Assessment of the Child Maltreatment Risk in Massachusetts

A thesis submitted by

Yuan Li

in partial fulfillment of the requirements for the degree of

Master of Arts

in

Urban and Environmental Policy and Planning

Tufts University

August 2017

Adviser: Dr. Sumeeta Srinivasan

Reader: Dr. Thomas J. Stopka

ABSTRACT

Child maltreatment is one of the most challenging public health issues facing families in Massachusetts. To help policymakers make informed decisions for the prevention of child maltreatment in Massachusetts, we identified hotspot clusters of child maltreatment counts and rates across the state and areas with a high risk of child maltreatment but low access to existing Family Centers that served prevent risks. To better understand mechanisms underlying these spatial patterns, we confirmed the presence of the spatial autocorrelation in regression residuals from the OLS regression model, improved the estimates by conducting spatial lag regression analysis, and identified areas with unusual characteristics that warrant further study. Our findings help portray the landscape of child maltreatment risks in Massachusetts and inform the strategic allocation of service resources in Massachusetts to ensure prevention programs are better targeted to under-served areas with a high risk of child maltreatment.

ACKNOWLEDGMENTS

I have been lucky to study in the Master degree program of Urban and Environmental Policy and Planning (UEP) at Tufts from September 2015 to May 2017, because the program has allowed me to get to know many wonderful people. Without their help, support, and encouragement, it would have been impossible to complete this thesis.

I would like to thank my thesis committee for their great contributions to this endeavor. I appreciate my thesis advisor Dr. Sumeeta Srinivasan for being the great mentor for me, teaching me a great deal about GIS-based skills, bringing assurance to me when I was not confident in the work, and contributing innovative input to my thesis. I am grateful to Dr. Thomas J. Stopka, my thesis reader and research advisor, for introducing me to the research project on Laying the Groundwork for a Child Maltreatment Prevention Initiative (LGCMPI) where my thesis topic comes from, for helping advance my academic development, and for providing constructive advice throughout my research and thesis and suggestions for my career development. I also appreciate his efforts of making his research lab where I was working over the past year a stimulating place to work, where I learned the way of thinking critically, advanced research methods and skills, and thought-provoking ideas. I thank my lab mates, “Turtle Team Friends”, for sharing with me the wonderful time in the lab: Rachel Hoh, Anna Kaplan, and Harsha Amaravadi. In a larger team of LGCMPI, I like to thank Dr. Rebecca C. Fauth, Dr. Jessica Goldberg, and Katie Aasland, for helping me develop the knowledge of child maltreatment and providing helpful feedback to my research work on child maltreatment.

My deep thanks also go to the greatest faculty, staff, and peers at UEP. The connections with them help me better understand the role of practical visionaries and commitment to diversity, which have greatly influenced me throughout my academic endeavors when working on public health issues especially for disadvantaged groups. I am honored to be part of the UEP community! My sincerest thanks are extended to Dr. Barbara M. Parmenter, lecturer and student affairs coordinator of UEP. She opened the door of GIS for me, recommended me to Dr. Thomas J. Stopka as a research assistant that gave me a chance to work on LGCMPI, and provided support and encouragement to help me come through my hardest time with dealing with a totally new culture as an international student. I greatly appreciate Christine Cousineau, lecturer of UEP, for helping me develop skills in academic writing in English, helping me better understand the sense of purpose for being an urban planner, and providing constructive suggestions on my professional development. I would like to thank Dr. Francine Jacobs, associate professor of UEP, for training me to think critically, being my advisor of the filed project, and helping lay the foundation for my understanding of child maltreatment issues.

Finally, I owe a huge debt of gratitude to my families who stand with me all the time and support me to make best to this endeavor, including my parents, Qingyou Li and Xin Tang, my parents-in-law, Zhongmin Wu and Xiujie Zhang, my uncle Qingzhong Li and aunt-in-law Guizhi Xie, and my aunt Hongying Li and uncle-in-law Yeqing Liu. I am deeply grateful to my husband, Xu Wu, for being with me over the past ten years even when I am living abroad. Having him to share my ups and downs, smiles and tears, and successes and struggles makes my life an exciting and pleasant journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Child maltreatment refers to the abuse and neglect that occurs to children less than 18 years old (WHO 2016) and is one of the most challenging public health issues facing families worldwide as well as in the United States. The annual Child Maltreatment report provided by the U.S. Department of Health and Human Services (HHS), notes that, in 2015, about 683,000 child victims were reported (U.S. Department of Health & Human Services 2017). Massachusetts had one of the highest child maltreatment rates nationally, with a 2015 rate of 22.4 incidents per 1,000 children, which was much higher than the national rate of 0.092 per 1,000 children (U.S. Department of Health & Human Services 2017). To best respond to this serious challenge in Massachusetts, it would be useful to better understand the landscape of child maltreatment risk and its indicators across the state.

A wide range of studies has confirmed that community context is an important component of child maltreatment. Child maltreatment tends to concentrate among socially, economically, and psychologically high-risk families (Garbarino and Sherman 1980) and past research indicates that socioeconomic factors are highly correlated with high rates of child maltreatment and that children at risk for maltreatment are often spatially clustered (Coulton et al. 2007, 1995; Drake and Pandey 1996; Garbarino and Sherman 1980; Spearly and Lauderdale 1983; van Dijken, Stams, and de Winter 2016). Typically, the highest maltreatment rates often occur in locations with high percentages of poverty, unemployment, female-headed households, racial segregation, abandoned housing, and population loss (Coulton et al. 1995). Particularly, poverty has the most significant effect on child maltreatment rates (Coulton, Korbin, and Su 1999; Coulton et al. 1995; Garbarino and Sherman 1980; Drake and Pandey 1996; Ernst 2000; Garbarino and Crouter 1978; Garbarino and Kostelny 1992; Kim 2004).

A higher child-care burden, measured by a combination of children per adult, a large proportion of female family headship, and unbalanced sex ratio, also relates to a high child maltreatment risk (Coulton et al. 1995). Especially, child abuse highly associates with a high density of low-income working single mothers who have parental stress regarding child care burden (Ha, Collins, and Martino 2015). Moreover, these associations vary in specific ethnic and racial groups. Garbarino and Kostelny (1992) found that different relationships between child maltreatment rates and percent of female family headship in their research areas, negative in the African-American areas and positive in the Hispanic areas, while Freisthler, Bruce, and Needell (2007) found that child maltreatment rates varied in black, Hispanic, and white children.

In addition to examining the relationship between the child maltreatment outcomes and these community-based risk factors, associations between child maltreatment risk and spatial effects are revealed by an increasing number of studies using a geographic information system (GIS). Child maltreatment referrals were found to be related to the geographic distribution of services (Freisthler 2011), while low levels of public awareness of child maltreatment were associated with a lack of placement of Coalition organizations (Cardazone and colleagues 2014). Moreover, the risk of a child living in poverty was found not to be randomly distributed in space (Voss et al. 2006), while three types of the child victimization: incidents of child abuse, child neglect, and juvenile assault, were found to have different spatial clustering patterns (Paulsen 2004). However, these spatial patterns of child maltreatment risk in Massachusetts are not well understood and few regression analysis models for estimating child maltreatment risk take these spatial effects into account.

Further, in response to the serious risk of child maltreatment, the Massachusetts Children's Trust (CT), the Commonwealth's leading family support organization, has tried to prevent child abuse by supporting parents and strengthening families. CT is currently developing an expansion plan of its existing Family Centers, which are place-based, family-centered, multi-generational service hubs that provide universal access to families with young children, in the form of play groups, parenting education, care coordination, and support groups. By 2016, CT had funded eight Family Centers connecting nearly forty cities and towns in Massachusetts serving 15,000 families statewide (Massachusetts Children's Trust 2017). In the future, CT's ambitious maltreatment prevention initiative plans to operate in 10-15 sites across the state, with the Family Centers serving as the focal points. However, little is known about the geospatial distribution of child maltreatment risk across Massachusetts. Analyses of such distributions can assist CT in evaluating access to existing Family Center services so as to allocate service resources effectively. This study seeks to fill these research gaps.

