levels of residual autocorrelation across methods rather than relative decreases in autocorrelation observed within each method.

Figure 15 shows that HLM Model 3 (level 2 PCV = .311, $R^2 = .132$) explained 11.5% more neighborhood-level variance and 7.0% more overall variance than GSM Model 3 (level 2 PCV = .196, $R^2 = .062$). This meant that HLM also left 3.8% less residual autocorrelation (ICC = .208, 95% CI = [.139, .295]) than GSM Model 3 (PSR =

.246, 95% CI = [.157, .357]), though there was 88% overlap in those credible intervals (see Figure 16). So, findings from these criteria conflict with the information gleaned from the DIC values and fail to support the latter part of H7, suggesting that when both methods measured crime within the same fixed cluster boundaries, HLM performed better than GSM.

However, using the optimal buffer for measuring crime instead of the cluster boundaries made an immense difference: HLM Model 3 explained 26.2% less neighborhood-level variance and 13.3% less overall variance than GSM Model 56 (level 2 PCV = .573, $R^2 = .265$). As a result, this HLM model left 5.7% more residual autocorrelation than GSM Model 56 (PSR = .151, 95% CI = [.092, .238]), though there was still a lot of overlap (68%) in those credible intervals (see Figure 16). Unlike with the cluster-based GSM model, all three of these criteria agree with the DIC and strongly support the first part of H7 by showing that a buffer-based GSM model fit the data significantly better than the HLM model.

The regression coefficients and t-statistics associated with the crime effects in these series of models also address H7, which predicted that larger effects would be observed in the GSM models. As Figure 18 shows, the crime coefficient and t-statistic for HLM Model 3 ($\gamma = 0.028$, $t = 4.11$) were noticeably larger than those for GSM Model 3 ($\beta = 0.013$, $t = 2.68$), but there was substantial overlap (47%) in their respective credible intervals (95% CIs = [0.014, 0.041] and [0.004, 0.023]). So, when crime was measured within cluster boundaries, its effect was weaker in GSM than in HLM, which fails to support H7. However, the story was quite different when crime was measured in the 1.1 km buffer: The crime coefficient ($\beta = 0.263$, 95% CI = [0.186, 0.338], $t = 6.88$) in GSM

Model 56 was an order of magnitude larger than it was in HLM Model 3, plus there was no overlap at all between those two credible intervals. That finding strongly supports H7.

Adding NSES alone. HLM Model 4 used NSES as the sole neighborhood-level predictor instead of crime, yielding a DIC of 3,505.70. This was a definite improvement over HLM Model 2 with respect to all the criteria of interest in this study (see Table 5). But, as Figure 15 shows, the DIC still favors the corresponding GSM models (Model 4 DIC = 1,573.47, Model 57 DIC = 1,566.95) by over 1900 points, suggesting support for H7. But, as with the crime models, other criteria revealed that this large discrepancy in DIC values does not tell the whole story about which method performs better.

Adding NSES in HLM Model 4 decreased the residual autocorrelation in the data by 8.4% as compared to HLM Model 2 (cf. Figures 9 and 16). For the GSM models, adding the cluster-based NSES measure in Model 4 decreased the PSR by only 3.6% compared to Model 2 and adding the buffer-based NSES measure in Model 57 decreased the PSR by 7.5% (cf. Figures 9 and 16). This meant that HLM Model 4 (ICC = .180) had a significantly lower level of remaining autocorrelation than GSM Model 4 (PSR = .247), but not GSM Model 57 (PSR = .208). It should also be noted that adding NSES as a predictor reduced the spatial range over which autocorrelation persisted in the GSM models from 3,058 m in Model 2 to 2,451 m in Model 4 and 1,840 m in Model 57 (cf. Figures 10 and 17). So, HLM did noticeably better at reducing the level of residual autocorrelation (as determined by comparing to a simpler model run with the same method) than GSM based on using the clusters but only marginally better than GSM based on using buffers. That conflicts with the conclusion that might be drawn from the DIC values because it fails to support either part of H7.

Once again, the relative decreases in residual autocorrelation are mostly attributable to changes in the neighborhood-level variance components (cf. the level 1 PCV values shown in Tables 5-6, plus the variance estimates in Figures 9 and 16). Figure 15 shows that HLM Model 4 (level 2 PCV = .425, $R^2 = .177$) performed better than GSM on both variance explained criteria when NSES was the only neighborhood-level predictor. It explained 23.7% more neighborhood level variance and 12.9% more overall variance than GSM Model 4 (level 2 PCV = .188, $R^2 = .048$). It also explained 6.7% more neighborhood-level variance and 6.1% more overall variance than GSM Model 57 (level 2 PCV = .358, $R^2 = .116$). So, for these models, the DIC and the variance explained criteria lead to conflicting conclusions about which model fits the data better because the former criterion suggests support for H7, while the latter criteria refute H7.

Hypothesis 7 also predicted that the NSES regression coefficients and t-statistics in these series of models would be larger for the GSM models than for the HLM models. Figure 19 shows that the NSES coefficient and t-statistic for HLM Model 4 ($\gamma = -0.028$, $t = -5.38$) were significant and much larger than the non-significant values observed for GSM Model 4 ($\beta = -0.008$, $t = -1.33$). In addition, there was only 5% overlap in their respective credible intervals (95% CIs = [-0.038, -0.018] and [-0.019, 0.004]). So, GSM found a weaker effect than HLM when NSES was measured within cluster boundaries, which fails to support H7. The difference was less extreme but still followed the same pattern when crime was measured in the 0.2 km buffer: GSM Model 57 produced a significant NSES coefficient and t-statistic ($\beta = -0.021$, 95% CI = [-0.031, -0.010], $t = -3.90$) that were much closer to the results from HLM Model 4, but still smaller. There

was 65% overlap between those two credible intervals. So, with NSES as the only neighborhood-level predictor, H7 was not supported because HLM always produced larger NSES effects.

Adding both crime and NSES. The pattern of the DIC favoring the two GSM models (Model 5 DIC = 1,573.32, Model 58 DIC = 1,564.06) over the HLM model (Model 5 DIC = 3,505.20) by over 1900 points is still evident in Figure 15 when both crime and NSES are used as neighborhood-level predictors. However the other criteria continued to reveal that this large discrepancy in DIC values does not tell the whole story about which method performs better.

Adding both neighborhood-level predictors in HLM Model 5 decreased the residual autocorrelation in the data by 10.7% (to ICC = .157) as compared to HLM Model 2 (cf. Figures 9 and 16). For the GSM models, adding the cluster-based measures in Model 5 decreased the PSR by only 6.6% (to PSR = .217) compared to Model 2 and adding the buffer-based measures in Model 58 decreased the PSR by 14.5% (to PSR = .138; cf. Figures 9 and 16). It should also be noted that adding crime and NSES together reduced the spatial range over which autocorrelation persisted in the GSM models from 3,058 m in Model 2 to 1,964 m in Model 5 and 505 m in Model 58 (cf. Figures 10 and 17). Overall, these comparisons reveal that HLM did somewhat better at reducing the residual autocorrelation (as determined by comparing to a simpler model run with the same method) than GSM based on using the clusters and somewhat worse than GSM based on using buffers. That conflicts with the conclusion that might be drawn from the DIC values because it supports only the first part of H7.

Figure 15 shows that when both crime and NSES are in the models, HLM Model 5 (level 2 PCV = .512, $R^2 = .207$) explained 19.7% more neighborhood-level variance and 12.2% more overall variance than GSM Model 5 (level 2 PCV = .315, $R^2 = .085$), but 11.0% less neighborhood-level variance and 7.5% less overall variance than GSM Model 58 (level 2 PCV = .622, $R^2 = .282$). Both sets of comparisons exceed the 5% difference criterion adopted for this study as defining what constituted a significant difference in performance. While this supports the H7 prediction that buffer-based GSM models would outperform HLM models, it fails to support the H7 prediction that cluster-based GSM models would also do so (but by a smaller margin).

When both neighborhood-level predictors were included in the model, the crime coefficient for HLM Model 5 decreased ($\gamma = 0.018$, $t = 2.80$), becoming more similar to, but remaining larger than, the coefficient for GSM Model 5 ($\beta = 0.013$, $t = 2.60$) which was largely unchanged. This increased the overlap in their respective credible intervals to 90% (95% CIs = [0.005, 0.030] and [0.003, 0.023], see Figure 18). Once again, the hypothesis that a GSM model relying on cluster boundaries would yield a larger crime effect than the corresponding HLM was not supported. Hypothesis 7 was still partially supported by the finding that the buffer-based crime effect in GSM Model 58 ($\beta = 0.200$, 95% CI = [0.125, 0.275], $t = 5.26$) was still an order of magnitude larger than the crime effect in HLM Model 5, with no overlap in their credible intervals (see Figure 18).

With both crime and NSES in the model, the NSES coefficient for HLM Model 5 decreased slightly ($\gamma = -0.022$, $t = -4.34$), but remained larger than the coefficient for GSM Model 5 ($\beta = -0.007$, $t = -1.18$), which was largely unchanged (see Figure 19). This

increased the overlap in their respective credible intervals to 30% (95% CIs = [-0.032, -0.012] and [-0.018, 0.005]). The NSES effect in GSM Model 58 ($\beta = -0.014$, 95% CI = [-0.023, -0.005], $t = -3.11$) remained smaller than the NSES coefficient from HLM Model 5 and also decreased in size when crime was present in the model, reducing the overlap between those two credible intervals to 61% (see Figure 19). With respect to the NSES effect, H7 was not supported by these model comparisons.

In summary, the balance of the evidence shows that when using cluster-based measurement of crime and NSES with both methods, HLM performed better than GSM. However, the DIC, variance explained, crime coefficient, and crime t-statistic all indicate that buffer-based GSM did indeed outperform HLM as expected when both crime and NSES were in the models. That result is apparently driven by how crime was measured because the NSES effect was actually somewhat weaker in the GSM model than in the HLM model.

Comparing CAR HLM to standard HLM. Before comparing the CAR HLM model to the GSM models, it is useful to first understand how it compared to HLM Model 5. As Table 5 and Figures 15-16 show, even though the CAR variance component in HLM Model 6 was itself extremely small (95% CI = [0.000, 0.002]), adding this spatially structured random effect to the model had a modest, favorable impact on the results. Relative to HLM Model 5 (DIC = 3505.20, level 2 PCV = .512, $R^2 = .207$, ICC = 0.157), the DIC for HLM Model 6 improved by about 1.5 points (DIC = 3503.70) and the overall variance explained increased by 3.2% ($R^2 = .239$). The CAR HLM model also explained about 10% more of the neighborhood-level variance (level 2 PCV = .610) than HLM

Mo

ber

aut

aff

we

(γ

als

-3

co

an

a

n

t

g

t

Model 5. However, there was no difference in the individual-level variance component between HLM Models 5 and 6. Thus, the amount of hierarchically structured residual autocorrelation decreased by 2.8% ($ICC = 0.129$) in the CAR HLM model.

Figures 18 and 19 show how adding the CAR structure to the HLM model affected the coefficients for crime and NSES. The CAR HLM Model 6 had a slightly weaker crime effect ($\gamma = 0.014$, 95% CI = [0.001, 0.026], $t = 2.08$) than HLM Model 5 ($\gamma = 0.018$, 95% CI = [0.005, 0.030], $t = 2.80$). Judging by the difference in t-statistics, it also had significantly weaker NSES effect ($\gamma = -0.018$, 95% CI = [-0.030, -0.007], $t = -3.16$) than HLM Model 5 ($\gamma = -0.022$, 95% CI = [-0.032, -0.012], $t = -4.34$).

Comparing CAR HLM to cluster-based GSM. Next, the CAR HLM model was compared to the cluster-based GSM Model 5. The boundaries used for measuring crime and NSES are identical for these two models, so the models differ only in how autocorrelation was represented. Figures 15-16 facilitate graphical comparison of the model fit indices and variance components for these models. Despite the fact that the DIC for HLM Model 6 is more than 1900 points larger than the DIC for GSM Model 5 (indicating poorer fit), the CAR HLM model explains 15.4% more overall variance and 29.5% more of the neighborhood-level variance than the GSM model. There was also 8.8% less residual autocorrelation in the CAR HLM model ($ICC = .129$) than in GSM Model 5 ($PSR = .217$).

Figures 18-19 show the crime and NSES coefficients for both models. The crime effect was virtually identical between the CAR HLM model ($\gamma = 0.014$, 95% CI = [0.001, 0.026], $t = 2.08$) and GSM Model 5 ($\beta = 0.013$, 95% CI = [0.003, 0.023], $t = 2.60$), though the GSM model produced a narrower credible interval that was completely

enveloped by the corresponding CAR HLM credible interval (100% overlap). However, there was a big difference in the size of the NSES effect, which was significant in the CAR HLM model ($\gamma = -0.018$, 95% CI = $[-0.030, -0.007]$, $t = -3.16$), but not in the GSM model ($\beta = -0.007$, 95% CI = $[-0.018, 0.005]$, $t = -1.18$). There was 48% overlap between these NSES credible intervals.

Despite the DIC value favoring the GSM model, the CAR HLM model seems to have performed better than the cluster-based GSM model according to most measures of model fit. This indicates a lack of support for H7's prediction that the GSM model would yield better model fit. Furthermore, this comparison also failed to support H7 because the cluster-based GSM model failed to produce stronger contextual effects of crime and NSES than the CAR HLM model.

Comparing CAR HLM to buffer-based GSM. The final test of H7 involved comparing the CAR HLM model to the buffer-based GSM Model 58. These models differ in both how boundaries used for measuring crime and NSES were defined and in how autocorrelation was represented. Figures 15-16 provide a graphical comparison of the model fit indices and variance components for these models. Once again, the DIC for HLM Model 6 is more than 1900 points larger than the DIC for GSM Model 58 (indicating poorer fit). However, switching to buffer-based contextual measures for the GSM model rather than cluster-based measures made the performance of the two models much more comparable. The CAR HLM model explains 4.3% less overall variance and 1.2% less of the neighborhood-level variance than the GSM model. However, there was also 0.9% less residual autocorrelation in the CAR HLM model ($ICC = .129$) than in GSM Model 58 ($PSR = .138$).

Figures 18-19 show the crime and NSES coefficients for both models. The crime effect was far smaller in the CAR HLM model ($\gamma = 0.014$, 95% CI = [0.001, 0.026], $t = 2.08$) than in GSM Model 58 ($\beta = 0.200$, 95% CI = [0.125, 0.275], $t = 5.26$), with no overlap at all in these credible intervals. However, there was little difference in the size of the NSES effect, which was significant in both the CAR HLM model ($\gamma = -0.018$, 95% CI = [-0.030, -0.007], $t = -3.16$) and in the GSM model ($\beta = -0.014$, 95% CI = [-0.023, -0.005], $t = -3.11$). This GSM model produced a narrower NSES credible interval that was completely enveloped by the CAR HLM credible interval for NSES (100% overlap).

So, here the DIC value again favored the GSM model, but the CAR HLM model seems to have performed nearly as well the buffer-based GSM model according to most measures of model fit. This indicates a lack of support for H7's prediction that the GSM model would yield better model fit. While these results did partially support H7 because the buffer-based GSM model produced a far stronger contextual effect of crime than the CAR HLM model, they failed to support H7 with respect to the size of the NSES effect.

Research Question 4. The final research question asked how the geographical scales on which crime and NSES operate compare to each other and to the size of the clusters used in the HLM models. The corresponding exploratory hypothesis, H8, stated that these two contextual effects would operate at different geographical scales, and that neither would operate at the scale of the average cluster used in the HLM models.

The geographical area enclosed by the 52 cluster boundaries used in the HLM models averaged 0.08 km^2 and ranged from 0.03 to 0.47 km^2 . Figure 20 illustrates the range of GSM buffer sizes tested in this study and highlights three key facts. First, H8 is supported by the fact that there is a very marked difference in the optimal buffer sizes for

crime (1.1 km radius, 3.80 km²) and NSES (0.2 km radius, 0.13 km²). There is little doubt that these are dramatically different spatial scales for measuring contextual conditions. Second, H8 is also supported by the fact that the buffer-based crime measure used in GSM models 56 and 58 is clearly operating on a far larger spatial scale than the cluster-based crime measures used in the HLM models. Finally, while the buffer-based NSES measure is still operating on a spatial scale slightly larger than the mean cluster size, it falls well within the range of cluster sizes. So, H8 was not fully supported because the spatial scale of the buffer-based NSES measure is quite similar to the sizes of the clusters used in the HLM models.