Chapter 2 will present the results of hotspot analysis in a GIS to examine whether child maltreatment counts and rates were clustered spatially and temporally at the zip code level across Massachusetts. The analysis will identify persistent hotspot clusters of child maltreatment that indicate a high risk of child maltreatment, examine if existing Family Centers are located in these areas with the high risk, and assess the types of the communities where the hotspot clusters of child maltreatment rates are located. Chapter 3 will present the results of OLS regression and spatial lag regression to test spatial effects on child maltreatment rates in Massachusetts, correct the OLS model by taking spatial autocorrelation into account in spatial regression analysis, obtain improved estimates of the relationship between child maltreatment risk and community-based factors, and identify unusual locations that warrant further study. Finally, Chapter 4 will summarize findings and limitations, and include suggestions for further research.

CHAPTER 2: USING HOTSPOT ANALYSIS TO ASSESS THE RISK OF CHILD MALTREATMENT IN MASSACHUSETTS

Introduction

As discussed in the previous chapter, a wide range of studies has confirmed that socio-economic factors are highly correlated with high rates of child maltreatment. In addition to poverty which has the greatest effect on child maltreatment risks, the availability, adequacy, and use of family support resources in the community have a direct relation to risks (Garbarino and Crouter 1978). A lack of resources, social isolation, and less involvement in community social organizations that offer formal and informal social networks and support, may make it hard for parents and children to obtain support in the times of need. (Coulton et al. 1995).

This important tie between the spatial access to service resources and the effectiveness of preventing child maltreatment is revealed by an increasing number of studies. Associations have been found between rates of child maltreatment referrals and the geographic distribution of services (Freisthler 2011). Locating intervention programs in places with easy access to public transport will facilitate the use of the services (Marsh, D'Aunno, and Smith 2000; Onyskiw et al. 1999). Cardazone and colleagues (2014) studied the relationship between the network of Hawaii Children's Trust Fund Coalition organizations and the public awareness of child maltreatment factors by using network analyses in a GIS, which helped identify further opportunities for reaching target populations by visualizing the areas with low levels of the public awareness and lack of placement of the Coalition organizations.

Just like geographic access to services, child maltreatment risk also clusters in space. Voss et al. (2006) found that the risk of a child living in poverty was not randomly distributed in space. Paulsen (2004) used hotspot analysis in GIS to find that the three types of the child victimization: incidents of child abuse, child neglect, and juvenile assault, have different spatial clustering patterns. However, the spatial patterns of child maltreatment risk in Massachusetts are not well understood. Little is known about child maltreatment clusters in Massachusetts and the spatial relationship between Family Center locations and spatial clusters of child maltreatment.

This chapter aims to conduct hotspot analysis to determine if child maltreatment counts and rates cluster in space by finding areas with statistically significant spatial hotspots. Our findings will allow policymakers and the CT to better understand the landscape of child maltreatment risks in Massachusetts, to assess whether existing locations of Family Centers are effectively placed to maximize access to services among target populations that are most in need, and to inform the geographic and strategic placement of future Family Centers in Massachusetts to ensure prevention programs are better targeted to under-served areas with a high risk of child maltreatment.

Methodology

Data and sources

Child maltreatment data. We based the assessment of child maltreatment risk on the rates of unduplicated child maltreatment counts per 1,000 children under 18 years at the zip code level (n=538) in Massachusetts. We obtained data on the counts of child maltreatment events from reports with supported allegations of child maltreatment following an investigation from 2009 to 2014 provided by the Massachusetts Department of Children and Families (DCF). To calculate child maltreatment rates per 1,000 children and assess sociodemographic factors associated with maltreatment, we obtained population denominators and community-level sociodemographic data from the 2010-2014 American Community Survey 5-Year Estimates of U.S. Census Bureau (US Census Bureau 2017). We obtained population estimates from children and youth under 18 years of age at the zip code level within Massachusetts (n=538).

We calculated the child maltreatment rates per 1,000 children under 18 years at the zip code level. We gave “0” to the field value of rates in a few zip codes with no DCF records. Generally, over 93% of the zip codes had records, except for 2012 when 70% of the zip codes had records of maltreatment (see Table 1).

Table 1. Child Maltreatment Counts by Massachusetts Zip Codes, 2009 to 2014

Estimate Year	Zip Codes		Percent: Records Contained/Total
	Total Number	Records Contained	
2009	538	509	93.54%
2010	538	484	94.18%
2011	538	462	95.69%
2012	538	441	69.50%
2013	538	461	98.48%
2014	538	486	98.32%

Source: Child maltreatment counts were provided by the Massachusetts Department of Children and Families.

Family Center catchment areas. To assess access to existing Family Centers in Massachusetts, we obtained Massachusetts Family Center addresses and catchment areas from the Children's Trust Massachusetts website (see Table 2). It allowed for the visualization of the boundaries of Family Center catchment areas based on the political boundary layer of towns in Massachusetts.

Table 2. Catchment Areas of Existing Family Centers in Massachusetts

Family Center Locations		Community Served (Catchment Areas)
1	Medford	Medford
2	Cambridge	Cambridge
3	Athol	Athol, Erving, New Salem, Orange, Petersham, Royalston, Warwick, Wendell and Phillipston.
4	North Eastham	Brewster, Chatham, Eastham, Harwich, Orleans, Provincetown, Truro and Wellfleet.
5	Springfield	Springfield
6	North Adams	Adams, Cheshire, Clarksburg, Florida, Hancock, Lanesboro, Monroe, New Ashford, North Adams, Savoy and Willamstown.
7	Vineyard Haven	Aquinnah, Chilmark, Edgartown, Oak Bluffs, Tisbury and West Tisbury
8	Greenfield	Greenfield and Turner Falls.

Source: Massachusetts Children's Trust website, available at http://childrenstrustma.org/our-programs/family-centers/family-centers-contact-list.

Analysis Techniques

We converted the addresses of eight existing Family Centers into to longitude and latitude measures for each address in a GIS so that the addresses can be placed on a map like pushpins. We also visualized the boundaries of catchment areas (cities or towns) for the existing eight Family Centers in a GIS. Based on child maltreatment rates in 2009, we developed descriptive maps by using the quintile classification method at the zip code level (n=538). We divided child maltreatment rates at the zip code level into five equal groups and each group (a quintile) represented one fifth of the range of the child maltreatment rates. Visualizing these zip codes by quintiles allowed us to see about how child maltreatment rates in 2009 differed at the zip code level across Massachusetts. To compare results across years, we developed descriptive maps for the three years, 2009, 2011, and 2014, based on 2009 reference quintiles.

Conduct hotspot analysis to visualize the landscape of child maltreatment risks

To test whether child maltreatment counts and rates cluster spatially in Massachusetts, we applied hotspot analysis (Getis-Ord Gi*) (Getis and Ord 1992; Ord and Getis 1995) to determine whether or not spatial autocorrelation of a variable exists. This approach allowed us to assess whether high values (statistically significant hotspots) or low values (statistically significant coldspots) clustered spatially (Mitchell 1999). Statistically significant hotpots of child maltreatment counts or rates at the zip code level indicated that those zip codes with high values of counts or rates of child maltreatment, together with their neighboring zip codes, had higher average local counts or rates of child maltreatment than the global average for all zip codes in the entire state. To conduct a systematic hotspot analysis, we employed the 5-step geoprocessing approach used by Stopka et al. (2014) as follows:

1) We controlled for outliers regarding polygon size. Our unit of analysis was zip codes (n=538) with areas ranging from 0.01 square kilometers (km^2) to 263.33 km^2, mean area of 40 km^2, and the Standard Deviation (SD) of polygon areas of 38 km^2. Large variation of polygon sizes could affect an analysis through bias in data visualization and challenges in defining neighborhood size (Stopka et al. 2014). For example, the visualized polygon with a much bigger size would be perceived as more notable than another polygon with a small size even though they were similar in the outcome value.

To handle these challenges, we: (1) temporarily removed 16 zip code polygons with no population counts from the database of 538 zip codes; (2) calculated the area of the resulting 522 zip code polygons; and (3) identified 38 zip code polygons with areas larger than 1.5 SD above the mean polygon area as outliers regarding polygon size (>97 km^2) and temporarily removed them from the database. Further, we temporarily removed 21 "island polygons" (see Appendix A) that had less than two neighboring polygons (i.e., the zip code of 02657 had only one adjacent zip code polygon). As a result, 463 zip code polygons that were similar in size and had at least two neighbors remained for the next steps in analyses.

2) We identified the beginning distance and distance increment values for the Incremental Spatial Autocorrelation (ISA) analysis (Step 3). Different from non-spatial statistics, spatial statistics combine space and spatial relationships and identify the conceptualization of spatial relationships (Mitchell 1999). We conducted hotspot analysis that applied a fixed distance band to identify the spatial relationships between zip codes in Massachusetts. Within this distance band, each zip code polygon would be analyzed in the context of its neighboring zip code polygons to reflect the influence of these neighbors by equally weighting these polygons within the band and not weighting those polygons outside the band.

To justify the distance band for analysis, we started by identifying an initial distance and an incremental distance value for the ISA analysis by calculating the maximum and the average distances between the geocentroid (the geographic center) of each zip code polygon and the geocentroid of its two nearest neighboring zip codes to ensure it had at least two neighbors. When determining the distance where the most pronounced clustering will be, using the actual maximum distance between neighboring zip codes can be too conservative if the database includes locational outliers (Stopka et al. 2014). Therefore, we used 2/3 of the maximum distance (round down to 9000m) as the beginning distance and 1/2 of the average distance (round down to 2000m) as the incremental distance for the ISA analysis in Step 3.

3) We justified a distance band reflecting maximum spatial autocorrelation by conducting ISA analysis. Through the ISA analysis, we calculated the spatial autocorrelation (Global Moran's I) based on a series of distances to get Z-scores indicating the intensity of spatial clustering and peaks (Mitchell 1999). We conducted ISA analysis by using the identified beginning distance and incremental distance in Step 2 to calculate Z-scores across 30 distance bands for each year from 2009 to 2014. To compare results across years, we identified the shortest distance parameter of 11,000 meters (see Table 3) with the first statistically significant peak Z-score ($p<0.05$ and 95% confidence level) for the distance parameter in the subsequent spatial weights matrix and hotspot analysis. This distance reflects the smallest distance at which spatial clusters of child maltreatment are most pronounced.

Table 3. Z-scores with the First Peak Distance Resulting from ISA Analysis, 2009 to 2014

Yeas	First Peak Distance	Z-Score
2014	17000 meters	4.71 (p=0.000002)
2013	11000 meters	6.89 (p<0.0000001)
2012	11000 meters	9.66 (p<0.0000001)
2011	13000 meters	9.33 (p<0.0000001)
2010	15000 meters	7.28 (p<0.0000001)
2009	13000 meters	6.13 (p<0.0000001)

4) We defined spatial relationships for analysis. To represent the spatial relationships among all zip codes (n=538), now including previously excluded outlier and island zip codes, across Massachusetts, we created a Spatial Weights Matrix (SWM) where each zip code polygon had at least 2 neighbors and based the conceptualization of spatial relationships on fixed distance and the distance parameter of 11,000 meters that was identified as the threshold distance in Step 3. The threshold distance parameter was overridden by the number of neighbor parameter when the minimum number of neighbors was not found (Mitchell 1999), so every zip code polygon was ensured to have at least two neighbors even if the neighbors were not within the threshold distance.

5) We conducted Hotspot Analyses (HSA) to determine spatial clusters. We calculated the Getis-Ord Gi* statistic to identify statistically significant hotspot and coldspot clusters of child maltreatment counts and rates across the state. Statistically significant hotspots indicated that the local mean child maltreatment counts or rates for a zip code, and its neighboring zip codes were significantly higher than the global mean child maltreatment counts or rates for all zip codes in Massachusetts. Statistically significant coldspots indicated that the local mean child maltreatment counts or rates for a zip code, and its neighboring zip codes were significantly lower than the global mean child maltreatment counts or rates for all zip codes across Massachusetts.

We selected the confidence level of 95% and defined the zip codes with a Z-score > 1.96 ($p<0.05$) as child maltreatment hotspots (represented in red) and the zip codes with a Z-score <- 1.96 ($p<0.05$) as coldspots (represented in blue). We defined zip codes with mean rates of child maltreatment as those with a Z-score >-1.96 and < 1.96, and these zip codes were represented in yellow. By adding a new field to each table of HSA layers, we selected zip codes with a "Z-score>1.96" and then gave them the value "1" (i.e., "hotspot zip code"). The added field in each HSA shapefile tabulation of 2009, 2011, and 2014, had two values: 1 (a zip code was defined as a hotspot) and 0 (a zip code was not defined as a hotspot). We then merged this information from the three years to get a new shapefile that had hotspot indicators for each year.

To better understand how hotspot clusters changed between 2009, 2011, and 2014, we defined five categories for analysis (see Table 4). Zip codes were defined as "No Hotspots" if "0" occurred in the HSA field during each year, which meant these zip codes were not a hotspot in 2009, 2011, and 2014; "Diminishing Hotspots" which meant these zip codes were a hotspot (identified by "1") in 2009 but not in the more recent years (i.e., 2011 or 2014); "Recent Hotspots" if "1" occurred in 2011 or 2014; "Intensifying Hotspots" if "1" occurred twice, which meant these zip codes were a hotspot for at least two years; "Persistent Hotspots" if "1" occurred in each year, which meant these zip codes were consistently identified as hotspots in 2009, 2011, and 2014.

Table 4. Typology for Visualizing Explanatory Time-Space Changes of Hotspot Clusters

<table>
<tr><th>Definition</th><th>2009 HSA</th><th>2011 HSA</th><th>2014 HSA</th><th>Years of Being Hotspots</th></tr>
<tr><td>No Hotspots</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>Diminishing Hotspots</td><td>1</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td rowspan="2">Recent Hotspots</td><td>0</td><td>1</td><td>0</td><td rowspan="2">1</td></tr>
<tr><td>0</td><td>0</td><td>1</td></tr>
<tr><td rowspan="3">Intensifying Hotspots</td><td>1</td><td>1</td><td>0</td><td rowspan="3">2</td></tr>
<tr><td>0</td><td>1</td><td>1</td></tr>
<tr><td>1</td><td>0</td><td>1</td></tr>
<tr><td>Persistent Hotspots</td><td>1</td><td>1</td><td>1</td><td>3</td></tr>
</table>

Results

Descriptive maps of child maltreatment rates

Figure 1 shows how child maltreatment counts and rates varied at the zip code level (n=538) across Massachusetts in 2009, 2011, and 2014. Based on 2009 reference quintiles, all statistics of child maltreatment counts or rates were divided into 5 classes. The darker colors in the legend mean that there is a higher value of child maltreatment counts or rates in those zip codes. Generally, we found that high child maltreatment counts and rates were concentrated in the zip codes of the central Massachusetts (i.e., Worcester County) as well as areas in Hampden County.

However, we identified different spatial distributions of child maltreatment when comparing maltreatment counts and rates. Child maltreatment counts tended to be higher in eastern and southeast Massachusetts, while the maps of child maltreatment rates presented a reversed result, with higher rates generally in western Massachusetts. These results provide a preliminary descriptive understanding of the geographic distribution of child maltreatment counts and rates across the state.