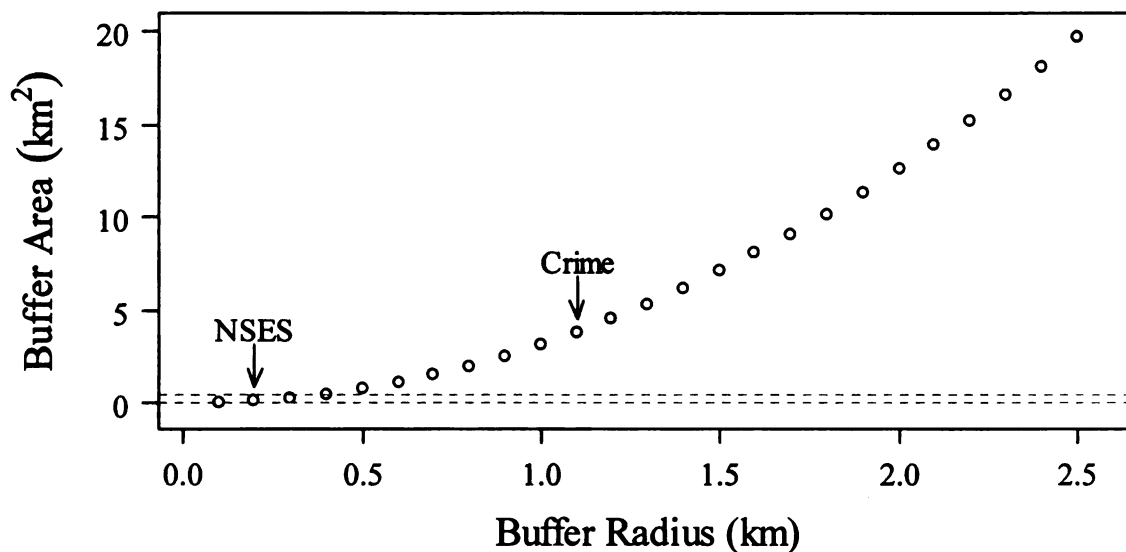


Figure 20: Scatterplot showing buffer area (km²) as a function of buffer radius (km), with annotations showing the optimal buffer sizes for crime and NSES. The dashed horizontal reference lines show the minimum (0.03 km²) and maximum (0.47 km²) areas among the 52 clusters. The mean cluster area was 0.08 km². Arrows show the optimal buffer sizes for crime (1.1 km radius, 3.80 km²), and NSES (0.2 km radius, 0.13 km²).

DISCUSSION

The fundamental premises underlying research on neighborhood effects are that neighborhoods are meaningful ecological contexts for the people who reside in them and that variation in neighborhood characteristics can explain at least some of the variation in resident outcomes. Studies pursuing questions about neighborhood effects are thus inherently multilevel studies and must conceptualize and operationally define neighborhoods as units of analysis and pay careful attention to measuring neighborhood-level constructs (Linney, 2000). Furthermore, they should utilize methods specifically designed for answering questions about contextual effects (Luke, 2005; Shinn & Rapkin, 2000). Suitable multilevel analysis methods must acknowledge and correct for the fact that if neighborhoods actually do affect residents, then the resident-level observations cannot all be independent (Raudenbush & Bryk, 2002; Roosa, et al., 2003). This autocorrelation is a consequence of neighborhood-level variability in the outcome, which may be caused by either compositional or contextual neighborhood effects (or both).

Broadly speaking, this study compared two methods for testing multilevel hypotheses about how much influence contextual characteristics of neighborhoods have on resident outcomes. The first method is well-established in the neighborhood effects literature: HLM has been applied by community psychologists and other social scientists to study a wide variety of phenomena (Beyers, et al., 2003; Caughey, et al., 2008; T. E. Duncan, et al., 2003; Sampson, et al., 1997; Sunder, et al., 2007). The second method—GSM—has only been applied a few times outside of its original applications in the earth sciences (e.g., geology and geography), mostly for epidemiological studies of neighborhood effects on health and healthcare utilization (Chaix, et al., 2006; Chaix,

Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian, et al., 2005). This study sought to discern whether GSM is a valuable alternative to HLM for studying neighborhood effects. To do that, it examined crime and neighborhood socioeconomic status (NSES) effects on residents' perceived levels of neighborhood problems using both methods.

HLM and GSM are grounded in different ways of conceptualizing neighborhoods and geographic space. These conceptualizations inform two key aspects of neighborhood studies: (a) how we group residents in order to detect neighborhood-level variability and model the resulting autocorrelation in outcomes, and (b) how we define the geographic area of the neighborhood that should be used when measuring neighborhood context. This study answered four research questions related to these issues by testing eight specific hypotheses (see Table 10 below) to examine whether the conceptual differences between these methods contributed to differences in scientific inferences about the phenomena under study that warrant further usage of GSM in community psychology.

Overall, the study found that while empty HLM and GSM models detected similar amounts of neighborhood-level variance and autocorrelation in perceived neighborhood problems, GSM provided a better description of the data from this sample because crucial HLM assumptions about the independence of the residuals were violated. In contrast, GSM assumptions about the residuals were not violated. This study also found that, for the present sample, circular buffers centered on residents' homes provided a better operational definition of the neighborhoods within which crime and NSES should be measured than that offered by the fixed cluster boundaries required by HLM. The specific boundaries used to measure the contextual variables had important implications for the size and statistical significance of the crime and NSES effects in this study.

Table 10: List of research questions and hypotheses

Research Questions	Hypotheses
1. How do GSM estimates of neighborhood-level variance and autocorrelation compare to HLM estimates?	H1: GSM estimates of neighborhood-level variance and the amount of autocorrelation for perceived neighborhood problems will be higher than the corresponding HLM estimates, both before and after controlling for neighborhood composition H2: The range of spatial autocorrelation in perceived neighborhood problems detected by GSM will be long enough to reach across the borders between at least some of the neighborhood units used in the HLM analyses.
2. Which method (HLM or GSM) is more effective at modeling the autocorrelation actually observed in data from neighborhood residents?	H3: An empty GSM will fit the perceived neighborhood problems data better than an empty HLM. Similarly, a GSM model of perceived neighborhood problems containing only individual-level predictors will fit better than a corresponding HLM model containing only individual-level predictors of perceived neighborhood problems. H4: HLM will not fully control for spatial autocorrelation in perceived neighborhood problems, so there will be evidence of residual spatial autocorrelation remaining in both the Level 1 and Level 2 residuals from HLM models. H5: GSM will fully control for within-neighborhood spatial autocorrelation in residents' perceptions of neighborhood problems, so there will be no evidence of hierarchical autocorrelation remaining in the individual-level residuals from GSM models. H6: Neighborhood-level GSM residuals from a model predicting perceived neighborhood problems will contain hierarchical autocorrelation when examined with HLM, but the ICC will be lower than the PSR.
3. How do GSM estimates of contextual effects and model fit compare to HLM estimates?	H7: GSM will yield models that fit better and have larger contextual effects of crime and NSES on perceived neighborhood problems than corresponding HLM models when they use contextual measures calculated within appropriately-sized buffers. Using HLM-style contextual measures of crime and NSES calculated within discrete neighborhood cluster boundaries in GSM analyses will yield models of perceived neighborhood problems that improve on HLM results, but not as much as when buffers are used.
4. In a dataset originally collected with use of HLM methods in mind, how do the geographical scales on which different contextual factors operate (as estimated by GSM) compare to each other and to the size of the neighborhood units used in HLM?	H8: The geographical scales on which crime and NSES influence resident perceptions of neighborhood problems will differ from one another and from the average size of the neighborhood areas used in the HLM analysis.

At least for the present dataset, HLM models overestimated the size and statistical significance of the effect of a cluster-based measure of NSES on residents' perceptions of neighborhood problems because of the violated assumptions about the independence of the residuals. GSM corrected for that mis-specified error structure and showed that while the cluster-based NSES measure did not affect residents' perceptions in these data, when NSES was instead measured in 0.2 km radius buffers around residents' homes, it did affect those perceptions. However, the NSES effect detected with the buffer-based GSM analysis was not as strong as the cluster-based NSES effect in the HLM analysis.

HLM also severely underestimated the strength of crime's effect on residents' perceptions in this study because the clusters used in the HLM analysis were far too small compared to the actual spatial scale on which crime mattered to the residents. Buffer-based GSM models showed that crime within 1.1 km of residents' homes had a much stronger effect on perceived neighborhood problems than the cluster-based crime effect observed with the HLM models.

The findings supported some, but not all of the hypotheses in this study. The discussion below synthesizes the findings with related literature, links them to key theoretical and conceptual issues, shows how the study contributes to neighborhood research, and describes what the results may mean for community psychologists.

Detecting Neighborhood-Level Variability in Perceived Neighborhood Problems

Before testing whether neighborhood-level characteristics like crime and NSES influence resident-level outcomes such as perceived neighborhood problems, one ought to first show that those outcomes do indeed vary from neighborhood to neighborhood and

quantify that neighborhood-level variance. Otherwise, there may be nothing for those contextual characteristics to explain.

HLM and GSM rely on distinct ways of grouping residents to detect and quantify neighborhood-level variability, but both ultimately divide the total variability in residents' perceptions of neighborhood problems into individual-level and neighborhood-level variance components. Those variance components can then be used to calculate directly comparable measures of autocorrelation (the intra-class correlation [ICC] for HLM and the partial sill ratio [PSR] for GSM) that represent the proportion of variability in perceived problems that is attributable to differences between neighborhoods. To put the current findings in perspective, the next section briefly describes how much autocorrelation has been detected in perceived neighborhood problems in previous HLM and GSM studies. The following section then interprets the current findings and discusses their importance.

Previous research. There is little consensus in previous HLM research about how much autocorrelation there is in perceived neighborhood problems and closely related constructs such as perceived crime or perceived disorder. For a sample of residents drawn from neighborhoods in ten different cities, Coulton et al. (2004) reported ICCs ranging from .04 to .10, depending on the size of the neighborhood units used in the HLM models (ICC = .09 for census tracts, and .10 for block groups). A measure of perceived neighborhood crime from a 1978 survey of Chicago residents had a tract-level ICC of .10 (Quillian & Pager, 2001), but a measure of perceived disorder from a 1995 survey of Chicago residents had a block-group level ICC of .35 (Sampson & Raudenbush, 2004). Meanwhile, recent survey data on perceived disorder from Baltimore found a block-

group level ICC of .40 (Franzini, et al., 2008). The latter two ICCs are likely slight underestimates because they were calculated from models that controlled for individual-level predictors rather than empty models. These HLM studies collectively show that the amount of autocorrelation in perceived neighborhood problems (and closely related constructs) varies from study to study and with the size of the neighborhood units.

There are only two previous studies that have applied variations of GSM to study perceived neighborhood problems or closely related constructs. Inspection of Figure 3 in Bass and Lambert's (2004) GSM-based work suggests that the PSR for perceptions of neighborhood disorder might be as high as .57 among Baltimore adolescents. This is somewhat higher than the ICC from Franzini et al.'s (2008) HLM study ($ICC = .40$), which was conducted in the same city. The only other previous study that used GSM methods to examine neighborhood-level variability in perceived neighborhood problems found a PSR of .32 (Pierce, 2006), but that was a preliminary analysis of the data from the present study and it used different estimation methods than were adopted here.

None of the prior research had ever used both HLM and GSM to quantify the level of autocorrelation in perceived problems in the same dataset. Hence, the first research question in this study asked: how do GSM estimates of neighborhood-level variance and autocorrelation compare to HLM estimates? Examining these estimates is important because the level of autocorrelation observed in an empty model (i.e., one with no substantive predictors) puts an upper bound on the amount of variance that can be explained by differences between neighborhoods. Re-examining those estimates after adjusting for individual-level predictors sheds light on whether neighborhood-level variability is primarily a result of compositional or contextual effects. The first

hypothesis (H1) predicted that GSM would detect larger neighborhood-level variances and higher levels of autocorrelation for perceived neighborhood problems than HLM did, both before and after controlling for neighborhood composition.

Current findings. Contrary to H1, the GSM estimates of neighborhood-level variance and autocorrelation in the present study were only trivially larger than the HLM estimates both before and after controlling for neighborhood demographic composition. The ICC of .28 in the present HLM analyses indicates that about 28% of the variability in neighborhood problems scores from this dataset can be attributed to differences between neighborhoods. A similar result was found with the GSM analyses ($PSR = .29$), which found that 29% of the variability in those scores is attributable to neighborhoods. These estimates are closer to the upper end of the range of values observed in the prior HLM studies focused on this outcome, but lower than the prior GSM estimates.

So, the answer to the first research question is that these two methods essentially agreed on how much neighborhood-level variance and autocorrelation existed in residents' perceptions of neighborhood problems in this study. Furthermore, controlling for several individual level characteristics made little difference in the autocorrelation estimates from these methods, reducing them to 26% for HLM and 28% for GSM. As a result, we can draw the same conclusion from using both methods. While it is possible that individual-level characteristics omitted from this study may still be important, the present results suggest that neighborhood composition effects arising from geographical clustering of similar persons do not provide a compelling theoretical explanation for most of the observed neighborhood-level variability in this sample. Other theoretical mechanisms must be at work here; those mechanisms are probably related to contextual

characteristics of the neighborhoods. The similarity of those revised autocorrelation estimates means we can also conclude from both the HLM and GSM analyses that there is substantial potential for neighborhood characteristics to exert contextual influences on these resident perceptions.

Although the two methods provided similar estimates in this study, researchers cannot take it for granted that HLM and GSM will always yield similar estimates of neighborhood-level variances and levels of autocorrelation. Whether or not they do may depend crucially on the outcome being studied. For example, Chaix et al. (2005) found much smaller neighborhood-level variances in their GSM models of two health care utilization measures than in corresponding HLM models, even when they varied the size of the geographic units used for grouping observations in the HLM models. The contrast between their findings and the present findings suggests that HLM and GSM variance estimates (and the autocorrelation estimates calculated from them) could differ from each other more for some constructs than others. Coulton et al.'s (2004) observation that the sensitivity of the ICC to changes in the size of the neighborhood units depended on the specific construct being examined suggests that both the specific outcome and the size of the neighborhood units in the HLM model may affect whether HLM and GSM yield similar or different neighborhood-level variances.

Future research should consider that possibility that HLM and GSM will yield different estimates of the variance components and levels of autocorrelation rather than assuming they will turn out to be similar as they have in this study. If we accumulate more empirical evidence from such comparisons, we may be able to better understand the conditions under which such estimates are likely to converge or diverge. Next, the focus

shifts to an issue that is frequently neglected in the HLM studies: The spatial scale on which autocorrelation in perceived neighborhood problems was observed.

Spatial Scale of Autocorrelation for Perceived Neighborhood Problems

Understanding how far apart residents need to be before we can expect their perceptions of neighborhood problems to be essentially independent of one another provides researchers with a rough upper bound on the potential size of the neighborhood areas that may influence residents. Comparing the spatial scale on which autocorrelation was observed in perceived neighborhood problems with HLM and GSM offers another way to answer the first research question. GSM methods routinely provide information about the spatial scale of autocorrelation in the data (Banerjee, et al., 2004; Chaix, et al., 2006; Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian, et al., 2005; Diggle & Ribeiro, 2007), but HLM studies are less likely to directly discuss this spatial aspect of the phenomena being analyzed.

Spatial scale in HLM. Even if we accept the conceptualization of geographic space and neighborhoods underlying HLM, we still must consider whether we are using neighborhood units that are properly sized to capture the neighborhood-level variability in outcomes. On one hand, a researcher who uses neighborhood units that are too large compared to the neighborhood areas that are relevant to residents will effectively be grouping together people who belong to different neighborhoods. On the other hand, using neighborhood units that are too small would divide people from the same neighborhood and place them in separate units. Either of these situations would dilute the ability to detect neighborhood-level variance and the associated autocorrelation; they also might make it harder to detect the effects of neighborhood-level variables. So, one

problem with HLM studies is that, with only a few notable exceptions (Coulton, et al., 2004), researchers rarely report the sensitivity of HLM results to the size of the neighborhood units.