Counts VS Rates

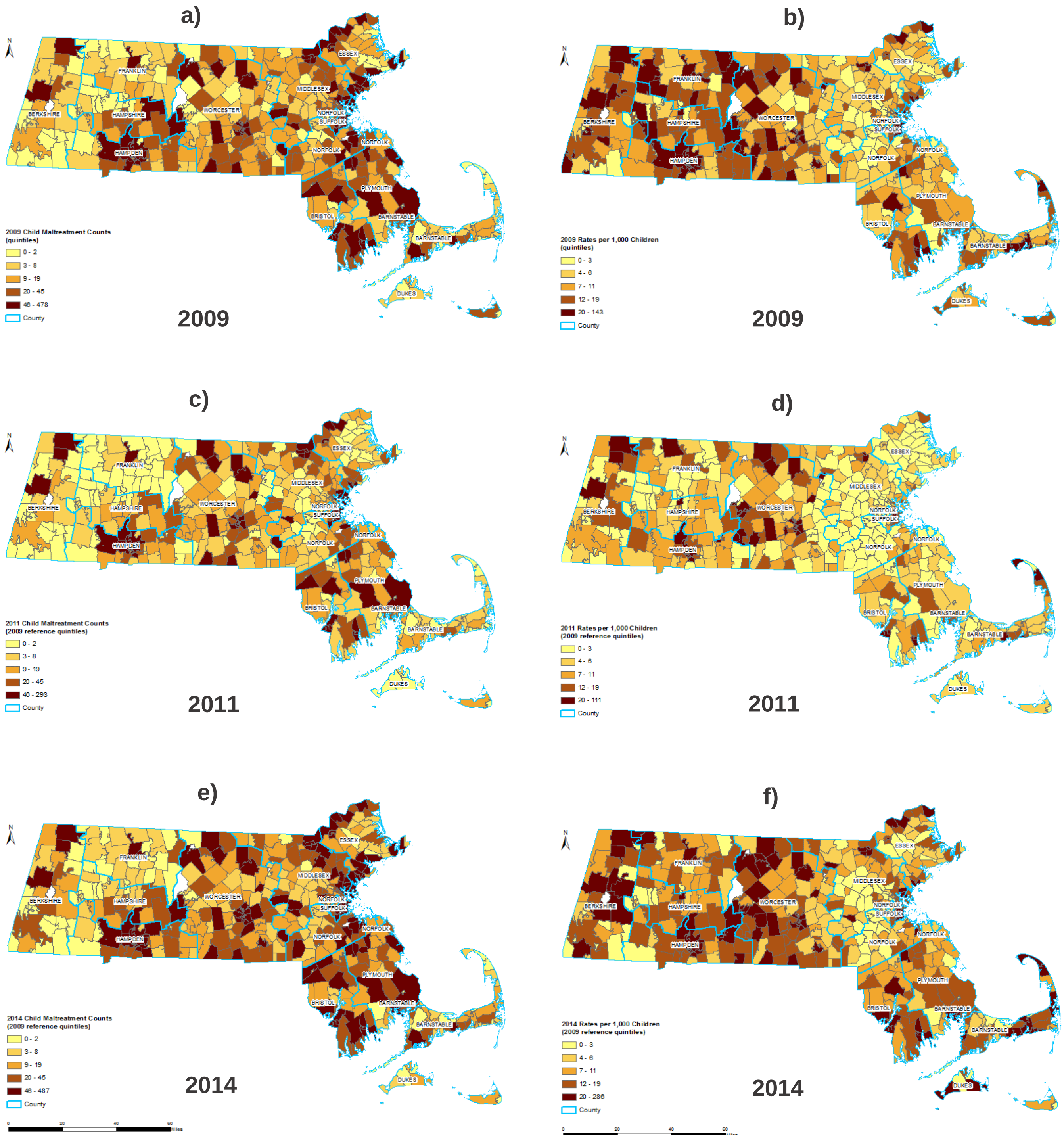


Figure 1. Child Maltreatment Counts and Rates by Massachusetts Zip Codes (n=538)

a): 2009 child maltreatment counts; **b): 2009** child maltreatment rates; **c): 2011** child maltreatment counts; **d): 2011** child maltreatment rates; **e): 2014** child maltreatment counts; **f): 2014** child maltreatment rates.
Data sources: Child maltreatment counts represent unduplicated counts of reports with supported allegations of child maltreatment following an investigation in Massachusetts from 2009 to 2014 provided by the Massachusetts Department of Children and Families; demographic estimates of population under 18 years come from 2010-2014 American Community Survey 5-Year Estimates from U.S. Census Bureau.

Spatial hotspot and coldspot clusters of child maltreatment

Figures 2 depicts statistically significant hotspot and coldspot clusters of child maltreatment counts and rates across Massachusetts zip codes in 2009, 2011, and 2014. In the hotspot maps for child maltreatment counts, we found hotspot clusters in nine counties including Berkshire, Hampden, Worcester, Middlesex, Suffolk, Norfolk, Plymouth, Bristol, and Barnstable, and coldspot clusters in two counties including Middlesex and Norfolk. In the hotspot maps for child maltreatment rates, we found hotspot clusters in seven counties including Berkshire, Franklin, Hampshire, Hampden, Worcester, Bristol, and Barnstable, and coldspot clusters in three counties Middlesex, Norfolk, and Essex.

We found some similarities when comparing hotspot and coldspot clusters for child maltreatment counts and rates. The maps of both counts and rates showed persistent hotspot clusters over the three years in Hampden, the areas between Hampden and Worcester and between Bristol and Plymouth. Also, the border regions between Middlesex and Norfolk counties showed consecutive coldspot clusters over the three years, although the cluster maps for counts presented fewer coldspot clusters than the maps for rates.

However, we also identified a number of differences in the clustering patterns of child maltreatment counts and rates. In the cluster maps based on counts we identified a high density of hotspot clusters in Middlesex, Suffolk, and Norfolk counties and a low density of hotspot clusters in Berkshire, Franklin, and Worcester, while the cluster maps for child maltreatment rates showed a low density of hotspot clusters in Middlesex, Suffolk, and Norfolk counties and a high density of hotspot clusters in Berkshire, Franklin, and Worcester. Also, the cluster maps for counts showed fewer hotspot clusters in the border regions between Berkshire and Franklin and between Worcester and Hampshire than the cluster maps for rates. Further, in the cluster maps for counts we noted fewer coldspot clusters in Middlesex and Norfolk Counties even though both the cluster maps of counts and rates showed coldspot clusters in these areas.

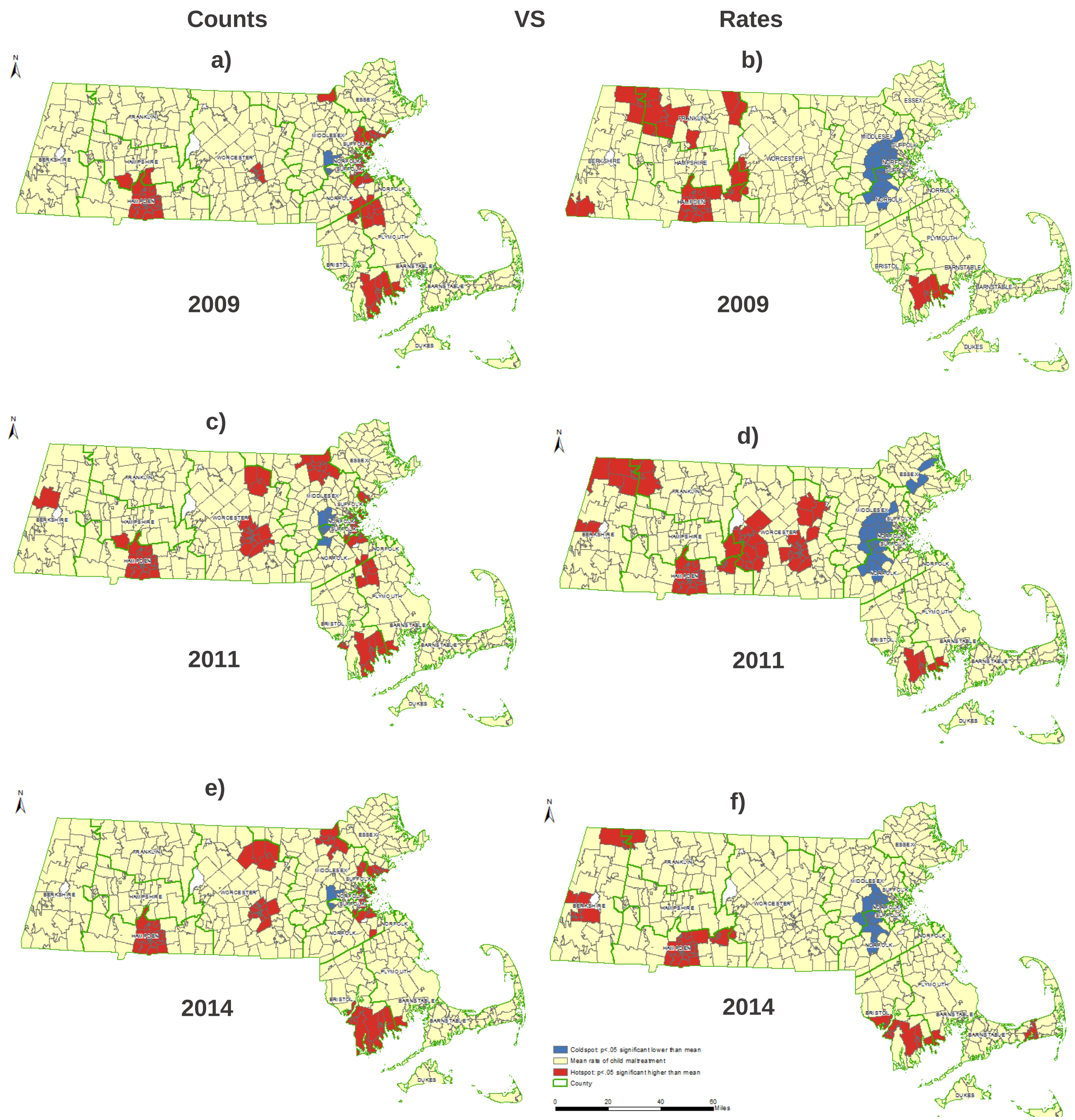


Figure 2. Child Maltreatment Spatial Clusters by Massachusetts Zip Codes (n=538)

a): 2009 spatial clusters of child maltreatment counts; **b): 2009** spatial clusters of child maltreatment rates; **c): 2011** spatial clusters of child maltreatment counts; **d): 2011** spatial clusters of child maltreatment rates; **e): 2014** spatial clusters of child maltreatment counts; **f): 2014** spatial clusters of child maltreatment rates.