Given the way HLM is typically used, the only way to describe the spatial scale on which autocorrelation exists in the outcome data being modeled is to describe the geographic size of the neighborhood units that were used to group the residents. Because those units are not required to be the same shape or size, they may vary in geographic size, so one can usually get only a rough description of how far autocorrelation might reach by looking at descriptive statistics about their size. That means HLM will rarely provide precise information about issues of spatial scale even when authors are paying close attention to that issue.

But, many HLM studies have not even directly reported the physical size of the neighborhood units they used (Franzini, et al., 2008; Sampson & Raudenbush, 2004). The common practice of using census tracts or block groups as neighborhood units usually means that at least crude estimates of the size of the neighborhoods can in principle be derived by analyzing GIS files available from the Census Bureau, but there is still a serious problem with using census units to describe spatial scale. Their boundaries are designed to contain approximately equal numbers of residents rather than approximately equal land area (U.S. Census Bureau, 1994). Census tracts are intended to contain 2,500 to 8,000 residents, so they can vary dramatically in size because of varying population densities. For example, the average of size of a census tract in Chicago is about 0.67 km^2 (McMillen, 2003), but in a smaller and less densely populated city like Battle Creek, it is 7.14 km^2 . So, stating there is neighborhood-level variability between neighborhoods

defined in terms of census-based units is at best an indirect and vague statement about the spatial scale of the autocorrelation in that may exist in a particular outcome measure.

The neighborhood clusters in this study were constructed by combining census blocks in Battle Creek with either whole adjacent blocks or with the parts of adjacent blocks facing the core block across the street serving as the boundary between them (Van Egeren, et al., 2007). They were smaller than the block groups in Battle Creek partly because there simply were not enough block groups in the city to achieve recommended sample size at the neighborhood level for doing HLM analyses. Even though the strict hierarchy in the physical size of census units (blocks are combined into block groups, which are combined to form tracts; U.S. Census Bureau, 1994, 2002) suggests that the clusters used in this study were far smaller than the tract-based neighborhoods used in other studies, the size difference is not really as large as it might seem from just naming the census units used to construct them. The clusters in this study were indeed physically quite small, ranging from 0.026 to 0.472 km² and averaging 0.083 km², but they are much closer in physical size to the tracts found in a large city like Chicago than they are to the size of tracts in Battle Creek. Chicago tracts are still larger than the clusters used here, but that size difference is not as large as the one between the local block groups and the clusters.

Spatial scale in GSM. GSM analyses provide a convenient method for learning about the spatial scale of autocorrelation in the data. Information about this aspect of the sample data is extracted from the same variogram model used to estimate the individual- and neighborhood-level variance components. Depending on the shape of the variogram used in the model, there is usually some type of range parameter that indicates how far

spatial autocorrelation actually reaches (Banerjee, et al., 2004; Chaix, et al., 2006; Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian, et al., 2005; Diggle & Ribeiro, 2007). The second hypothesis (H2) for this study predicted that the autocorrelation found with GSM would extend far enough to reach across the borders between the clusters used in the HLM models, which were for the most part not directly adjacent to each other.

That hypothesis was strongly supported. The GSM models for the present study showed that the practical range of spatial autocorrelation detectable in these residents' perceptions of neighborhood problems is about 3.0 km. This is slightly longer than the median pairwise distance between the clusters used in the HLM analyses and it stretches nearly one third of the distance across the study region. A circle with a 3.0 km radius would enclose an area of 28.27 km^2 . Thus, spatial autocorrelation in this dataset persists over a far larger geographic area than the typical size of the neighborhood units used in the HLM models, which had an average area of just 0.08 km^2 .

In terms of the spatial scale of autocorrelation in perceived neighborhood problems, HLM and GSM provided rather different answers in this study. Using the cluster boundaries to group residents and characterize the structure of autocorrelation in these data without considering alternative possibilities would have ignored an important spatial pattern in the data for this study. Relying solely on the published advice that HLM researchers should use the smallest geographic units available to represent neighborhoods (Roosa, et al., 2003) appears to be potentially risky: It can easily lead to a dramatic mismatch between the spatial scale on which autocorrelation and spatial variability exist in the data and the size of the clusters used to represent neighborhood settings that

supposedly account for that autocorrelation. Knowing the spatial scale on which autocorrelation is detectable is useful because it provides an initial (though not definitive) clue about the potential size of the neighborhood areas that may be relevant to the outcomes among residents.

Modeling Autocorrelation: Spatial Versus Hierarchical Structure

Having established how much autocorrelation there was in residents' perceived problems and how far that autocorrelation reached, we still need to answer the second research question: Which method (HLM or GSM) is more effective at modeling the autocorrelation actually observed in data from neighborhood residents? The answer to this is important because the simplest possible HLM and GSM models (called the empty or null models) describe the essential structure of the data that we hope to explain by adding predictors in subsequent models. Starting from a poor or inaccurate description of the data does not position a researcher to draw the best possible scientific conclusions about the phenomena of interest. Answering this question is also important because evidence that the assumptions underlying a statistical model were violated should reduce our confidence in the accuracy of its results, especially when a competing model does not suffer from similar problems.

Four hypotheses were tested to answer this question (see Table 10). Overall, this study found that some model fit criteria consistently indicated that GSM models fit better than HLM, but others indicated very little difference. Examination of the residuals from both methods revealed serious violations of the HLM assumptions but not of the GSM assumptions.

Comparing model fit. The evidence with respect to H3 was somewhat mixed. The DIC values strongly and consistently favored the GSM models over the HLM models, indicating support for this hypothesis. However, there are two reasons to be cautious about the DIC values. First, the GSM models consistently had larger numbers of effective parameters (pD values) than the corresponding HLM models, so they were more complex models. Perhaps the difference in fit is a result of that additional complexity, rather than of a true difference in the structure of the underlying pattern of autocorrelation. While the DIC incorporates a term to penalize complex models for using additional effective parameters (Spiegelhalter, et al., 2002), the penalty term may not be a perfect solution to the issue of deciding whether the difference in model fit is entirely an artifact of the additional complexity. The present results are based on real data where the true population parameters and underlying autocorrelation structure are unknown, so while the present conclusions appear to be reasonable, controlled simulation studies will be necessary to investigate this issue more fully.

Second, comparisons of other model fit indices after controlling for resident characteristics indicated that the difference in fit between HLM and GSM was not so stark. While HLM explained about 3% more neighborhood-level variance, GSM explained about 0.4% more overall variance than HLM. These are surprisingly small differences given the incredibly large differences in DIC values observed when testing H3. The conflicting evidence offered by the various fit indices was unexpected, but simulation work has shown that the DIC can successfully distinguish between alternative covariance structures in longitudinal datasets (Barnett, Koper, Dobson, Schmiegelow, & Manseau, 2010). It is possible that in this study, the DIC is demonstrating high sensitivity

to the different covariance structures implied by HLM and GSM. That discrepancy between the DIC and some of the other model fit indices may be partly explained by the additional insights obtained from examining the HLM and GSM residuals to test hypotheses H4-H6.

Diagnostic analyses of HLM residuals. The diagnostic analyses of the HLM residuals supported H4 by showing that there was still a substantial amount of spatial autocorrelation remaining in both the individual- and neighborhood-level HLM residuals for this sample. In fact, the PSRs for the level 1 HLM residuals were substantially *higher* than the ICCs in the HLM models that produced those residuals. If a hierarchical structure fit the data better than a spatial structure, then the level 1 HLM residuals should represent only random error and they should not contain spatial patterns. But, adjusting the raw data to account for which neighborhood cluster each resident lived in did not leave behind only random error in those residuals: Residents who lived very close to each other had more similar level 1 HLM residuals than residents who lived farther apart. At the shortest distances, the autocorrelation in those residuals was actually higher than the autocorrelation in the raw data!

Although that spatial autocorrelation had a rather limited range (less than 20 m), this suggests that a hierarchical structure was not accounting for all the autocorrelation that exists within neighborhood clusters. Given the very short range of this phenomenon, one might suspect that using data from multiple people in the same household account for this and that the HLM model should actually be a three-level model (residents nested within households, which are then nested within neighborhood clusters). However, the sampling design for this study rules out that explanation: The sample included only one

resident per household. So why might there be such short range spatial autocorrelation in residents' perceptions even after adjusting for the overall effect of living in a specific neighborhood cluster?

One possibility is that residents and their next-door neighbors influence each other's perceptions through social interaction more than residents who live farther apart. Neighborly social contact provides ample opportunity for informal information sharing about neighborhood issues that might shape people's perceptions (Unger & Wandersman, 1985). Previous research has found that social ties in neighborhood networks decline with increasing distance and that residents report that high proportions of their closest friends live very close by (Greenbaum & Greenbaum, 1985). Furthermore, Skogan and Maxfield (1981) observed that strong neighborhood social ties led people to talk more to their neighbors about local crime. They also found that those who talked to their neighbors about crime were more likely to personally know local victims of crime and to fear crime more, presumably as a result of what they called "vicarious victimization". So, while the level 2 HLM residuals in this study may be capturing information about the average level of social connectedness and information sharing in each cluster, they may be leaving behind some residual spatial structure in the level 1 residuals that could be driven by residents exchanging more information with the neighbors who live closest to them than they do with other neighbors in the cluster.

The level 2 HLM residuals in this study represent centered estimates of neighborhood-level means for perceived neighborhood problems. Additional support for H4 was provided by evidence that the neighborhood-level HLM residuals in this sample also contained substantial spatial autocorrelation that decreased with increasing distance

between clusters. Thus, there was strong evidence that the standard HLM assumption of independent neighborhood-level residuals had been violated in this study. This is completely consistent with what one would expect given the long range spatial autocorrelation detected by the initial GSM models.

The present study is the first to report such an analysis of the residuals from an HLM model that sought to predict residents' perceptions of neighborhood problems. That makes it difficult to definitively determine whether this finding is generalizable or more specific to this outcome, this sample, this study region, or some combination of these factors. However, studies of other outcomes have also found evidence of spatial autocorrelation in neighborhood-level residuals obtained from HLM analyses of Chicago residents' informal neighboring activities and their participation in neighborhood organizing activities (Swaroop & Morenoff, 2006), French residents' usage of specialist physicians (Chaix, Merlo, & Chauvin, 2005), and substance abuse disorders among residents of a Swedish city (Chaix, Merlo, Subramanian, et al., 2005). These other studies suggest that the present findings may not be unique, but replication of this finding is certainly advisable before drawing strong conclusions about generalizability.

Again, a question arises about what could cause this neighborhood-level spatial autocorrelation in perceived neighborhood problems. Just as the distance-decay pattern in neighborhood social network connections (Greenbaum & Greenbaum, 1985) could combine with information sharing between neighbors (Skogan & Maxfield, 1981; Unger & Wandersman, 1985) to explain such a pattern on a smaller scale within neighborhoods, the same thing can happen across the boundaries between clusters. Another possibility is that adopting researcher-defined cluster boundaries effectively introduces measurement

error because those boundaries do not coincide with how residents would define their neighborhoods (Swaroop & Morenoff, 2006). Finally, to the extent that crime, NSES, and other neighborhood-level characteristics do in fact predict residents' perceptions, then the spatial distribution of those contextual characteristics and the spatial arrangement of the clusters themselves should also play a role in inducing spatial autocorrelation in those perceptions.

The key finding here is that, at least for this sample, the independence assumption for the HLM residuals was violated at both levels of analysis. Thus, the DIC values favoring the GSM models over the HLM models may be picking up on a mis-specified error structure in the HLM models. That might be a finding specific to the present dataset, but even if that is the case, it remains important to the current study because relying on a mis-specified model is not a sound statistical practice when we have an alternative model that may be better suited to analyzing the data at hand. The conclusion that the DIC is discriminating between the two models on the basis of legitimate differences in how well they match the underlying covariance structure in the data can be bolstered by showing that the GSM model assumptions were not similarly violated, so next the discussion turns to interpreting the diagnostic analyses of the GSM residuals.

Diagnostic analyses of GSM residuals. The GSM models did a much better job than HLM of producing level 1 residuals that were purged of autocorrelation, providing partial support for H5. It is only partial support because H5 predicted that there would be no evidence at all of remaining autocorrelation in those residuals. Instead, there was a significant, but small amount of hierarchically structured autocorrelation remaining in the individual-level GSM residuals. However, the ICC for those residuals was more than 20

percentage points lower than the PSR from the models that produced them in the first place. So, we can conclude that the variograms in the GSM models were accounting for most of the autocorrelation and did an excellent job of separating individual-level and neighborhood-level variance.

In H6, it was predicted that the neighborhood-level residuals from the GSM models would appear to contain hierarchical autocorrelation, but at a lower level than the amount of spatial autocorrelation observed in the GSM model that produced the residuals. HLM was unable to detect much of any individual-level, within-cluster variance in those GSM neighborhood-level residuals (this variance component was very close to zero). As a result, even though HLM actually detected a smaller amount of neighborhood-level variance than was present in the original GSM model, the ICC for the residuals was very close to 1 (indicating almost perfect hierarchical autocorrelation). This means that a researcher relying only on HLM to analyze the present data could easily, but mistakenly, conclude that the data conform to the hierarchical structure assumed in HLM, when in fact a distance-decay pattern of spatial autocorrelation provides a more accurate description of the data.

In hindsight, this should not be surprising though because the observations within a cluster are generally closer to each other than they are to observations in most of the other clusters. Thus having a dataset that closely matched the pattern modeled by the variogram built into the GSM approach virtually guaranteed this result. Still, H6 was only partly supported because the initial prediction failed to anticipate that looking for hierarchical structure in a set of GSM residuals already purged of individual-level variance would increase rather than decrease the apparent levels of autocorrelation.

Summary. Overall, the evidence from testing H3 – H6 suggest that the GSM models do more accurately describe the data than the HLM models, both before and after controlling for neighborhood composition. Because the GSM residuals behaved largely as expected, but the HLM residuals did not, we can conclude that the answer to the second research question is that GSM provided a better model for the autocorrelation in these data than HLM. Thus, the standard HLM models are mis-specified because they ignore the information conveyed by the spatial arrangement of the residents and neighborhood units in this sample: They focus too much on place and neglect the role of the space in which those people and places are embedded.

While this result may be specific to the present data, there are examples in the literature of other studies that have also observed spatial patterns remaining after HLM analyses (Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian, et al., 2005; Swaroop & Morenoff, 2006). At a minimum, researchers using HLM to examine neighborhood effects should test for residual spatial autocorrelation. If they find it, they should strongly consider at least modifying the default HLM assumption that neighborhoods are independent (Beard, 2008; W. Browne & Goldstein, in press). Failure to do so incurs a risk of inflated Type I error that parallels the risk associated with using traditional OLS regression models instead of HLM models when the data are hierarchically structured.

Because adding spatial autocorrelation to an HLM model is not supported in many HLM software packages, doing so may require switching to software that supports a flexible, fully Bayesian approach to estimating HLM models. WinBUGS (Lunn, et al., 2000) is one option because it supports adding a conditional autoregressive (CAR)

structure to the neighborhood-level part of an HLM model (Fagg, et al., 2008; Thomas, et al., 2004). However, moving from a standard HLM model to a CAR HLM model is not the only option. Researchers observing residual spatial autocorrelation in their HLM models should also consider analyzing their data with a model such as GSM that is explicitly designed to handle spatial patterns of autocorrelation. This may only be feasible if they also have (or can obtain) precise location information for the residents (e.g., addresses that can be geocoded to point-level spatial coordinates).

The discussion so far has focused on interpreting the results of comparing HLM and GSM with respect to modeling neighborhood-level variance and autocorrelation, but has not yet addressed what we can learn from comparing them with respect to testing the effects of neighborhood-level predictors on resident outcomes. Exploring the former issue was a necessary precursor, but the latter part of this study has more interesting implications for neighborhood research. Hence, now the discussion moves on to interpret the findings related to testing crime and NSES effects on residents' perceptions of neighborhood problems.

Testing Crime and NSES Effects

One of the primary aims in this study was to assess whether GSM provides a valuable alternative to HLM for testing hypotheses about neighborhood effects. More specifically, this study examined whether and to what extent two specific ecological characteristics of the neighborhood context (neighborhood crime and NSES) explain the levels of neighborhood problems reported by the residents in the sample. Thus, the study examines a multilevel phenomenon with two levels of analysis: residents and neighborhoods. We must conceptualize and operationally define neighborhoods before

we can measure crime and NSES for those neighborhoods (Linney, 2000). The crux of this study is that how we do that has important consequences for what we can learn about the phenomena we are studying.

Conceptualizing and defining neighborhoods. The introduction and literature review above noted that HLM and GSM draw on different conceptualizations of geographic space and neighborhoods as places within that space. HLM relies on a discontinuous view of geographic space that treats neighborhoods as ecological settings that occupy geographic places with fixed, non-overlapping boundaries and possess contextual characteristics reflecting local conditions inside those boundaries. Most neighborhood studies using HLM adopt census tracts or block groups to operationalize neighborhoods (Leventhal & Brooks-Gunn, 2000; Roosa et al., 2003; Sampson et al., 2002), thereby inheriting a convenient, well-known, and hierarchically organized boundary system that is thoroughly grounded in that discontinuous view of space (U.S. Census Bureau, 1994, 2002).

There are several potential problems with the discontinuous conceptualization of space and neighborhoods. First, the specific boundaries chosen to define neighborhoods can affect the values of contextual characteristics associated with them and the statistical results obtained from analyses—this is the modifiable areal unit problem, or MAUP (Bailey & Gatrell, 1995; Coulton, et al., 2004; Downey, 2006; Mowbray, et al., 2007). Second, it fosters a modeling approach that ignores spatial proximity between residents and proximity or contiguity between the neighborhood units (Downey, 2006; Mowbray, et al., 2007). Third, it makes a strong assumption that the selected boundaries are meaningful to (and agreed upon) by residents, and are equally appropriate for measuring

all neighborhood-level characteristics. Fourth, it ignores potential spatial variation in contextual conditions that may occur within the boundaries of the neighborhood unit (Roosa, et al., 2003). Finally, it has little flexibility to address questions about the spatial scale of the phenomena being studied.

GSM relies on a continuous view of geographic space that attends to spatial proximity and spatial relationships between places (Chaix, et al., 2006; Chaix, Merlo, & Chauvin, 2005; Chaix, Merlo, Subramanian, et al., 2005; Downey, 2006). Adopting Galster's (2001) conceptualization of neighborhoods as "bundles of spatially-based attributes associated with clusters of residences" (p. 2112) is consistent with this view of space and emphasizes that neighborhoods are places that can be described as ecological settings that are tied to geographic locations and possesses contextual characteristics reflecting local conditions in the geographic areas surrounding those locations.

This conceptualization allows GSM to offer us more flexibility than HLM in how we define neighborhoods for measuring constructs like crime and NSES because a neighborhood no longer needs to have a single, fixed, and unambiguous geographic boundary (Galster, 2001; Guo & Bhat, 2007). For example, we can use fixed boundaries like those required in HLM studies, but we can also allow neighborhoods to partially overlap, use different boundaries for measuring crime than we use to measure NSES, or easily change the size of a neighborhood.

Allowing neighborhoods to partially overlap is consistent with research showing that residents often disagree about neighborhoods boundaries (Coulton, et al., 2010; Coulton, et al., 2004; Coulton, et al., 2001) and that the boundaries of places may really be rather fuzzy and vague (Montello, et al., 2003). It is also consistent with the

observation that most residents describe themselves as living in the center of their own neighborhoods (Coulton, et al., 2001). For some of the GSM models in this study, neighborhoods were defined with circular buffers centered on residents' homes. This is a simple method for creating "sliding neighborhoods" (Guo & Bhat, 2007) or "bespoke neighborhoods" (Galster, 2008) that may be more closely aligned with these prior findings from the literature and address some of the problems with how neighborhoods are defined for use in HLM (Guo & Bhat, 2007; Kruger, 2008; Meersman, 2005).

Very little previous research has explored the consequences of switching from a discontinuous conceptualization of geographic space and neighborhoods to a continuous one. This study is a step toward filling that gap in the literature.

Comparing HLM and GSM. The HLM models in this study always measured crime and NSES within fixed neighborhood cluster boundaries, but two types of boundaries were used in the GSM models. Cluster-based GSM models measured crime and NSES within the same boundaries used by the HLM models, while buffer-based GSM models measured them in circular buffers. This enabled the study to pursue the third research question for this study, which asked how GSM estimates of contextual effects and model fit compare to HLM estimates. This is effectively also a question about whether one conceptualization and operational definition of neighborhoods works better than the other in practice.

The corresponding hypothesis (H7) predicted that both model fit and the size of the crime and NSES effects on perceived neighborhood problems would fall into a rank-order with buffer-based GSM models performing best, followed by cluster-based GSM models, and then HLM models. The three-way comparison used to test H7 was crucial

for isolating whether the differences between the two methods were driven primarily by how autocorrelation was modeled or by how neighborhoods were defined for measurement purposes. In addition, the study also examined whether switching from a standard HLM model to a CAR HLM model impacted the results.

The study also sought to answer a fourth research question, which asked how the geographical scales on which different contextual factors operate (as estimated by GSM) compare to each other and to the size of the neighborhood units used in HLM. The prediction (H8) was that the optimal buffer sizes for measuring crime and NSES would differ from one another and from the average size of the neighborhood areas used in the HLM analysis.

Current findings. Even though the evidence was mixed with respect to support for H7, this study found that the GSM models produced more credible analyses of the present data than the HLM models because the latter depended on assumptions that were violated while the former did not. Overall, the analyses showed that the circular buffers used in some of the GSM models provided better operational neighborhood definitions for measuring crime and NSES than the fixed cluster boundaries. In terms of predicting residents' perceptions of neighborhood problems in these data, the standard HLM models overestimated the effect of NSES, but underestimated the effect of crime. The CAR HLM model appeared to mostly correct the overestimation of the NSES effect, but failed to correct for the underestimation of the crime effect.

Comparing model fit. Once again, the DIC strongly and consistently indicated that the GSM models fit the data better than the HLM models (including the CAR HLM model). For most other indices of model fit, the findings were a bit more nuanced. GSM

based on using the cluster boundaries to measure crime and NSES had either similar or somewhat worse performance than either HLM or CAR HLM models. Correcting for the mis-specified error structure in the HLM models by using cluster-based GSM models instead appears to provide a more conservative assessment of model performance.

Meanwhile, buffer-based GSM generally performed better than standard HLM as long as crime was in the model (alone or together with NSES) and it performed about the same as CAR HLM. This suggests that crime may have been a more potent influence on residents' perceptions than NSES. It may also be a sign that the differences in performance between GSM and HLM are likely to be greater when there is a large disparity in the sizes of the neighborhood clusters used in the HLM models and the optimal buffers used in the GSM models, as there was with crime in this study. Further research (perhaps based on controlled simulations) could either support or refute that possibility.

Spatial scale of crime and NSES. Because the results with respect to H8 are crucial to the interpretation of the findings from testing H7, they are summarized first. For this dataset, H8 was partially supported because GSM modeling revealed that the optimal spatial scale for measuring crime (1.1 km radius) was far larger than the clusters used in the HLM. However, it also revealed that the optimal spatial scale for measuring NSES was approximately the same size as the clusters (a 0.2 km radius). A theoretical interpretation of why these two contextual characteristics seem to be operating on such disparate spatial scales is integrated into the next section.

Contextual effect of NSES. At least for the present dataset, HLM models overestimated the size and statistical significance of the effect of a cluster-based measure

of NSES on residents' perceptions of neighborhood problems because of the violated assumptions about the independence of the residuals. Applying GSM instead resolved the mis-specified error structure and showed that while the cluster-based NSES measure did not affect residents' perceptions in these data, the buffer-based NSES with a 0.2 km radius did affect those perceptions (just not as strongly as indicated in the HLM analysis). Furthermore, switching from a standard HLM model to a CAR HLM model reduced the discrepancy between the HLM and buffer-based GSM estimates of the NSES coefficient. This suggests that the mis-specified neighborhood-level error structure in the standard HLM models may have artificially inflated the NSES effect in the standard HLM analyses.

That the two methods led to different conclusions about the effect of a cluster-based measure of NSES in this study has important implications for our theoretical understanding of what shapes residents' perceptions of their neighborhoods. They are disagreeing about the importance of the stigma associated with poor neighborhoods as a mechanism linking neighborhoods to resident perceptions of neighborhood problems. With GSM, we were able to directly test whether the fixed cluster boundaries or buffers better approximate the neighborhood settings that inform residents' perceptions. The answer here appears to be that the buffers work better for this purpose, but we could not have even done such an analysis with HLM.

What could explain why these buffers appeared to be more psychologically meaningful neighborhood areas than the clusters when it comes to measuring NSES? The optimal buffers for NSES were similar in size to clusters (though the latter varied in size somewhat), so it is unlikely that this is purely a matter of measuring NSES on the wrong

spatial scale. Instead, we should consider that this study measured NSES in terms of median residential property values. How could residents even perceive that neighborhood characteristic and why would it influence their perceptions?

Property values are directly related to the physical quality of the housing, which likely serves as a symbolic cue (Unger & Wandersman, 1985) to residents about the socioeconomic status of their neighbors because families tend to move into better housing when their incomes increase (Schill & Wachter, 1995). Presumably, residents are very familiar with the quality of the housing immediately surrounding their own home because they see it daily. They may also be more familiar with the housing occupied by the people with whom they interact frequently than they are with the housing occupied by people with whom they socialize less often.

If so, then the present results make sense in light of the fact that residents' social network connections and social travel appear to decline with increasing distance (Greenbaum, 1982; Greenbaum & Greenbaum, 1985; Stutz, 1973; Wheeler & Stutz, 1971). Using the cluster boundaries implicitly assumes that only the housing within the cluster is relevant (and that it is all equally relevant), but residents living on the edges of the cluster may well be interacting with people just outside the border more often than they interact with people on the far side of their cluster. There may also simply be some important spatial variability in the local median housing values within some of the clusters. The buffers might therefore be better capturing the group of people that a resident interacts with and use to form their impressions about NSES, which subsequently prime residents who live in poorer neighborhoods to perceive greater problems because of the stigma that has accumulated as a result of the historical

association between poverty and neighborhood problems (Franzini, et al., 2008; Sampson & Raudenbush, 2004).

Contextual effect of crime. Unlike with NSES, HLM severely underestimated the strength of crime's effect on residents' perceptions in this study. Indeed, the cluster-based GSM models also severely underestimated the crime effect on perceived neighborhood problems for the same reason. Buffer-based GSM models showed that crime within 1.1 km of residents' homes had a much stronger effect on perceived neighborhood problems than the cluster-based crime effect observed with the HLM and cluster-based GSM models. Switching to the CAR HLM model did not substantially change the HLM estimate of the crime effect, so this difference in coefficients cannot be explained by the difference in how autocorrelation was modeled. Consequently, we can conclude that the clusters used in the HLM analysis were simply far too small compared to the actual spatial scale on which crime mattered to the residents. This of course raises the question of why the spatial scale for crime is so large, especially in comparison to the spatial scale for NSES.

Again, we need to consider the nature of the phenomenon to better understand this. Crime, especially the kind of violent crime represented by the measure used here, is an extreme form of social disorder (Sampson & Raudenbush, 1999). Its presence in the local neighborhood is highly salient to residents because it is a potential threat to their well-being. People often fear being victimized by criminals, so they are motivated to protect themselves by avoiding places and situations that would expose them to crime (Gates & Rohe, 1987).

But, to successfully avoid exposure to crime, residents need to know where it has been occurring. Residents use multiple sources of information to learn about local crime, including media reports, gossip from friends and neighbors, plus their own experience of crime and direct observation of the local environment (Sampson & Raudenbush, 1999; Skogan & Maxfield, 1981). Furthermore, the theory underlying behavioral geography suggests that as people go about their daily life, they construct an “awareness space” that expands outward beyond the area in which they engage in their activities (i.e., their “activity space”) to encompass adjacent and surrounding areas as well (McCord, et al., 2007). It makes sense that this awareness space would be quite expansive because residents frequently travel outside their own neighborhood boundaries to shop, go to work, visit friends, and so on (Sastry, et al., 2002).

The larger spatial scale associated with the buffer-based crime effect in the GSM models suggests that, with respect to crime, residents are attending to broader neighborhoods than they use to inform their assessment of NSES. This is consistent with the idea that residents can and do think about their neighborhoods at multiple spatial scales (Galster, 2001; Kearns & Parkinson, 2001; Suttles, 1972). Kearns and Parkinson (2001) suggested that these different spatial scales of neighborhood serve different functions for residents, so it should not be surprising that different neighborhood characteristics are more relevant at one scale than at another.

Implications for Defining Neighborhoods

O’Campo (2003) suggested that neighborhood researchers might need to try using multiple operational definitions of neighborhoods within the same study. This study did that in two different ways: it varied whether each neighborhood-level characteristic was

measured within discrete cluster boundaries or buffer-based boundaries, and it allowed the two neighborhood characteristics to be measured within different size buffers so that multiple neighborhood definitions were in use in the context of a single model. The best results were obtained when crime and NSES were measured in buffers that differed dramatically in size. Although HLM can in principle handle using multiple sizes of neighborhood units (by using more than two levels of analysis), this is rarely done in practice. It appears to be much easier to do this with GSM.

So, what can we learn about how to think about and define neighborhoods from this study? One implication of the findings from this study is that researchers may need to pay greater attention to how they define neighborhood boundaries for measuring neighborhood-level characteristics. These results suggest that there are at least some circumstances when it is useful to discard the constraints associated with taking a discontinuous view of geographic space and embrace instead the ideas that neighborhoods can sometimes overlap and that the most relevant neighborhood boundaries for measuring contextual characteristics may depend on what you want to measure (Galster, 2001; Kruger, 2008).

Doing that enabled this study to demonstrate that the neighborhood area most meaningful for measuring the effect of NSES on these residents was not identical to the area most meaningful for measuring the effect of crime. While NSES was most influential when measured in a small area surrounding a resident's home, crime had to be measured over a much larger area. This supports the value of adopting a more flexible conceptualization of neighborhoods that does not demand that the same boundaries be used to measure all neighborhood characteristics (Galster, 2001).

Future researchers may be able to use the optimal buffer sizes for crime and NSES observed in this study to inform what range of spatial scales might be worth examining in their studies, but they should exercise caution in doing so. It seems reasonable to expect that—in relative terms—NSES should perhaps be measured over a smaller spatial area than crime, but the precise buffer sizes that worked in this study may not work as well for other study regions. Researchers may benefit more from adopting the strategy that was used in this study to determine the buffer sizes than from directly trying to use the optimal buffer sizes reported here.

Implications for Community Interventions

The findings in the present study are not just about comparing two statistical methods as an abstract, academic exercise. Comparing HLM and GSM as tools for studying neighborhood effects is important because the differences in their performance could have practical implications for how research findings can inform community intervention efforts. To explore those implications for the design of a hypothetical community intervention set in the study region, we can examine the coefficients from the models and translate them into estimates of how much change in either crime or NSES would be required to achieve some substantively important amount of change in the mean level of perceived neighborhood problems.

The outcome in this study has a standard deviation of 1.48 points (on a scale ranging from 1 to 6). Let us assume the intervention team had determined that reducing the mean level of perceived problems in a particular neighborhood by half a standard deviation ($0.5 * 1.48 = 0.74$ points) would produce some desired benefit for the residents. If we want to know how much impact this hypothetical intervention would need to have

on a contextual factor to achieve that goal, we can examine Table 11. This table uses the model coefficients to predict how much crime and NSES would need to change to produce such a shift in residents' perceptions, depending on how the neighborhoods were defined for measurement purposes and whether one uses the results from the HLM, CAR HLM, or GSM models from this study.

Table 11: Amount of change required on each predictor to reduce mean perceived problems by half a standard deviation (0.74 Points).

Method	Model	Neighborhood Definition	Coef. ¹	Change Goal	Target Area ² (km ²)	No. Crimes To Prevent
Crime effect						
HLM	5	Clusters	0.018*	-411 crimes/km ²	0.083	34
CAR HLM	6	Clusters	0.014*	-529 crimes/km ²	0.083	44
GSM	5	Clusters	0.013*	-569 crimes/km ²	0.083	47
GSM	58	1.1 km buffer	0.200*	-39 crimes/km ²	3.838	142
NSES effect						
HLM	5	Clusters	-0.022*	\$ 33,636	0.083	
CAR HLM	6	Clusters	-0.018*	\$ 41,111	0.083	
GSM	5	Clusters	-0.007	\$ 105,714	0.083	
GSM	58	0.2 km buffer	-0.014*	\$ 52,857	0.126	

Note: The outcome variable had $SD = 1.48$, so $0.5 * SD = 0.74$ points. The change goal values represent how much change on a given predictor would be required to observe a 0.5 SD decrease in mean perceived neighborhood problems based on the estimated model coefficients. Coef = Coefficient.