Note: A **hotspot** is defined as a cluster of zip codes with significantly higher child maltreatment counts or rates than the global mean child maltreatment counts or rates for all zip codes in Massachusetts (Z-score > 1.96, p<0.05) and a **coldspot** is defined as a cluster of zip codes with significantly lower rates than the global mean (Z-score < -1.96, p<0.05).

Data sources: Child maltreatment counts represent unduplicated counts of reports with supported allegations of child maltreatment following an investigation in Massachusetts from 2009 to 2014 provided by the Massachusetts Department of Children and Families; demographic estimates of population under 18 years come from 2010-2014 American Community Survey 5-Year Estimates from U.S. Census Bureau.

Explanatory space-time changes of hotspot clusters (i.e., Hotspot typology)

Based on the hotspot clusters of child maltreatment rates at the zip code level in 2009, 2011, and 2014, we identified 32 zip codes that were persistently located in child maltreatment hotspot clusters for the three years (i.e., Persistent Hotspots=32); 11 zip codes that were hotspot clusters for at least for two years (i.e., Intensifying Hotspots=11); 49 zip codes that were hotspot clusters in 2011 or 2014 (i.e., Recent Hotspots=49); 17 zip codes that were hotspot clusters in 2009 but not in the recent years, 2011 or 2014 (i.e., Diminishing Hotspots=17).

In Figure 3 and Table 5, we present the locations for different types of hotspot clusters based on child maltreatment rates over the three years in eight Massachusetts counties: Berkshire, Franklin, Hampshire, Hampden, Worcester, Barnstable, Bristol, and Plymouth. Hampden County has the largest number of zip codes located within Persistent Hotspots (n=23), where clustering is evident in 12 towns including Chicopee, East Longmeadow, Feeding Hills, Holyoke, Ludlow, Thorndike, West Springfield, Springfield, Longmeadow, Indian Orchard, Agawam, and Bondsville. Especially, Chicopee had 3 zip codes located within Persistent Hotspots and Springfield had 10. Also, 5 Intensifying Hotspots clustered in towns of Hampden County, including Hampden, Southwick, Three Rivers, Westfield, Wilbraham. These areas presented a higher density of spatial hotspots of child maltreatment than the rest areas across the state.

As one of Hampden's neighbors, Worcester County had the most of Recent Hotspots (n=33) clustering in 33 towns like Gilbertville, Hardwick, South Barre, Wheelwright, Petersham and Barre, Auburn, Brookfield, East Brookfield, Holden. Moreover, the city of Worcester had a particularly high density of child maltreatment spatial clusters that involved 9 zip codes. In the areas between Hampshire and Worcester counties, both Warren and Ware showed an Intensifying Hotspot; in the areas between Berkshire and Franklin counties, the towns of Monroe Bridge, Rowe, and Drury had spatial clusters of child maltreatment; in the areas between Bristol and Plymouth Counties, Mattapoisett and Fairhaven had a Persistent Hotspot.

To prevent child maltreatment, eight Family Centers have been built in Massachusetts (see Figure 3). Family Centers within Athol, North Adams, Greenfield, and Springfield are located in close proximity to areas with a high density of hotspot clusters of child maltreatment. However, no Family Centers are established in the areas with a high density of hotspot clusters in Worcester and Bristol counties. Special attention is needed in these areas for providing access to future Family Center locations.

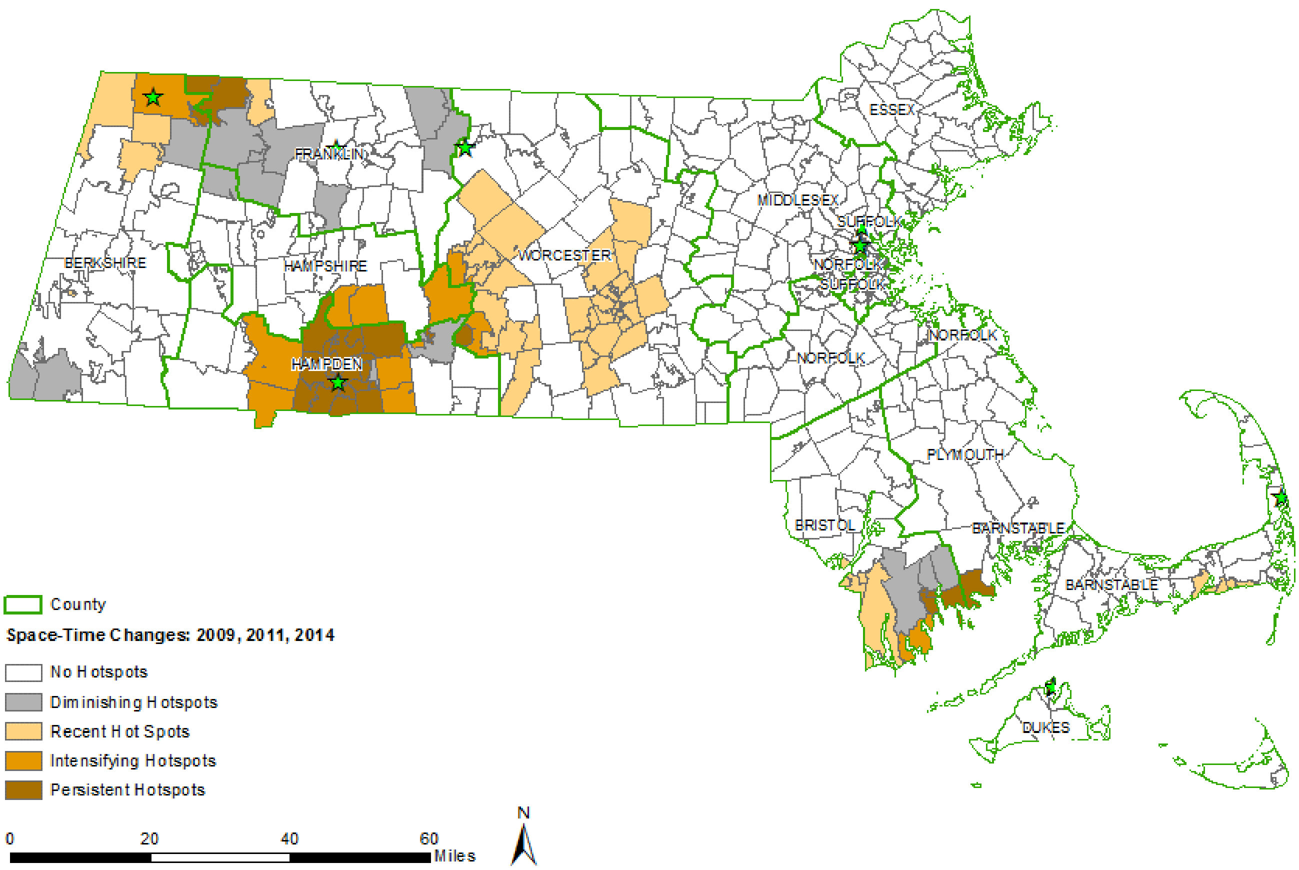


Figure 3. Explanatory Space-Time Changes of Hotspot Clusters for Child Maltreatment Rates by Massachusetts Zip Codes (n=538)

Note: No Hotspots are defined as zip codes that are not hotspot clusters in 2009, 2011, or 2014; **Diminishing Hotspots** are defined as zip codes that are hotspot clusters in 2009 but not in the recent years (2011 or 2014); **Recent Hotspots** are defined as zip codes that are hotspot clusters in 2011 or 2014; **Intensifying Hotspots** are defined as zip codes that are hotspot clusters at least for two years; **Persistent Hotspots** are defined as zip codes that are persistently stayed as hotspot clusters for the three years.