¹ In the models, crime was measured in units of 10 crimes/km² and NSES was measured in \$1,000 units.

* $p < .05$.

Crime effect. Table 11 shows that if we used the results of HLM Model 5 to plan a crime prevention effort, we might set a target of reducing crime density by 411 crimes/km². Assuming we are working in a neighborhood cluster of average size (0.083 km²), this model would tell us that we need to prevent 34 crimes inside that cluster boundary (over the course of a year) to achieve the intended impact on residents'

perceptions. The cluster-based GSM model yields a different answer, suggesting instead that we would need to reduce crime density by 569 crimes/km². In a neighborhood of the same size, we would need to actually prevent 47 crimes per year. Relying on the standard HLM model therefore would put us at risk of setting the prevention target too low, which could then lead to an under-powered intervention that is less likely to achieve the intended outcome.

While it is useful to see that the CAR HLM model mostly corrects the underestimation of the crime effect observed in HLM Model 5, Table 10 shows that the findings of GSM Model 58 tell a vastly different story about what would need to happen in the prevention effort to achieve the desired outcome. First, the geographic scope of the prevention effort would need to be dramatically expanded because crimes occurring far outside the borders of a particular cluster still influence residents' perceptions. Although the target decrease in crime density is much smaller (39 crimes/km²), that decrease would have to happen over a far larger area (3.838 km²) because the optimal buffer size is so much larger than the clusters. As a result, a total of 142 crimes would need to be prevented within this larger area to achieve the intended effect.

NSES effect. NSES was measured in terms of median housing value, so an intervention would generally require engaging in home improvement efforts that would raise property values. An intervention relying on the findings from HLM Model 5 would see that the model predicts that increasing the median housing value within a neighborhood cluster by \$33,636 would reduce the mean of perceived neighborhood

problems by half a standard deviation (see Table 11). Such an intervention would clearly be a large and expensive undertaking.

Unfortunately, the results from GSM Model 5 suggest that even if that goal were achieved, it would not have the desired effect because the latter model found that the cluster-based measure of NSES did not significantly influence perceived problems. According to this GSM Model 5, it would take an increase in value more than three times that size (\$105,714) inside the cluster to achieve the intended improvement in residents' perceptions. However, the buffer-based GSM Model 58 suggests that the situation is not quite so grim: Increasing the median housing value by \$52,857 over an area about 1.5 times the size of the typical cluster would achieve the intended improvement in residents' perceptions. So, compared to the buffer-based GSM model, a prevention goal set according to the HLM model results would both aim for too small a change in NSES and target a geographic area that would be slightly too small. While relying on the CAR HLM model instead would lead to adopting a change target (\$41,111) somewhat closer to that of the buffer-based GSM model, it would still lead the intervention to aim for too small a change over too small an area to really achieve the intended effect.

Summary. The way we conceptualize and operationally define neighborhood boundaries for the purpose of measuring contextual characteristics like crime and NSES would have important consequences if we wished to use the present research findings to inform the planning and execution of a community intervention designed to change residents' perceptions of neighborhood problems. The HLM and GSM models reported here produced rather different pictures of what it would take to shift those perceptions by the same amount.

With this sample, the buffer-based GSM models provided more credible results than the HLM models. At least for this study region, relying on the HLM results would put us at risk of setting intervention goals that are too low to have the desired impact on outcomes. The essential message emerging from this comparison is that how we define neighborhoods and how we test the effects of neighborhood characteristics could matter a great deal when we go to apply those findings.

Feasibility of Applying GSM in Community Psychology Research

Although GSM did generate interesting and important findings that differed from the HLM findings in this study, implementing this method was a significant challenge on several fronts. First, there was a substantial amount of sophisticated data management work involved in linking various GIS shapefiles to prepare the dataset. Second, it was necessary to spend time learning about the Bayesian approach to statistics and statistical inference because the software used (Finley, et al., 2007, 2009) relied on a Bayesian modeling framework. Bayesian models are rarely used in our discipline, so there were few examples that could serve as models for some of the methodological choices involved (e.g., choosing appropriate prior distributions).

Third, applying GSM to a dataset with large sample of residents ($N = 1,049$ in this study) was especially computationally demanding. On average, running three MCMC chains for a single GSM model consumed over 6 days' worth of total computing time on a server or a powerful desktop computer. That can be done in two calendar days if the chains are run on different computers or on a server with multiple processors. The series of GSM models presented above cumulatively consumed over 120 days of computing time on a fairly new and powerful network server. This may pose a particular challenge

for most community psychologists, who may not have easy access to computers powerful enough to make running GSM models on large datasets feasible, especially when the MCMC chains must be run for many thousands of iterations.

Although the present study did not take advantage of one, some universities have established high-performance computing centers. It is possible that the computer hardware and software infrastructure offered by such a facility could have decreased the amount of time it would take to run the analyses. The tradeoff associated with using the parallel processing capabilities of such facilities is that doing so often involves more complex and specialized computer programming. However, some such centers offer technical assistance or consulting services that may make using their facilities easier.

To make applying GSM more feasible for community psychologists, the best strategy may be to engage in interdisciplinary collaboration (Maton, et al., 2006) with researchers who have expertise in working with GIS tools, spatial data analysis techniques, and the Bayesian modeling framework. Such individuals will probably be found in academic disciplines such as geography, ecology, natural resources, and statistics.

Limitations

As the literature review illustrated, taking a close look at how geographic space and neighborhoods are conceptualized raises a host of issues, many of which are difficult to disentangle. This study was only able to address some of them, and even there it was only investigating a single outcome measure in a single sample. This restricts the generalizability of the results in several ways.

This study cannot tell us whether applying both HLM and GSM to other outcomes will generate similar results. Perceived neighborhood problems was selected as the target outcome for this study specifically because prior research had found evidence of spatial autocorrelation in adolescents' perceptions of neighborhood disorder (Bass & Lambert, 2004). It is possible that this outcome is the exception rather than the rule and that other constructs will not demonstrate such a clear pattern of spatial autocorrelation that decays as a function of distance. Because so few social science outcomes have yet been examined with geostatistical methods of any kind, the literature is currently too sparse to provide community psychologists with a reliable guide to which outcomes might be best analyzed with HLM and which would be better analyzed with GSM.

Another limit on the generalizability of these results is that most neighborhood research is conducted with residents of much larger cities (Franzini, et al., 2008; Sampson & Raudenbush, 2004; Sampson, et al., 1997), but this study focused on a sample from a single, small city. Battle Creek is not the same kind of place as a large city like Chicago or Baltimore. The experience of neighborhood life in a small city may simply be quite different than it is in larger, more densely populated urban environments. Replicating this study with data from additional study sites (i.e., different cities) would provide insight into the generalizability of its findings to other samples and geographic areas.

This study serves more as a proof-of-concept that GSM can outperform HLM under certain circumstances than as a thorough assessment of whether GSM will reliably do so for a wide range of outcomes and samples. There is still a great deal of work to do to establish the conditions under which each of the two statistical methods works best.

One inherent limitation of GSM relative to HLM is that it cannot easily be used to simultaneously analyze data from multiple cities. For example, Coulton et al. (2004) pooled data from neighborhoods in ten different cities that are scattered across the US and analyzed the combined data with HLM. This is not really feasible with GSM. The vast distances between cities would dramatically skew the distribution of pairwise distances between residents. Spatial autocorrelation is usually considered a small-scale phenomenon relative to the size of the study region. Any spatial autocorrelation between people living in separate cities could not reasonably be construed as reflecting neighborhood effects (it might be more properly be considered to reflect regional effects). Overall, future researchers should probably look at GSM as a technique better suited for studying neighborhoods within a single city.

Another issue limitation of GSM relative to HLM is that it requires precise location information for every resident. Where privacy issues or practical concerns make it difficult to obtain precise location data for each observation, HLM may be the better methodological choice because it does not matter precisely where in a neighborhood each resident lives: The only thing that matters is that the resident lives somewhere in that neighborhood.

Directions for Future Research

There are a couple promising directions for future research that could build upon the present study. One option is to pursue formal simulation studies to more rigorously compare HLM and GSM. Another option is to explore alternative methods for defining buffers in GSM models.

Use simulation studies to compare HLM and GSM. Well-designed simulation studies have exceptional value for comparing different statistical methods. Creating datasets with known parameters and then analyzing them with both HLM and GSM would allow us to draw strong conclusions about the conditions under which each method performs well. A variety of factors could be manipulated in such simulations, such as: the true underlying structure in the data; the level of autocorrelation; the number, shapes, and spatial arrangement of neighborhood clusters; the numbers and spatial arrangements of residents within those clusters; the spatial distributions of the predictors and so on. This will certainly prove a challenging task. In the meantime, applying GSM techniques to other existing datasets may be valuable because it will help us better understand the method's capabilities and limitations.

Explore alternative methods for defining buffers. While the present study used circular buffers as a simple alternative to fixed cluster boundaries, doing that was only one of many possible options for defining sliding neighborhoods (Guo & Bhat, 2007). Having demonstrated a proof-of-concept that buffers have the potential to outperform fixed neighborhood boundaries, it is worth asking whether there are yet better ways to represent neighborhoods. Circular buffers are a rather crude approach to defining sliding neighborhood boundaries because they implement a very simple rule to determine where to place the edges of the buffer: From a target location, they simply travel outward along a straight line in every direction to enclose the area within a specific distance threshold.

More sophisticated methods for creating buffers have been explored by researchers in other academic disciplines. For example, a "network band" approach defines the buffer as the area within a specific *travel distance along the street network*

from the resident's home (this replaces a simple straight-line distance threshold with one based on the configuration of the local streets) (Guo & Bhat, 2007). Network band buffers can be asymmetrical if the layout of the street network facilitates travel in some directions more readily than others.

Another option might be to actually collect resident-defined boundaries, which will almost certainly vary from resident to resident (Coulton, et al., 2001), and use those in GSM models instead of algorithmically-defined buffers. Although labor intensive, this would take residents' word about the neighborhood area that matters to them at face value. It would be quite interesting to explicitly test whether this would produce better statistical results than other possible approaches to operationalizing neighborhoods.

Finally, Lee's observation that individual-level characteristics are related to the size of residents' self-reported neighborhoods (Lee, 2001) suggests that it may be interesting to pursue testing whether the optimal buffer size for specific neighborhood-level characteristics is moderated by individual-level characteristics. Doing that was outside the scope of this study, but it could be a fruitful direction for future research.

Conclusion

GSM proved to be a valuable alternative to HLM in this study. This new method allowed the study to precisely quantify the distance over which autocorrelation in residents' perceptions of neighborhood problems persisted (3 km). Using it was crucial in establishing that the cluster boundaries selected for use in the original data collection effort provided neither the best method of grouping residents to detect and model spatial variability in perceived neighborhood problems, nor the best operational definition of neighborhoods for the purposes of measuring crime and NSES.

Residents appear to have been influenced by the residential property values associated with housing located within about 0.2 km of their own homes, but it would take fairly radical changes in median housing values to effect substantial changes in perceived neighborhood problems. Meanwhile, residents's perceptions were quite sensitive to the spatial density of violent crime occurring within 1.1 km of their homes. The amount of change in crime that would be required to decrease perceived neighborhood problems appears to be quite feasible: Preventing 147 crimes over the course of one year seems like an achievable goal.

Perhaps the single strongest reason to consider further use of GSM in community psychology is that it allows us to question the conventional assumption that census tracts and other arbitrary neighborhood units are good proxies for meaningful neighborhoods and test new ways of representing neighborhoods that may be more closely aligned with what we know about how residents think about their own neighborhoods.

APPENDIX

Table 12: Parameter estimates and model fit statistics for GSM Models 15-17 and 31-33.

Parameter	Model 15 (1.0 km crime buffer)			Model 16 (1.1 km crime buffer)		
	P. Mean	95% CI	t	P. Mean	95% CI	t
L2 fixed effects						
Intercept	3.745	[3.489, 3.949]	32.28	3.754	[3.517, 3.961]	32.52
Crime (buffer)	0.227	[0.148, 0.293]	6.22	0.263	[0.187, 0.337]	6.98
NSES (buffer)						
L1 fixed effects						
Age(years)						
36-55	0.145	[-0.045, 0.337]	1.49	0.144	[-0.045, 0.336]	1.48
≥ 56	0.168	[-0.089, 0.425]	1.28	0.176	[-0.081, 0.433]	1.34
Female	0.120	[-0.062, 0.303]	1.29	0.123	[-0.061, 0.308]	1.30
Race						
Black	-0.252	[-0.471, -0.039]	-2.29	-0.251	[-0.470, -0.034]	-2.26
Hispanic	-0.030	[-0.427, 0.364]	-0.15	-0.040	[-0.434, 0.353]	-0.20
Other	-0.078	[-0.658, 0.501]	-0.26	-0.075	[-0.647, 0.493]	-0.25
Marital status						
Married	0.033	[-0.183, 0.249]	0.29	0.034	[-0.184, 0.255]	0.30
Divorced	0.029	[-0.217, 0.280]	0.23	0.020	[-0.228, 0.267]	0.16
Widowed	-0.169	[-0.507, 0.166]	-0.98	-0.163	[-0.501, 0.173]	-0.96
Education						
< High school	-0.017	[-0.237, 0.201]	-0.15	-0.014	[-0.232, 0.202]	-0.13
Undergraduate	0.111	[-0.107, 0.332]	0.99	0.115	[-0.103, 0.335]	1.04
Postgraduate	0.270	[-0.295, 0.834]	0.94	0.285	[-0.276, 0.841]	1.00
Employed	-0.081	[-0.262, 0.100]	-0.88	-0.082	[-0.263, 0.098]	-0.89
Income (\$1,000s)						
< 15	0.160	[-0.139, 0.457]	1.05	0.163	[-0.135, 0.458]	1.07
15-25	-0.008	[-0.296, 0.274]	-0.06	-0.007	[-0.292, 0.275]	-0.05
25-45	-0.014	[-0.279, 0.253]	-0.11	-0.008	[-0.273, 0.252]	-0.06
Home owner	-0.240	[-0.432, -0.046]	-2.42	-0.235	[-0.428, -0.041]	-2.37
Children present	0.167	[-0.024, 0.361]	1.69	0.161	[-0.030, 0.353]	1.65
Random effects						
	P. Mean	95% CI	PCV	P. Mean	95% CI	PCV
L2 intercept	0.275	[0.161, 0.472]	0.559	0.266	[0.151, 0.480]	0.573
L1 residuals	1.487	[1.342, 1.640]	0.018	1.482	[1.338, 1.632]	0.021
PSR	0.155	[0.095, 0.243]		0.151	[0.092, 0.245]	
Spatial parameter						
	P. Mean	95% CI		P. Mean	95% CI	
Phi (ϕ) x 1000	3.972	[0.901, 8.603]		4.068	[0.731, 8.357]	
Range (km)	0.754	[0.348, 3.324]		0.736	[0.358, 4.096]	
Model fit index						
	DIC	Deviance	R^2	DIC	Deviance	R^2
Statistic	1,571.31	1,465.90	0.248	1,567.90	1,463.73	0.266
pD	105.41			104.16		
Note: Estimates obtained with Bayesian Markov chain Monte Carlo estimation via Gibbs sampling.						
95% CI = central 95% credible interval; DIC = deviance information criterion; L1 = level 1						
(individual); L2 = level 2 (neighborhood); P. Mean = posterior mean; PCV = proportional change in						
variance from Model 1 (level-specific R^2); pD = effective number of parameters; Phi = rate of						
decrease in autocorrelation (multiplied by 1,000 for display); PSR = partial sill ratio; R^2 = overall						
proportion of variance explained; Range = practical range of variogram.						
Table 10 (cont'd)						

Table 12 (cont'd)

Parameter	Model 17 (1.2 km crime buffer)			Model 31 (0.1 km NSES buffer)		
	P. Mean	95% CI	t	P. Mean	95% CI	t
L2 fixed effects						
Intercept	3.743	[3.524, 3.947]	34.86	3.563	[3.078, 3.967]	15.99
Crime (buffer)	0.273	[0.189, 0.352]	6.66			
NSES (buffer)				-0.015	[-0.023, -0.006]	-3.50
L1 fixed effects						
Age(years)						
36-55	0.149	[-0.041, 0.341]	1.53	0.141	[-0.050, 0.335]	1.44
≥ 56	0.174	[-0.083, 0.430]	1.33	0.178	[-0.079, 0.435]	1.36
Female	0.125	[-0.059, 0.307]	1.35	0.141	[-0.041, 0.326]	1.51
Race						
Black	-0.253	[-0.472, -0.039]	-2.29	-0.259	[-0.479, -0.040]	-2.32
Hispanic	-0.044	[-0.438, 0.349]	-0.22	-0.042	[-0.443, 0.357]	-0.20
Other	-0.079	[-0.652, 0.495]	-0.27	-0.102	[-0.675, 0.476]	-0.35
Marital status						
Married	0.028	[-0.189, 0.247]	0.25	0.055	[-0.161, 0.273]	0.50
Divorced	0.018	[-0.229, 0.270]	0.15	0.018	[-0.230, 0.265]	0.14
Widowed	-0.171	[-0.509, 0.167]	-1.00	-0.163	[-0.499, 0.174]	-0.96
Education						
< High school	-0.022	[-0.240, 0.194]	-0.20	-0.041	[-0.258, 0.173]	-0.37
Undergraduate	0.107	[-0.112, 0.325]	0.96	0.134	[-0.081, 0.352]	1.21
Postgraduate	0.288	[-0.277, 0.853]	1.00	0.305	[-0.253, 0.869]	1.07
Employed	-0.079	[-0.260, 0.103]	-0.86	-0.079	[-0.259, 0.100]	-0.86
Income (\$1,000s)						
< 15	0.154	[-0.143, 0.452]	1.01	0.124	[-0.174, 0.425]	0.81
15-25	-0.014	[-0.305, 0.272]	-0.10	-0.017	[-0.300, 0.273]	-0.12
25-45	-0.021	[-0.286, 0.242]	-0.15	-0.045	[-0.305, 0.216]	-0.34
Home owner	-0.244	[-0.438, -0.050]	-2.45	-0.229	[-0.423, -0.032]	-2.28
Children present	0.158	[-0.036, 0.350]	1.60	0.156	[-0.034, 0.351]	1.59
Random effects						
	P. Mean	95% CI	PCV	P. Mean	95% CI	PCV
L2 intercept	0.269	[0.160, 0.431]	0.568	0.434	[0.254, 0.742]	0.303
L1 residuals	1.485	[1.343, 1.634]	0.019	1.504	[1.365, 1.650]	0.006
PSR	0.152	[0.095, 0.230]		0.221	[0.143, 0.333]	
Spatial parameter						
	P. Mean	95% CI		P. Mean	95% CI	
Phi (ϕ) x 1000	3.690	[1.304, 7.757]		1.381	[0.723, 2.644]	
Range (km)	0.812	[0.386, 2.298]		2.169	[1.133, 4.145]	
Model fit index						
	DIC	Deviance	R^2	DIC	Deviance	R^2
Statistic	1,568.41	1,465.56	0.256	1,568.29	1,478.95	0.083
pD	102.85			89.34		

Note: Estimates obtained with Bayesian Markov chain Monte Carlo estimation via Gibbs sampling. 95% CI = central 95% credible interval; DIC = deviance information criterion; L1 = level 1 (individual); L2 = level 2 (neighborhood); P. Mean = posterior mean; PCV = proportional change in variance from Model 1 (level-specific R^2); pD = effective number of parameters; Phi = rate of decrease in autocorrelation (multiplied by 1,000 for display); PSR = partial sill ratio; R^2 = overall proportion of variance explained; Range = practical range of variogram.

Table 12 (cont'd)

Parameter	Model 32 (0.2 km NSES buffer)			Model 33 (0.3 km NSES buffer)		
	P. Mean	95% CI	t	P. Mean	95% CI	t
L2 fixed effects						
Intercept	3.606	[3.159, 3.968]	17.79	3.614	[3.203, 3.947]	19.55
Crime (buffer)						
NSES (buffer)	-0.021	[-0.031, -0.010]	-3.92	-0.022	[-0.034, -0.010]	-3.66
L1 fixed effects						
Age(years)						
36-55	0.149	[-0.040, 0.339]	1.53	0.146	[-0.044, 0.337]	1.50
≥ 56	0.185	[-0.075, 0.441]	1.41	0.163	[-0.094, 0.418]	1.25
Female	0.124	[-0.059, 0.307]	1.34	0.123	[-0.060, 0.304]	1.32
Race						
Black	-0.261	[-0.481, -0.040]	-2.34	-0.266	[-0.485, -0.046]	-2.37
Hispanic	-0.044	[-0.440, 0.347]	-0.22	-0.054	[-0.455, 0.340]	-0.27
Other	-0.111	[-0.682, 0.465]	-0.38	-0.106	[-0.686, 0.478]	-0.36
Marital status						
Married	0.056	[-0.161, 0.271]	0.51	0.053	[-0.162, 0.271]	0.48
Divorced	0.030	[-0.215, 0.277]	0.24	0.025	[-0.222, 0.272]	0.20
Widowed	-0.160	[-0.498, 0.176]	-0.94	-0.148	[-0.482, 0.189]	-0.87
Education						
< High school	-0.037	[-0.255, 0.182]	-0.33	-0.040	[-0.258, 0.178]	-0.36
Undergraduate	0.128	[-0.093, 0.345]	1.15	0.127	[-0.092, 0.345]	1.15
Postgraduate	0.299	[-0.270, 0.866]	1.04	0.317	[-0.251, 0.875]	1.11
Employed	-0.083	[-0.265, 0.097]	-0.89	-0.078	[-0.257, 0.102]	-0.85
Income (\$1,000s)						
< 15	0.128	[-0.170, 0.429]	0.84	0.142	[-0.153, 0.441]	0.94
15-25	-0.003	[-0.287, 0.282]	-0.02	-0.004	[-0.287, 0.284]	-0.03
25-45	-0.032	[-0.299, 0.232]	-0.24	-0.018	[-0.279, 0.243]	-0.13
Home owner	-0.233	[-0.429, -0.038]	-2.33	-0.235	[-0.432, -0.038]	-2.35
Children present	0.163	[-0.030, 0.355]	1.66	0.160	[-0.034, 0.352]	1.63
Random effects						
	P. Mean	95% CI	PCV	P. Mean	95% CI	PCV
L2 intercept	0.397	[0.226, 0.676]	0.361	0.387	[0.227, 0.657]	0.378
L1 residuals	1.498	[1.361, 1.644]	0.010	1.497	[1.359, 1.646]	0.011
PSR	0.207	[0.128, 0.312]		0.203	[0.128, 0.308]	
Spatial parameter						
	P. Mean	95% CI		P. Mean	95% CI	
Phi (ϕ) x 1000	1.659	[0.693, 3.526]		1.891	[0.773, 3.851]	
Range (km)	1.805	[0.850, 4.326]		1.584	[0.778, 3.874]	
Model fit index						
	DIC	Deviance	R^2	DIC	Deviance	R^2
Statistic	1,566.96	1,475.20	0.117	1,569.42	1,473.67	0.121
pD	91.77			95.75		

Note: Estimates obtained with Bayesian Markov chain Monte Carlo estimation via Gibbs sampling. 95% CI = central 95% credible interval; DIC = deviance information criterion; L1 = level 1 (individual); L2 = level 2 (neighborhood); P. Mean = posterior mean; PCV = proportional change in variance from Model 1 (level-specific R^2); pD = effective number of parameters; Phi = rate of decrease in autocorrelation (multiplied by 1,000 for display); PSR = partial sill ratio; R^2 = overall proportion of variance explained; Range = practical range of variogram.

REFERENCES

REFERENCES

- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage Publications, Inc.
- Anderson, L. S., Cooper, S., Hassol, L., Klein, D. C., Rosenblum, G., & Bennett, C. C. (1966). *Community psychology: A report of the Boston Conference on the Education of Psychologists for Community Mental Health*. Boston, MA: Boston University.
- Bailey, T. C., & Gatrell, A. C. (1995). *Interactive spatial data analysis*. Harlow, England: Prentice Hall.
- Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2004). *Hierarchical modeling and analysis for spatial data*. In V. Isham, N. Keiding, T. Louis, N. Reid, R. Tibshirani & H. Tong (Series Eds.), *Monographs on Statistics and Applied Probability*, Vol. 101. Retrieved from http://www.statsnetbase.com/ejournals/books/book_km.asp?id=1285
- Barcikowski, R. S. (1981). Statistical power with group mean as the unit of analysis. *Journal of Educational Statistics*, 6(3), 267-285. Retrieved from <http://jeb.sagepub.com>
- Barnett, A. G., Koper, N., Dobson, A. J., Schmiegelow, F., & Manseau, M. (2010). Using information criteria to select the correct variance–covariance structure for longitudinal data in ecology. *Methods in Ecology and Evolution*, 1(1), 15-24. doi: 10.1111/j.2041-210X.2009.00009.x
- Bass, J. K., & Lambert, S. F. (2004). Urban adolescents' perceptions of their neighborhoods: An examination of spatial dependence. *Journal of Community Psychology*, 32(3), 277-293. doi: 10.1002/jcop.20005
- Beard, J. R. (2008). New approaches to multilevel analysis. *Journal of Urban Health*, 85(6), 805-806. doi: 10.1007/s11524-008-9314-7
- Beyers, J. M., Bates, J. E., Pettit, G. S., & Dodge, K. A. (2003). Neighborhood structure, parenting processes, and the development of youths' externalizing behaviors: A multilevel analysis. *American Journal of Community Psychology*, 31(1-2), 35-53. doi: 10.1023/A:1023018502759
- Bingenheimer, J. B., & Raudenbush, S. W. (2004). Statistical and substantive inferences in public health: Issues in the application of multilevel models. *Annual Review of Public Health*, 25, 53-77. doi: 10.1146/annurev.publhealth.25.050503.153925
- Bivand, R. S., Pebesma, E. J., & Gómez-Rubio, V. (2008). *Applied spatial data analysis with R*. New York, NY: Springer Science+Business Media.

- Block, R. (2000). Gang activity and overall levels of crime: A new mapping tool for defining areas of gang activity using police records. *Journal of Quantitative Criminology*, 16(3), 369-383. doi: 10.1023/A:1007579007011
- Bowes, D. R., & Ihlanfeldt, K. R. (2001). Identifying the impacts of rail transit stations on residential property values. *Journal of Urban Economics*, 50(1), 1-25. doi: 10.1006/juec.2001.2214
- Boyd, H. A., Flanders, W. D., Addiss, D. G., & Waller, L. A. (2005). Residual spatial correlation between geographically referenced observations: A Bayesian hierarchical modeling approach. *Epidemiology*, 16(4), 532-541. doi: 10.1097/01.ede.0000164558.73773.9c
- Brodsky, A. E., O'Campo, P. J., & Aronson, R. E. (1999). PSOC in community context: Multi-level correlates of a measure of psychological sense of community in low-income, urban neighborhoods. *Journal of Community Psychology*, 27(6), 659-679. doi: 10.1002/(SICI)1520-6629(199911)27:6%3C659::AID-JCOP3%3E3.3.CO%3B2-R
- Brooks-Gunn, J., Duncan, G. J., Leventhal, T., & Aber, J. L. (1997). Lessons learned and future directions for research on the neighborhoods in which children live. In J. Brooks-Gunn, G. J. Duncan & J. L. Aber (Eds.), *Neighborhood Poverty* (Vol. I: Context and Consequences for Children, pp. 279-297). New York, NY: Russell Sage Foundation.
- Browne, W., & Goldstein, H. (in press). MCMC sampling for a multilevel model with non-independent residuals within and between cluster units. *Journal of Educational and Behavioral Statistics*.
- Browne, W. J., & Draper, D. (2006a). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, 1(3), 473-514. doi: 10.1214/06-BA117
- Browne, W. J., & Draper, D. (2006b). Rejoinder. *Bayesian Analysis*, 1(3), 547-550. doi: 10.1214/06-BA117REJ
- Browning, C. R., & Cagney, K. A. (2002). Neighborhood structural disadvantage, collective efficacy, and self-rated physical health in an urban setting. *Journal of Health and Social Behavior*, 43(4), 383-399. doi: 10.2307/3090233
- Burton, L. M., Price-Spratlen, T., & Spencer, M. B. (1997). On ways of thinking about measuring neighborhoods: Implications for studying context and developmental outcomes for children. In J. Brooks-Gunn, G. J. Duncan & J. L. Aber (Eds.), *Neighborhood poverty: Policy implications in studying neighborhoods* (Vol. II, pp. 132-144). New York, NY: Russell Sage Foundation.
- Caughy, M. O. B., Hayslett-McCall, K. L., & O'Campo, P. J. (2007). No neighborhood is an island: Incorporating distal neighborhood effects into multilevel studies of

child developmental competence. *Health & Place*, 13(4), 788-798. doi: 10.1016/j.healthplace.2007.01.006

Caughy, M. O. B., Nettles, S. M., & O'Campo, P. J. (2008). The effect of residential neighborhood on child behavior problems in first grade. *American Journal of Community Psychology*, 42(1-2), 39-50. doi: 10.1007/s10464-008-9185-9

Caughy, M. O. B., & O'Campo, P. J. (2006). Neighborhood poverty, social capital, and the cognitive development of African American preschoolers *American Journal of Community Psychology*, 37(1-2), 141-154. doi: 10.1007/s10464-005-9001-8

Chainey, S., Tompson, L., & Uhlig, S. (2008). The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal*, 21(1), 4-28. doi: 4-28. doi:10.1057/palgrave.sj.8350066

Chaix, B., Leyland, A. H., Sabel, C. E., Chauvin, P., Råstam, L., Kristersson, H., et al. (2006). Spatial clustering of mental disorders and associated characteristics of the neighbourhood context in Malmö, Sweden, in 2001. *Journal of Epidemiology and Community Health*, 60(5), 427-435. doi: 10.1136/jech.2005.040360

Chaix, B., Merlo, J., & Chauvin, P. (2005). Comparison of a spatial approach with the multilevel approach for investigating place effects on health: The example of healthcare utilisation in France. *Journal of Epidemiology and Community Health*, 59(6), 517-526. doi: 10.1136/jech.2004.025478

Chaix, B., Merlo, J., Subramanian, S. V., Lynch, J., & Chauvin, P. (2005). Comparison of a spatial perspective with the multilevel analytical approach in neighborhood studies: The case of mental and behavioral disorders due to psychoactive substance use in Malmö, Sweden, 2001. *American Journal of Epidemiology*, 162(2), 171-182. doi: 10.1093/aje/kwi175

Chaskin, R. J. (1997). Perspectives on neighborhood and community: A review of the literature. *Social Service Review*, 71(4), 521-547. doi: 10.1086/604277

Chaskin, R. J. (1998). Neighborhood as a unit of planning and action: A heuristic approach. *Journal of Planning Literature*, 13(1), 11-30 doi: 10.1177/088541229801300102

Chilès, J.-P., & Delfiner, P. (1999). *Geostatistics: Modeling spatial uncertainty*. New York, NY: John Wiley & Sons.

Cleveland, W. S. (1993). *Visualizing data*. Summit, NJ: Hobart Press.