Table 5. Zip Codes and Cities/Towns Located in Hotspot Clusters in 2009, 2011, and 2014

County	Zip Codes Located in Diminishing Hotspots		Zip Codes Located in Recent Hotspots		Zip Codes Located in Intensifying Hotspots		Zip Codes Located in Persistent Hotspots	
	No.	City/Town Involved	No.	City/Town Involved	No.	City/Town Involved	No.	City/Town Involved
Barnstable	0		5	West Harwich, West Dennis, South Yarmouth, Harwich Port, Dennis Port	0		0	
Berkshire	3	Savoy, Sheffield, South Egremont	5	Adams, Cheshire, Lenox Dale, Williamstown, Glendale	1	North Adams	1	Drury
Bristol	3	Acushnet, New Bedford, North Dartmouth	5	Fall River (3), Somerset, Westport	1	South Dartmouth	4	Fairhaven, New Bedford (3)
Franklin	7	Ashfield, Buckland, Charlemont, Orange, Shelburne Falls, South Deerfield, Warwick	1	Heath	0		2	Monroe Bridge, Rowe
Hampden	2	Palmer, Springfield	0		5	Hampden, Southwick, Three Rivers, Westfield, Wilbraham	23	Chicopee (3), East Longmeadow, Feeding Hills, Holyoke, Ludlow, Thorndike, West Springfield, Springfield (10), Longmeadow, Indian Orchard, Agawam, Bondsville
Hampshire	2	Plainfield, Amherst	0		3	Granby, South Hadley, Ware	0	
Plymouth	0		0		0		1	Mattapoisett
Worcester	0		33	Gilbertville, Hardwick, South Barre, Wheelwright, Petersham, Barre, Auburn, Brookfield, East Brookfield, Holden, Jefferson, Leicester, Millbury, New Braintree, North Oxford, Oxford, Rochdale, Shrewsbury, South Lancaster, Sterling, Sturbridge, West Boylston, West Brookfield, Worcester (9), Cherry Valley	1	Warren	1	Warren
Suffolk	0		0		0		0	
Essex	0		0		0		0	
Nantucket	0		0		0		0	
Dukes	0		0		0		0	
Middlesex	0		0		0		0	
Norfolk	0		0		0		0	

Note. "City/Town Involved" means cities or towns where the hotspots are located; "No." means the total number of the zip codes that are located in hotspots.

Types of the community with hotspot clusters in Massachusetts

To better understand the types of the communities with hotspot clusters presented in Figure 3, we overlaid the Figure 3 and the classification system of Massachusetts community types developed by the Metropolitan Area Planning Council (MAPC). Based on the classification system (see Figure 4), we found that there are no hotspot clusters in inner core (including metropolitan core communities and streetcar suburbs). Almost all mature suburbs (including mature suburban towns and established suburbs and Cape Cod) had no hotspot clusters, except for just a few zip codes near Harwich with several zip codes located within recent hotspot clusters. Most hotspot clusters were in newly developing suburbs (including maturing New England towns and country suburbs) and regional urban centers (including major regional urban centers and sub-regional urban centers). Based on the MAPC's classification system of Massachusetts community types, we found more hotspot clusters were located in urban communities than in rural communities, especially in developing suburbs and regional urban centers rather than well built-out metropolis, suburbs near the metropolitan core, or maturing suburbs that were nearly built out.

a) b)

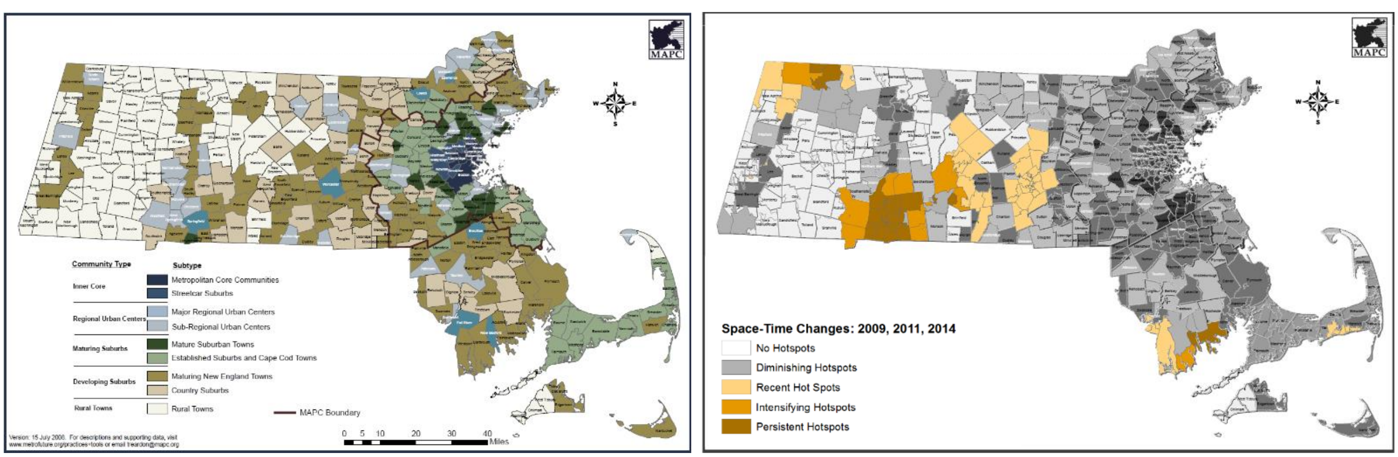


Figure 4. Explanatory Types of the Communities with Hotspot Clusters

a) Map of Massachusetts municipalities by type; **b)** Map of explanatory space-time changes of Hotspot Clusters for child maltreatment rates (Figure 3) overlaid on the map a).

Note: The Metropolitan Area Planning Council (MAPC), has created a classification system of municipalities in Massachusetts, five basic community types across the state, four of which can be subdivided further to yield nine sub-types. The criteria used to define Community Types include land use and housing patterns, recent growth trends, and projected development patterns. The classification system developed by (MAPC) is summarized: **1) Inner Core** includes Metropolitan Core Communities and Streetcar Suburbs; **2) Regional Urban Centers** includes Major Regional Urban Centers and Sub-Regional Urban Centers; **3) Maturing Suburbs** includes Mature Suburban Towns and Established Suburbs and Cape Cod Towns; **4) Developing Suburbs** includes Maturing New England Towns and Country Suburbs; and **5) Rural Towns**.

Source: Map of Massachusetts municipalities by type was obtained from Metropolitan Area Planning Council, available at http://www.mapc.org/sites/default/files/Massachusetts_Community_Types_-_July_2008.pdf.

Discussion

In this chapter, we examined whether child maltreatment counts and rates clustered spatially and temporally at the zip code level across Massachusetts by conducting hotspot analysis in a GIS. Our findings presented statistically significant clusters in space in terms of both child maltreatment counts and rates with a 95 percent confidence level ($p<0.05$). Generally, eight counties including Berkshire, Franklin, Hampshire, Hampden, Worcester, Barnstable, Bristol, and Plymouth, presented a high density of hotspot clusters, while four counties including Essex, Middlesex, Suffolk, and Norfolk presented a high density of coldspot clusters.

However, the comparison of these spatial clusters based on child maltreatment counts and rates revealed substantial differences. Some areas had a high density of hotspot clusters when analyzing counts but a low density of hotspot clusters when analyzing rates. Therefore, it is important to examine the child maltreatment risk by considering both variables. Hotspots based on counts may help to depict the overall burden of child maltreatment, while hotspots based on rates, control for population differences across zip codes.

Combining the hotspot clusters in 2009, 2011, and 2014, we showed how persistent these spatial hotspots could be over time. Identifying cities or towns with a persistently high density of child maltreatment clusters helped identify the areas at high risk. In selecting future Family Center locations, policy makers should pay attention to the identified persistent hotspot clusters of child maltreatment.