Cohen, D. A., Ashwood, J. S., Scott, M. M., Overton, A., Evenson, K. R., Staten, L. K., et al. (2006). Public parks and physical activity among adolescent girls. *Pediatrics*, 118(5), e1381-e1389. doi: 10.1542/peds.2006-1226

- Coulton, C. J., Chan, T., & Mikelbank, K. (2010). *Finding place in Making Connections communities: Applying GIS to residents' perceptions of their neighborhoods*. Washington, DC: The Urban Institute. <http://www.urban.org/url.cfm?ID=412057>.
- Coulton, C. J., Cook, T., & Irwin, M. (2004). *Aggregation issues in neighborhood research: A comparison of several levels of census geography and resident defined neighborhoods*. Retrieved from Case Western Reserve University, Mandel School of Applied Social Sciences, Center on Urban Poverty and Social Change website: <http://digitalcase.case.edu:9000/fedora/get/ksl:2006052511/Cook-Agression-2004.pdf>
- Coulton, C. J., Korbin, J. E., Chan, T., & Su, M. (2001). Mapping residents' perceptions of neighborhood boundaries: A methodological note. *American Journal of Community Psychology*, 29(2), 371-383. doi: 10.1023/A:1010303419034
- Coulton, C. J., Korbin, J. E., & Su, M. (1996). Measuring neighborhood context for young children in an urban area. *American Journal of Community Psychology*, 24(1), 5-32. doi: 10.1007/BF02511881
- Cumming, G. (2009). Inference by eye: Reading the overlap of independent confidence intervals. *Statistics in Medicine*, 28(2), 205-220. doi: 10.1002/sim.3471
- Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist*, 60(2), 170-180. doi: 10.1037/0003-066X.60.2.170
- Dalton, J. H., Elias, M. J., & Wandersman, A. (2001). *Community psychology: Linking individuals and communities*. Belmont, CA: Wadsworth/Thomson Learning.
- Dietz, R. D. (2002). The estimation of neighborhood effects in the social sciences: An interdisciplinary approach. *Social Science Research*, 31(4), 539-575. doi: 10.1016/S0049-089X(02)00005-4
- Diez Roux, A. V. (2001). Investigating neighborhood and area effects on health. *American Journal of Public Health*, 91(11), 1783-1789. Retrieved from <http://ajph.aphapublications.org/>
- Diez Roux, A. V., Mujahid, M. S., Morenoff, J. D., & Raghunathan, T. (2007). Mujahid et al. respond to "Beyond the metrics for measuring neighborhood effects". *American Journal of Epidemiology*, 165(8), 872-873. doi: 10.1093/aje/kwm039
- Diggle, P. J., & Ribeiro, P. J. (2007). *Model-based geostatistics*. New York, NY: Springer Science+Business Media. doi:10.1007/978-0-387-48536-2.
- Downey, L. (2006). Using geographic information systems to reconceptualize spatial relationships and ecological context. *American Journal of Sociology*, 112(2), 567-612. doi: 10.1086/506418

- Duffy, K. G., & Wong, F. Y. (2002). *Community psychology* (3rd ed.). Needham Heights, MA: Allyn & Bacon.
- Duncan, C., Jones, K., & Moon, G. (1998). Context, composition, and heterogeneity: Using multilevel models in health research. *Social Science & Medicine*, 46(1), 97-117. doi: 10.1016/S0277-9536(97)00148-2
- Duncan, S. C., Duncan, T. E., & Strycker, L. A. (2002). A multilevel analysis of neighborhood context and youth alcohol and drug problems. *Prevention Science*, 3(2), 125-133. doi: 10.1023/A:1015483317310
- Duncan, T. E., Duncan, S. C., Okut, H., Strycker, L. A., & Hix-Small, H. (2003). A multilevel contextual model of neighborhood collective efficacy. *American Journal of Community Psychology*, 32(3/4), 245-252.
- Dupéré, V., & Perkins, D. D. (2007). Community types and mental health: a multilevel study of local environmental stress and coping [Electronic version]. *American Journal of Community Psychology*, 39(1-2), 107-119. doi: 10.1007/s10464-007-9099-y
- Edwards, L. J., Muller, K. E., Wolfinger, R. D., Qaqish, B. F., & Schabenberger, O. (2008). An R2 statistic for fixed effects in the linear mixed model. *Statistics in Medicine*, 27(29), 6137-6157. doi: 10.1002/sim.3429
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12(2), 121-138. doi: 10.1037/1082-989X.12.2.121
- Fagg, J., Curtis, S., Clark, C., Congdon, P., & Stansfeld, S. A. (2008). Neighbourhood perceptions among inner-city adolescents: Relationships with their individual characteristics and with independently assessed neighbourhood conditions. *Journal of Environmental Psychology*, 28(2), 128-142. doi: 10.1016/j.jenvp.2007.10.004
- Finley, A. O., Banerjee, S., & Carlin, B. P. (2007). spBayes: An R package for univariate and multivariate hierarchical point-referenced spatial models. *Journal of Statistical Software*, 19(4), 1-24. Retrieved from <http://www.jstatsoft.org/v19/i04>
- Finley, A. O., Banerjee, S., & Carlin, B. P. (2009). spBayes: Univariate and multivariate spatial modeling (Version 0.1-3) [Computer program, R package]. East Lansing, MI: Author. Retrieved from <http://cran.r-project.org/package=spBayes>
- Foster-Fishman, P. G., Cantillon, D., Pierce, S. J., & Van Egeren, L. (2007). Building an active citizenry: The role of neighborhood problems, readiness, and capacity for change. *American Journal of Community Psychology*, 39(1-2), 91-106. doi: 10.1007/s10464-007-9097-0

- Foster-Fishman, P. G., Pierce, S. J., & Van Egeren, L. (2009). Who participates and why: Building a process model of citizen participation. *Health Education & Behavior*, 36(3), 550-569. doi: 10.1177/1090198108317408
- Fox, J. (1997). *Applied regression analysis, linear models, and related methods*. Thousand Oaks, CA: Sage Publications.
- Franzini, L., Caughy, M., Spears, W., & Esquer, M. E. F. (2005). Neighborhood economic conditions, social processes, and self-rated health in low-income neighborhoods in Texas: A multilevel latent variables model. *Social Science & Medicine*, 61(6), 1135-1150.
- Franzini, L., Caughy, M. O. B., Nettles, S. M., & O'Campo, P. J. (2008). Perceptions of disorder: Contributions of neighborhood characteristics to subjective perceptions of disorder. *Journal of Environmental Psychology*, 28(1), 83-93. doi: 10.1016/j.jenvp.2007.08.003
- Galster, G. C. (2001). On the nature of neighborhood. *Urban Studies*, 38(12), 2111-2124. doi: 10.1080/00420980120087072
- Galster, G. C. (2008). Quantifying the effect of neighbourhood on individuals: Challenges, alternative approaches, and promising directions. *Schmollers Jahrbuch*, 128(1), 7-48. doi: 10.3790/schm.128.1.7
- Gates, L. B., & Rohe, W. M. (1987). Fear and reactions to crime: A revised model. *Urban Affairs Review*, 22(3), 425-453. doi: 10.1177/004208168702200305
- Gee, G. C. (2002). A multilevel analysis of the relationship between institutional and individual racial discrimination and health status. *American Journal of Public Health*, 92(4), 615-623. Retrieved from <http://ajph.aphapublications.org>
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3), 515-534. doi: 10.1214/06-BA117A
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.
- Gelman, A., & Pardoe, I. (2006). Bayesian measures of explained variance and pooling in multilevel (hierarchical) models. *Technometrics*, 48(2), 241-251. doi: 10.1198/004017005000000517
- Gephart, M. A. (1997). Neighborhoods and communities as contexts for development. In J. Brooks-Gunn, G. J. Duncan & J. L. Aber (Eds.), *Neighborhood Poverty* (Vol. I:

Context and Consequences for Children, pp. 1-43). New York, NY: Russell Sage Foundation.

Gill, J. (2008). *Bayesian methods: A social and behavioral sciences approach* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. New York, NY: Oxford University Press.

Grannis, R. (1998). The importance of trivial streets: Residential streets and residential segregation. *American Journal of Sociology*, 103(6), 1530-1564. doi: 10.1086/231400

Greenbaum, S. D. (1982). Bridging ties at the neighborhood level. *Social Networks*, 4(4), 367-384. doi: 10.1016/0378-8733(82)90019-3

Greenbaum, S. D., & Greenbaum, P. E. (1985). The ecology of social networks in four urban neighborhoods. *Social Networks*, 7(1), 47-76. doi: 10.1016/0378-8733(85)90008-5

Guest, A. M., & Lee, B. A. (1984). How urbanites define their neighborhoods. *Population and environment: Behavioral and social issues*, 7(1), 32-56. doi: 10.1007/BF01257471

Guo, J. Y., & Bhat, C. R. (2007). Operationalizing the concept of neighborhood: Application to residential location choice analysis. *Journal of Transport Geography*, 15(1), 31-45. doi: 10.1016/j.jtrangeo.2005.11.001

Haining, R. (2003). *Spatial data analysis: Theory and practice*. Cambridge, United Kingdom: Cambridge University Press.

Haurin, D. R., Dietz, R. D., & Weinberg, B. A. (2003). The impact of neighborhood homeownership rates: A review of the theoretical and empirical literature. *Journal of Housing Research*, 13(2), 119-151. doi: 10.2139/ssrn.303398

Hayes, A. F. (2006). A primer on multilevel modeling. *Human Communication Research*, 32(4), 385-410. doi: 10.1111/j.1468-2958.2006.00281.x

Hofmann, D. A., Griffin, M. A., & Gavin, M. B. (2000). The application of hierarchical linear modeling to organizational research. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 467-511). San Francisco, CA: Jossey-Bass.

Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299-314.

Isaaks, E. H., & Srivastava, R. M. (1989). *An introduction to applied geostatistics*. New York, NY: Oxford University Press.

- James, L. R., & Williams, L. J. (2000). The cross-level operator in regression, ANCOVA, and contextual analysis. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 382-424). San Francisco, CA: Jossey-Bass.
- Kastellac, J. P., & Leoni, E. I. (2007). Using graphs instead of tables in political science. *Perspectives on Politics*, 5(4), 755-771. doi: 10.1017/S1537592707072209
- Kearns, A., & Parkinson, M. (2001). The significance of neighborhood. *Urban Studies*, 38(12), 2103-2110. doi: 10.1080/00420980120087063
- Kingsley, G. T., Coulton, C. J., Barndt, M., Sawicki, D. S., & Tatian, P. (1997). *Mapping your community: Using geographic information to strengthen community initiatives*. Washington, D.C.: U.S. Department of Housing and Urban Development.
- Korbin, J. E., & Coulton, C. (1997). Understanding the neighborhood context for children and families: Combining epidemiological and ethnographic approaches. In J. Brooks-Gunn, G. J. Duncan & J. L. Aber (Eds.), *Neighborhood poverty: Policy implications in studying neighborhoods* (Vol. II, pp. 65-79). New York: Russell Sage Foundation.
- Kramer, M. (2005). R^2 statistics for mixed models. *Proceedings of the Conference on Applied Statistics in Agriculture*, 17, 148-160. Retrieved from http://www.ars.usda.gov/sp2UserFiles/ad_hoc/12000000SpatialWorkshop/19KramerSupplRsq.pdf
- Kruger, D. J. (2008). Verifying the operational definition of neighborhood for the psychosocial impact of structural deterioration. *Journal of Community Psychology*, 36(1), 53-60. doi: 10.1002/jcop.20216
- Kruger, D. J., Reischl, T. M., & Gee, G. C. (2007). Neighborhood social conditions mediate the association between physical deterioration and mental health. *American Journal of Community Psychology*, 40(3/4), 261-271. doi: 10.1007/s10464-007-9139-7
- Land, K. C., & Deane, G. (1992). On the large-sample estimation of regression models with spatial- or network-effects terms: A two-stage least squares approach. *Sociological Methodology*, 22, 221-248. Retrieved from <http://www.jstor.org>
- Laraia, B. A., Messer, L., Kaufman, J. S., Dole, N., Caughy, M., O'Campo, P. J., et al. (2006). Direct observation of neighborhood attributes in an urban area of the US south: characterizing the social context of pregnancy. *International Journal of Health Geographics*, 5(11), 1-11. doi: 10.1186/1476-072X-5-11
- Lebel, A., Pampalon, R., & Villeneuve, P. Y. (2007). A multi-perspective approach for defining neighbourhood units in the context of a study on health inequalities in

the Quebec City region. *International Journal of Health Geographics*, 6(27), 1-15. doi: 10.1186/1476-072X-6-27

- Lee, B. A. (2001). Taking neighborhoods seriously. In A. Booth & A. C. Crouter (Eds.), *Does it take a village?: Community effects on children, adolescents, and families* (pp. 31-40). Mahwah, N.J.: Lawrence Erlbaum Associates.
- Lee, B. A., & Campbell, K. E. (1997). Common ground? Urban neighborhoods as survey respondents see them. *Social Science Quarterly*, 78(4), 922-936.
- Lery, B. (2008). A comparison of foster care entry risk at three spatial scales. *Substance Use and Misuse*, 43(2), 223-237. doi: 10.1080/10826080701690631
- Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin*, 126(2), 309-337. doi: 10.1037/0033-2909.126.2.309
- Linney, J. A. (2000). Assessing ecological constructs and community context. In J. Rappaport & E. Seidman (Eds.), *Handbook of community psychology* (pp. 647-668). New York, NY: Kluwer Academic/Plenum Publishers.
- Livert, D. E., & Hughes, D. L. (2002). The ecological paradigm: persons in settings. In T. A. Revenson, A. R. D'Augelli, S. E. French, D. Hughes, D. E. Livert, E. Seidman, M. Shinn & H. Yoshikawa (Eds.), *A quarter century of community psychology: Readings from the American Journal of Community Psychology* (pp. 51-64). New York, NY: Kluwer Academic/Plenum Publishers.
- Luke, D. A. (2005). Getting the big picture in community science: Methods that capture context. *American Journal of Community Psychology*, 35(3/4), 185-200. doi: 10.1007/s10464-005-3397-z
- Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10(4), 325-337. doi: 10.1023/A:1008929526011
- Lusch, D. P. (2005). A primer on coordinate systems commonly used in Michigan Retrieved January 22, 2007, from http://www.rsgis.msu.edu/pdf/mi_coordinate_systems.pdf
- Maton, K. I., Perkins, D. D., & Saegert, S. (2006). Community psychology at the crossroads: Prospects for interdisciplinary research. *American Journal of Community Psychology*, 38(1/2), 9-22. doi: 10.1007/s10464-006-9062-3
- McCord, E. S., & Ratcliffe, J. H. (2007). A micro-spatial analysis of the demographic and criminogenic environment of drug markets in Philadelphia. *The Australian and New Zealand Journal of Criminology*, 40(1), 43-63. doi: 10.1375/acri.40.1.43

- McCord, E. S., Ratcliffe, J. H., Garcia, R. M., & Taylor, R. B. (2007). Nonresidential crime attractors and generators elevate perceived neighborhood crime and incivilities [Electronic version]. *Journal of Research in Crime and Delinquency*, 44(3), 295-320. doi: 10.1177/0022427807301676
- McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing data: A gentle introduction*. New York, NY: Guilford Press.
- McMillen, D. P. (2003). Neighborhood house price indexes in Chicago: a Fourier repeat sales approach. *Journal of Economic Geography*, 3(1), 57-73.
- Meersman, S. C. (2005). Objective neighborhood properties and perceptions of neighborhood problems: Using a geographic information system (GIS) in neighborhood effects and aging research. *Ageing International*, 30(1), 63-87. doi: 10.1007/BF02681007
- Merlo, J. (2003). Multilevel analytical approaches in social epidemiology: Measures of health variation compared with traditional measures of association. *Journal of Epidemiology and Community Health*, 57(8), 550-552. doi: 10.1136/jech.57.8.550
- Merlo, J., Chaix, B., Ohlsson, H., Beckman, A., Johnell, K., Hjerpe, P., et al. (2006). A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of Epidemiology and Community Health*, 60(4), 290-297. doi: 10.1136/jech.2004.029454
- Merlo, J., Chaix, B., Yang, M., Lynch, J., & Råstam, L. (2005a). A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon. *Journal of Epidemiology and Community Health*, 59(6), 443-449. doi: 10.1136/jech.2004.023473
- Merlo, J., Chaix, B., Yang, M., Lynch, J., & Råstam, L. (2005b). A brief conceptual tutorial on multilevel analysis in social epidemiology: interpreting neighbourhood differences and the effect of neighbourhood characteristics on individual health. *Journal of Epidemiology and Community Health*, 59(12), 1022-1029. doi: 10.1136/jech.2004.028035
- Merlo, J., Yang, M., Chaix, B., Lynch, J., & Råstam, L. (2005). A brief conceptual tutorial on multilevel analysis in social epidemiology: investigating contextual phenomena in different groups of people. *Journal of Epidemiology and Community Health*, 59(9), 729-736. doi: 10.1136/jech.2004.023929
- Messer, L. C. (2007). Invited commentary: beyond the metrics for measuring neighborhood effects. *American Journal of Epidemiology*, 165(8), 868-871. doi: 10.1093/aje/kwm038