Also, relating the Hotspot results to the types of the communities suggested that hotspot clusters were found in developing suburbs and regional urban centers rather than the built-out metropolis or their nearby suburbs, maturing suburbs, or rural communities. The community characteristics may confirm that child maltreatment is associated with a high childcare burden but a lack of supportive services for families or children. The next chapter will take a closer look at community characteristics that may contribute to a high risk of child maltreatment.

CHAPTER 3: SPATIAL REGRESSION ANALYSIS OF CHILD MALTREATMENT RISK IN MASSACHUSETTS

Introduction

In order to understand the association between child maltreatment risk and community-based factors, a number of studies have employed multivariate regression analysis based on the Ordinary Least Squares (OLS) model. (Ha, Collins, and Martino 2015; Spearly and Lauderdale 1983; Coulton et al. 1995; Garbarino and Crouter 1978; Berger et al. 2009; Drake and Pandey 1996). However, very few of them takes spatial effects into account. Spatial effects can refer to spatial dependence or the spatial clustering of the dependent variable in a regression model and a spatial structure underlining this spatial cross-correlation (L. Anselin 1988). When spatial dependency (measured by spatial autocorrelation) exists in the data for regression analysis, its exclusion violates the assumptions of the classic OLS regression specification that requires all disturbance terms have the same variance and not be correlated with one another (Kennedy 1998, 43). Consequently, the use of the classic OLS regression will produce biased parameter estimates or incorrect statistical inference.

Furthermore, not considering spatial effects on child maltreatment may be misleading for policymakers who will use these results to design policies or programs for preventing child maltreatment. Planning for a prevention program largely relies on indicator data that are aggregated from localities with dramatic variations across regions of the state, but the planned program may apply uniformly across the state without considering the needs of specific localities (Anselin, Sridharan, and Gholston 2007). Some studies have revealed the importance of incorporating spatial effects on child-relevant issues into policy making. By using spatial regression procedures, Freisthler, Bruce, and Needell (2007) found child maltreatment rates varied in three different child groups (black, Hispanic, and white), while Voss et al.

(2006) improved the estimates of child poverty by considering the joint effects of place-influences and family-influences.

Chapter 2 has suggested the presence of the spatial autocorrelation of child maltreatment rates at the zip code level across Massachusetts, but not the mechanisms underlying these spatial patterns. This chapter aims to: 1) test variables that affect child maltreatment rates in Massachusetts using an OLS regression model; 2) correct the OLS model for spatial effects by estimating spatial regression models to take into account spatial dependency; 3) obtain an improved estimate of the relationship between child maltreatment risk and community-based variables by using the spatial regression models; and 4) identify unusual locations that warrant further study. Our findings will help confirm the presence and importance of spatial effects on estimating child maltreatment outcomes through an improved regression model and identify locations that have issues with "underprediction" based on the spatial clusters of regression residuals. These locations warrant further study so as to tailor policy for preventing child maltreatment to better fit with local characteristics.

Methodology

Data and sources

To examine the relationship between child maltreatment risk and community-based risk factors for the state, we defined child maltreatment rates as our outcome variable (dependent variable) and nine community-based factors, including poverty level, unemployment rate, child dependency ratio, sex ratio, family headship, tenure ratio, residential stability, population density of Hispanic or Latino Americans, and population density of African Americans, as covariates (independent variables) for regression analysis (see Table 6). The choice of the independent variables was based on previous literature reviewed for this study.

We based the assessment of child maltreatment risk on 2014 child maltreatment rates per 1,000 children under 18 years at the zip code level (n=538) in Massachusetts. We obtained data on the counts of child maltreatment events from reports with supported allegations of child maltreatment following an investigation in 2014 provided by the Massachusetts Department of Children and Families (DCF). To calculate child maltreatment rates per 1,000 population, we obtained population estimates for children under 18 years of age at the zip code level within Massachusetts (n=538) from the 2010-2014 American Community Survey 5-Year Estimates of U.S. Census Bureau (2010-2014 ACS).

To assess sociodemographic factors associated with child maltreatment, we also obtained community-based sociodemographic data from the 2010-2014 ACS (see Table 6), including percent of income in the past 12 months below the poverty level, unemployment rate (16 years old and over), child dependency ratio of people aged 0-14 per 100 people aged 15-64, sex ratio of females per 100 males, percent of female householder with children (no husband present) out of the total families with children under 18 years, tenure ratio of renters per 100 owners, percent of population living in the same house more than one year, percent of population of Hispanic or Latino Americans, and percent of African Americans.

Table 6. Outcome Variables and Covariates for Regression Analysis

Variable	Definition
Outcome Variable (Dependent Variable)	
2014 Child Maltreatment Rates	2014 Child maltreatment counts per 1,000 children under 18 years
Covariates (Independent Variables)	
Poverty Status	
• Poverty Level	% Income in the past 12 months below poverty level
• Unemployment Rate	% Resident unemployed (16 years and over)
Child care burden	
• Child Dependency Ratio	People aged 0-14 per 100 people aged 15-64
• Sex Ratio	Females per 100 males
• Family Headship	% Female householder with children (no husband present)
Residential instability	
• Tenure Ratio	Renters per 100 owners
• Residential Stability	% Living in the same house more than 1 year
Immigrant concentration	
• Population Density of Hispanic or Latino Americans	% Population of Hispanic or Latino Americans
• Population Density of African Americans	% Population of African Americans

Source: Child maltreatment counts represent unduplicated counts of reports with supported allegations of child maltreatment following an investigation in Massachusetts in 2014 provided by the Massachusetts Department of Children and Families; estimates of nine covariates come from 2010-2014 American Community Survey 5-Year Estimates from U.S. Census Bureau.

Analysis Software

We used a GIS (ArcGIS v10.4.1) to create the shapefile that included all variables at the zip code level within Massachusetts (n=538) and then used a geostatistical software package (Geoda v1.8.16.4) to conduct regression analyses (OLS regression analysis and spatial regression analysis) and develop maps (histograms, spatial cluster maps, Moran's I scatter plot, and significance maps).

Analysis Techniques

Classic OLS Regression (Model 1)

As the most popular method for doing empirical work, the OLS regression helps examine the statistic relationship between dependent and independent variables through the parameters of the linear regression model of minimizing the sum of squared residuals (Kennedy 1998, 45). To make the variable closely conform to the regression assumption of normality and to improve linearity between the dependent variable and the independent variables (Aldrich & Nelson, 1984), we firstly conducted log transformation based on "e" (approximately equal to 2.71828) for the outcome variable (2014 child maltreatment rates) to reduce its skewness in distribution and developed histograms to visualize the reduced skewness. Therefore, the outcome variable is log-normal conditional on all the covariates.

Based on the log transformed outcome variable, we developed the OLS model for analysis as follow (see formula I):

$$Log(y_i) = \alpha_0 + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \cdots + \alpha_9 X_{i9} + \varepsilon_i \qquad \text{(I)}$$

Where $\mathrm{Log}(y_i)$= an i×1 vector (i=538) with log transformed observations on the dependent variable (2014 child maltreatment rates), $X_{i1,i2,\dots,i9}$ = an i×9 matrix of observations on the independent variables (nine community-based factors), α_0 = a constant by having all the independent variables equal zero, $\alpha_{1,2,3,\dots,9}$ = the matching 9×1 vector of regression coefficients of the independent variables, and ε_i = an i×1 vector of regression errors or residuals. The assumption of OLS regression analysis requires regression residuals have uniform variance and be uncorrelated (Kennedy 1998, 45), so we set the null hypothesis that no dependence would exist in the test results of OLS regression residuals.

Regression coefficients of the independent variables indicate that the relationship with the log transformed dependent variable is positive or negative. We expected that the coefficients for the eight independent variables were positive, including poverty status factors (poverty level and unemployment rate), child care burden factors (child dependency ratio, sex ratio, and percent of female householder), residential instability factor (tenure ratio), and immigrant concentration factors (percent of population of Hispanic and percent of population of African Americans). These expected positive associations mean that an increase in log transformed child maltreatment rates is presumed to be associated with an increase in the values of the eight independent variables. On the other hand, we expected that the coefficient for residential stability was negative, which means log transformed child maltreatment rates would tend to increase as the percent of population staying in the same house more than one year decreases.