- Montello, D. R., Goodchild, M. F., Gottsegen, J., & Fohl, P. (2003). Where's downtown?: Behavioral methods for determining referents of vague spatial queries. *Spatial Cognition and Computation*, 3(2 & 3), 185-204. doi: 10.1207/S15427633SCC032&3_06
- Morenoff, J. D. (2003). Neighborhood mechanisms and the spatial dynamics of birth weight. *American Journal of Sociology*, 108(5), 976-1017. doi: 10.1086/374405
- Morenoff, J. D., Sampson, R. J., & Raudenbush, S. W. (2001). Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence. *Criminology*, 39(3), 517-559. doi: 10.1111/j.1745-9125.2001.tb00932.x
- Mowbray, C. T., Wolley, M. E., Grogan-Kaylor, A., Gant, L. M., Gilster, M. E., & Shanks, T. R. W. (2007). Neighborhood research from a spatially oriented strengths perspective. *Journal of Community Psychology*, 35(5), 667-680. doi: 10.1002/jcop.20170
- Nicotera, N. (2007). Measuring neighborhood: A conundrum for human services researchers and practitioners. *American Journal of Community Psychology*, 40(1/2), 26-51. doi: 10.1007/s10464-007-9124-1
- Ntzoufras, I. (2009). *Bayesian modeling using WinBUGS*. Hoboken, NJ: John Wiley & Sons, Inc.
- O'Campo, P. J. (2003). Invited commentary: Advancing theory and methods for multilevel models of residential neighborhoods and health. *American Journal of Epidemiology*, 157(1), 9-13. doi: 10.1093/aje/kwf171
- Orbell, J. M., & Uno, T. (1972). A theory of neighborhood problem solving: Political action vs. residential mobility [Electronic version]. *The American Political Science Review*, 66(2), 471-489.
- Orelien, J. G., & Edwards, L. J. (2008). Fixed-effect variable selection in linear mixed models using R2 statistics. *Computational Statistics & Data Analysis*, 52(4), 1896-1907. doi: 10.1016/j.csda.2007.06.006
- Paccagnella, O. (2006). Centering or not centering in multilevel models? The role of the group mean and the assessment of group effects. *Evaluation Review*, 30(1), 66-85. doi: 10.1177/0193841X05275649
- Pampalon, R., Hamel, D., De Koninck, M., & Disant, M.-J. (2007). Perception of place and health: Differences between neighbourhoods in the Québec City region. *Social Science & Medicine*, 65(1), 95-111. doi: 10.1016/j.socscimed.2007.02.044
- Papachristos, A. V., & Kirk, D. S. (2006). Neighborhood effects and street gang behaviors. In J. F. Short & L. Hughes (Eds.), *Studying youth gangs* (pp. 63-84). Lanham, MD: AltaMira Press.

- Perkins, D. D., Meeks, J. W., & Taylor, R. B. (1992). The physical environment of street blocks and resident perceptions of crime and disorder: implications for theory and measurement. *Journal of Environmental Psychology*, 12(1), 21-34. doi: 10.1016/S0272-4944(05)80294-4
- Perkins, D. D., & Taylor, R. B. (1996). Ecological assessments of community disorder: Their relationship to fear of crime and theoretical implications. *American Journal of Community Psychology*, 24(1), 63-107. doi: 10.1007/BF02511883
- Perkins, D. D., Wandersman, A. H., Rich, R. C., & Taylor, R. B. (1993). The physical environment of street crime: Defensible space, territoriality and incivilities. *Journal of Environmental Psychology*, 13(1), 29-49. doi: 10.1016/S0272-4944(05)80213-0
- Peterson, N. A., & Reid, R. J. (2003). Paths to psychological empowerment in an urban community: Sense of community and citizen participation in substance abuse prevention activities. *Journal of Community Psychology*, 31(1), 25-38. doi: 10.1002/jcop.10034
- Peugh, J. L., & Enders, C. K. (2005). Using the SPSS mixed procedure to fit cross-sectional and longitudinal multilevel models. *Educational and Psychological Measurement*, 65(5), 717-741. doi: 10.1177/0013164405278558
- Pierce, S. J. (2006, November). *Kriging: A spatial analysis tool for visualizing survey data and mapping community conditions*. Demonstration session presented at Evaluation 2006: The Consequences of Evaluation, the annual conference of the American Evaluation Association, Portland, OR.
- Pierce, S. J. (2008). *Using survival analysis to examine outreach effects on survey response*. Presented to the Ecological-Community Psychology Interest Group, Department of Psychology, Michigan State University, East Lansing, MI.
- Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, 6(1), 7-11. Retrieved from http://cran.r-project.org/doc/Rnews/Rnews_2006-1.pdf
- Plummer, M., Best, N., Cowles, K., & Vines, K. (2009). coda: Output analysis and diagnostics for MCMC (Version 0.13-4) [Computer program, R package]. Lyon, France: International Agency for Research on Cancer. Retrieved from <http://cran.r-project.org>
- Quane, J. M., & Rankin, B. H. (2006). Does it pay to participate? Neighborhood-based organizations and the social development of urban adolescents. *Children and Youth Services Review*, 28(10), 1229-1250. doi: 10.1016/j.chldyouth.2006.01.004
- Quillian, L., & Pager, D. (2001). Black neighbors, higher crime? The role of racial stereotypes in evaluations of neighborhood crime. *American Journal of Sociology*, 107(3), 717-767. doi: 10.1086/338938

- R Development Core Team. (2009). R: A language and environment for statistical computing (Version 2.9.2) [Computer program]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>
- Ratcliffe, J. H. (2004). Geocoding crime and a first estimate of a minimum acceptable hit rate. *International Journal of Geographical Information Science*, 18(1), 61-72. doi: 10.1080/13658810310001596076
- Ratcliffe, J. H., & McCullagh, M. J. (1999). Hotbeds of crime and the search for spatial accuracy. *Journal of Geographical Systems*, 1(4), 385-398. doi: 10.1007/s101090050020
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Relph, E. (1976). *Place and placelessness*. London: Pion Limited.
- Roberts, J. K., & Monaco, J. P. (2006, April). *Effect size measures for the two-level linear multilevel model*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA. Retrieved from http://www.hlm-online.com/papers/HLM_effect_size.pdf
- Roosa, M. W., Jones, S., Tein, J.-Y., & Cree, W. (2003). Prevention science and neighborhood influences on low-income children's development: Theoretical and methodological issues. *American Journal of Community Psychology*, 31(1/2), 55-72. doi: 10.1023/A:1023070519597
- Ross, C. E., & Mirowsky, J. (1999). Disorder and decay: The concept and measurement of perceived neighborhood disorder. *Urban Affairs Review*, 34(3), 412-432. doi: 10.1177/107808749903400304
- Rountree, P. W., & Land, K. C. (1996). Burglary victimization, perceptions of crime risk, and routine activities: A multilevel analysis across Seattle neighborhoods and census tracts. *Journal of Research in Crime and Delinquency*, 33(2), 147-180. doi: 10.1177/0022427896033002001
- Sampson, R. J. (2001). How do communities undergird or undermine human development? Relevant contexts and social mechanisms. In A. Booth & A. C. Crouter (Eds.), *Does it take a village? Community effects on children, adolescents, and families* (pp. 3-30). Mahwah, NJ: Lawrence Erlbaum Associates.
- Sampson, R. J. (2004). Networks and neighborhoods: The implications of connectivity for thinking about crime in the modern city. In H. McCarthy, P. Miller & P. Skidmore (Eds.), *Network logic: Who governs in an interconnected world?* London: Demos.
- Sampson, R. J., & Morenoff, J. (1997). Ecological perspectives on the neighborhood context of urban poverty: Past and present. In J. Brooks-Gunn, G. J. Duncan & J.

- L. Aber (Eds.), *Neighborhood poverty: Policy implications in studying neighborhoods* (Vol. II, pp. 1-22). New York, NY: Russell Sage Foundation.
- Sámpson, 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. doi: 10.1146/annurev.soc.28.110601.141114
- Sampson, R. J., & Raudenbush, S. W. (1999). Systematic social observation of public spaces: A new look at disorder in urban neighborhoods. *American Journal of Sociology*, 105(3), 603-651. doi: 10.1086/210356
- Sampson, R. J., & Raudenbush, S. W. (2004). Seeing disorder: Neighborhood stigma and the social construction of "broken windows". *Social Psychology Quarterly*, 67(4), 319-342. doi: 10.2307/3649091
- Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, 277, 918-924. doi: 10.1126/science.277.5328.918
- Sastry, N., Pebley, A. R., & Zonta, M. (2002). Neighborhood definitions and the spatial dimension of daily life in Los Angeles [Electronic version] *Labor and Population Working Paper Series*. Santa Monica, CA: Rand Corporation.
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Boca Raton, FL: Chapman & Hall/CRC.
- Schafer, J. L. (2009). mix: Estimation/multiple imputation for mixed categorical and continuous data (Version 1.0-7) [Computer program, R package]. University Park, PA: Author. Retrieved from <http://cran.r-project.org>
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Bulletin*, 7(2), 147-777. doi: 10.1037//1082-989X.7.2.147
- Schill, M. H., & Wachter, S. M. (1995). Housing market constraints and spatial stratification by income and race. *Housing Policy Debate*, 6(1), 141-167. Retrieved from <http://www.mi.vt.edu/web/page/580/sectionid/580/pagelevel/1/interior.asp>
- Shinn, M., & Rapkin, B. D. (2000). Cross-level research without cross-ups in community psychology. In J. Rappaport & E. Seidman (Eds.), *Handbook of community psychology* (pp. 669-695). New York, NY: Kluwer Academic/Plenum Publishers.
- Shinn, M., & Toohey, S. M. (2003). Community contexts of human welfare. *Annual Review of Psychology*, 54, 427-459. doi: 10.1146/annurev.psych.54.101601.145052
- Skogan, W. G., & Maxfield, M. G. (1981). *Coping with crime: Individual and neighborhood reactions*. Beverly Hills, CA: Sage Publications.

- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583-639. doi: 10.1111/1467-9868.00353
- Spiegelhalter, D. J., Thomas, A., Best, N., & Lunn, D. J. (2007). WinBUGS (Version 1.4.3) [Computer program]. Cambridge, UK: Medical Research Council, Biostatistics Unit. Retrieved from <http://www.mrc-bsu.cam.ac.uk/bugs>
- Steptoe, A., & Feldman, P. J. (2001). Neighborhood problems as sources of chronic stress: Development of a measure of neighborhood problems, and associations with socioeconomic status and health. *Annals of Behavioral Medicine*, 23(3), 177-185. doi: 10.1207/S15324796ABM2303_5
- Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software*, 12(3), 1-16. Retrieved from <http://www.jstatsoft.org>
- Stutz, F. P. (1973). Distance and network effects on urban social travel fields. *Economic Geography*, 49(2), 134-144. Retrieved from <http://www.jstor.org>
- Sunder, P. K., Grady, J. J., & Wu, Z. H. (2007). Neighborhood and individual factors in marijuana and other illicit drug use in a sample of low-income women. *American Journal of Community Psychology*, 40(3-4), 167-180. doi: 10.1007/s10464-007-9135-y
- Suttles, G. D. (1972). *The social construction of communities*. Chicago, IL: University of Chicago Press.
- Swaroop, S., & Morenoff, J. D. (2006). Building community: The neighborhood context of social organization. *Social Forces*, 84(3), 1665-1695. doi: 10.1353/sof.2006.0058
- Taylor, R. B. (1998). Crime and small-scale places: What we know, what we can prevent, and what else we need to know. *Crime and place: Plenary papers of the 1997 Conference on Criminal Justice Research and Evaluation Research Forum* (pp. 1-22). Washington, D.C.: National Institute of Justice. Retrieved from <http://www.ncjrs.gov/pdffiles/168618.pdf>.
- Thomas, A., Best, N., Lunn, D., Arnold, R., & Spiegelhalter, D. (2004). *GeoBUGS user manual, version 1.2*. Cambridge, UK: Medical Research Council, Biostatistics Unit. <http://www.mrc-bsu.cam.ac.uk/bugs>.
- Tobler, W. R. (1970). A computer model simulation of urban growth in the Detroit region. *Economic Geography*, 46(2), 234-240.
- U.S. Census Bureau. (1994). *Geographic Areas Reference Manual*. Washington, D.C.: U.S. Government Printing Office. Retrieved from U.S. Department of Commerce,

Economics and Statistics Administration, Bureau of the Census website:
<http://www.census.gov/geo/www/garm.html>.

U.S. Census Bureau. (2002). *Census 2000 basics*. (MSO/02-C2KB). Washington, D.C.: U.S. Government Printing Office Retrieved from U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census website:
<http://www.census.gov/mso/www/c2000basics/00Basics.pdf>.

U.S. Census Bureau. (2007a). *2007 TIGER/Line® shapefile of Census 2000 block boundaries for Calhoun County, Michigan [Data file]*. Retrieved from U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census website: <http://www.census.gov/cgi-bin/geo/shapefiles/county-files?county=26025>

U.S. Census Bureau. (2007b). *2007 TIGER/Line® shapefile of Census 2000 block group boundaries for Calhoun County, Michigan [Data file]*. Retrieved from U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census website: <http://www.census.gov/cgi-bin/geo/shapefiles/county-files?county=26025>

U.S. Census Bureau. (2007c). *2007 TIGER/Line® shapefile of Census 2000 census tract boundaries for Calhoun County, Michigan [Data file]*. Retrieved from U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census website: <http://www.census.gov/cgi-bin/geo/shapefiles/county-files?county=26025>

Unger, D. G., & Wandersman, A. (1983). Neighboring and its role in block organizations: An exploratory report. *American Journal of Community Psychology* 11(3), 291-300. doi: 10.1007/BF00893369

Unger, D. G., & Wandersman, A. (1985). The importance of neighbors: The social, cognitive, and affective components of neighboring. *American Journal of Community Psychology*, 13(2), 139-169. doi: 10.1007/BF00905726

Uniform Crime Reporting Program. (2000). *National Incident-Based Reporting System Volume 1: Data Collection Guidelines*. Clarksburg, WV: U.S. Department of Justice, Federal Bureau of Investigation, Criminal Justice Information Services Division. Retrieved from <http://www.fbi.gov/ucr/nibrs/manuals/v1all.pdf>.

Van Dongen, S. (2006). Prior specification in Bayesian statistics: Three cautionary tales. *Journal of Theoretical Biology*, 242, 90-100. doi: 10.1016/j.jtbi.2006.02.002

Van Egeren, L., Huber, M. S. Q., Foster-Fishman, P. G., Pierce, S. J., & Law, K. (2007). *A multi-level sampling strategy for conducting neighborhood surveys in small cities*. Unpublished manuscript. Department of Psychology, Michigan State University, East Lansing, MI.

- Wellman, B. (1996). Are personal communities local? A Dumptarian reconsideration. *Social Networks*, 18(4), 347-354. doi: 10.1016/0378-8733(95)00282-0
- West, B. T., Welch, K. B., Galecki, A. T., & Gillespie, B. W. (2007). *Linear mixed models: A practical guide using statistical software* Retrieved from http://www.statsnetbase.com/ejournals/books/book_summary/summary.asp?id=6409
- Wheeler, J. O., & Stutz, F. P. (1971). Spatial dimensions of urban social travel. *Annals of the Association of American Geographers*, 61(2), 371-386. doi: 10.1111/j.1467-8306.1971.tb00789.x
- Wilson, J. Q., & Kelling, G. L. (1982, March). Broken windows: The police and neighborhood safety [Electronic version]. *The Atlantic Monthly*, 249, 29-38.
- Wilson, W. J. (1987). *The truly disadvantaged: The inner city, the underclass, and public policy*. Chicago, IL: University of Chicago Press.
- Wyant, B. R. (2008). Multilevel impacts of perceived incivilities and perceptions of crime risk on fear of crime: Isolating endogenous impacts. *Journal of Research in Crime and Delinquency*, 45(1), 39-64. doi: 10.1177/0022427807309440
- Xu, R. (2003). Measuring explained variation in linear mixed effects models. *Statistics in Medicine*, 22(22), 3527-3541. doi: 10.1002/sim.1572
- Zucker, D. M. (1990). An analysis of variance pitfall: The fixed effects analysis in a nested design. *Educational and Psychological Measurement*, 50(4), 731-738. doi: 10.1177/0013164490504002

MICHIGAN STATE UNIVERSITY LIBRARY



3 1293 03063 5928