Furthermore, we were interested in the relationship between the independent variables with the dependent variable rather than just the log transformed dependent variable. To do so, we calculated the exponentiated regression coefficients (exp (α)) for each independent variable that had a statistically significant coefficient, since exponentiation is the inverse of logarithm function and these exponentiated values correspond to changes in the ratio of the expected geometric means of the original outcome variable (UCLA Institute for Digital Research and Education 2017).

After running the OLS regression (Model 1) in Geoda, we examined key results of the regression, including measure of fit, non-spatial diagnostics (Normality and Heteroskedasticity), and spatial diagnostics (Lagrange Multiplier Test, Robust versions of the Lagrange Multiplier Tests, and Moran's I). We tested for spatial autocorrelation of regression residuals by examining Moran's I, but Moran's I would not suggest guidance to decide which spatial regression specification to use, spatial lag regression specification or spatial error regression specification (Luc Anselin and Rey 2014, 108). Therefore, we took a closer look at Lagrange Multiplier Test (LM) for the choice of alternative specification. When the standard versions of LM tests (LM-Lag or LM-Error) were significant, the selection of alternative specification to use would only consider the Robust versions of LM tests (Luc Anselin and Rey 2014, 104–9).

We also saved the results of Model 1 residuals (differences between observed and estimated values) for each observation of the dependent variable. Positive Moran's I indicates positive spatial autocorrelation (the possible presence of spatial heterogeneity in space); negative residuals indicate a tendency to overpredict (observed values smaller than predicted values); positive residuals indicate a tendency to under-predict (observed values larger than predicted values) (L. A. Anselin 2017). To visualize these spatial autocorrelation patterns, we mapped the spatial clusters of Model 1 residuals at the zip code level (n=538) across Massachusetts to see how Model 1 worked differently in space.

Spatial Lag Regression (Model 2)

Based on Lagrange Multiplier Tests results (Robust LM-Lag rather than Robust LM-Error statistic was significant), we selected spatial lag regression as our spatial regression specification. The objective of the spatial lag regression model is to relate the target variable at one point in space to the observations for this variable in other spatial units (L. Anselin 1988, 22). Using Model 2, we associated the value of child maltreatment rates (dependent variable) for a zip code to child maltreatment rates for its neighboring zip codes that were defined as "neighbors" in a structure expressed by a spatial weight matrix. According to L. Anselin (1988)'s specification, we expressed the spatial lag regression model as follows (see Formula II).

$$Log(y) = \rho W_y + \alpha X + u \qquad \text{(II)}$$

Compared to the OLS regression specification of Model 1 (see Formula I), the spatial lag regression specification of Model 2 included a spatially lagged dependent variable W_y with spatial autoregressive parameter ρ and W expressed by a spatial weight matrix. To construct the spatial weight matrix for 538 zip codes within Massachusetts, we created a queen's contiguity matrix of the first order, which defined neighbors as spatial units sharing a common edge or vertex (Luc Anselin and Rey 2014, 36) and then based the Model 2 on this contiguity matrix.

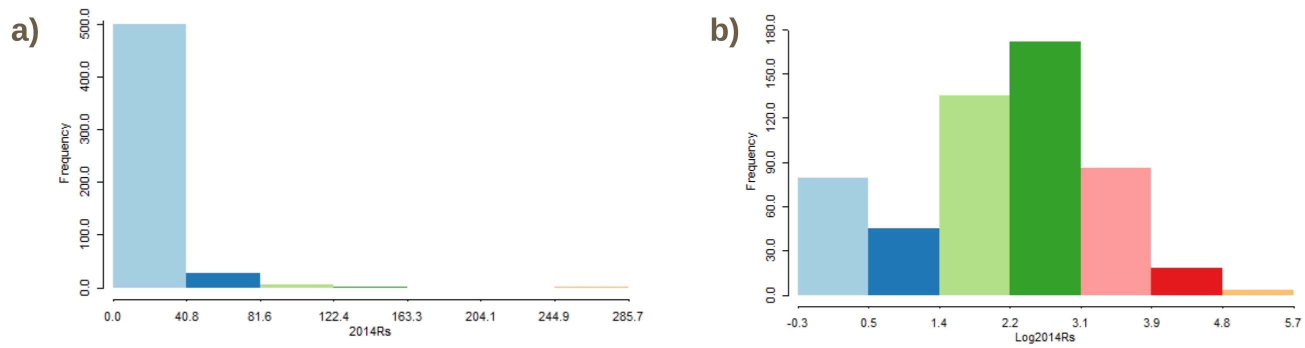
As expected for Model 1 regression coefficients, in Model 2 we expected the log transformed dependent variable to have a positive coefficient that indicates a positive relationship with eight independent variables (poverty level, unemployment rate, child dependency ratio, sex ratio, percent of female householder, tenure ratio, percent of Hispanic population, and percent of African Americans population) and to have a negative coefficient that indicates a negative relationship with the independent variable of residential stability. Moreover, we expected that the coefficient for the spatially lagged dependent variable was positive, which means that the value of the dependent variable for a zip code increases as the value of the dependent variable in neighboring zip codes increases.

We examined measure of fit to see whether Model 2 would be an improvement on Model 1. The standard R^2 measure of fit was not valid for model that incorporated spatial dependence (Luc Anselin and Rey 2014, 193), so we used the maximized log-likelihood and Akaike Information Criterion (AIC) to compare fit between the two models: a model would be better with a larger value for the maximized log-likelihood and a smaller value for AIC. We conducted a Moran's I test on Model 2 residuals to examine whether including the spatially lagged dependent variable in Model 2 would help eliminate the spatial autocorrelation that was found in Model 1 residuals.

Results

Results of OLS Regression Analysis (Model 1)

We log-transformed the outcome variable to get log-transformed child maltreatment rates in 2014 and found that the log transformation reduced the skewness of distribution compared to the original distribution of the outcome variable (see Figure 5). Therefore, the log-transformed outcome variable better conformed with the assumption of normality for OLS regression analysis, and we used it as the outcome variable for regression analysis in this study.



from	0.000	40.816	81.633	122.449	163.265	204.082	244.898
to	40.816	81.633	122.449	163.265	204.082	244.898	285.714
#obs	503	27	5	2	0	0	1
% of total	93.494	5.019	0.929	0.372	0.000	0.000	0.186
sd from mean	0.000	1.251	3.236	5.222	7.207	9.192	11.177

min: 0, max: 285.714, median: 9.93654, mean: 15.0937, s.d.: 20.5598, #obs: 538

from	-0.329	0.526	1.381	2.236	3.091	3.945	4.800
to	0.526	1.381	2.236	3.091	3.945	4.800	5.655
#obs	79	45	135	172	86	18	3
% of total	14.684	8.364	25.093	31.970	15.985	3.346	0.558
sd from mean	-1.362	-0.633	0.000	0.095	0.824	1.553	2.282

min: -0.328584, max: 5.65499, median: 2.29622, mean: 2.12381, s.d.: 1.17308, #obs: 538

Figure 5. Log-Transformation of the Outcome Variable

a) Distribution based on original values of 2014 child maltreatment rates; **b)** Distribution based on log-transformed values of 2014 child maltreatment rates.

We found that Model 1 could generally account for 31.95% (R^2= 0.3195) of variability in the log transformed outcome variable while account for 30.79% (adjusted R^2=0.3079) when considering the number of variables of 10 (see Table 7). At the 99% confidence level ($p<0.01$), we found three independent variables had statistically significant relationships with the log transformed dependent variable: poverty level ($p<0.01$), family headship ($p<0.01$), and residential stability ($p<0.01$).

As expected, the independent variables of poverty level (coefficient=0.0192) and family headship (coefficient=0.0297) presented a positive relationship with the dependent variable. That reflected every 1-unit increase in the percent of income in the past 12 months below poverty level associated to 1.93% increase in child maltreatment rates since exp (0.0192) equals 1.0193, while every 1-unit increase in the percent of female householder with children associated to 3.01% increase in child maltreatment rates since exp (0.0297) equals 1.0301. Different from our expectation, residential stability (coefficient=0.0156) had a positive relationship with the dependent variable, which meant every 1-unit increase in the percent of staying at the same house more than one year related to 1.57% increase in child maltreatment rates since exp (0.0156) equals 1.0157.

However, we found that these estimates based on Model 1 were biased. At a 99% confidence level ($p<0.01$), the results of Breusch-Pagan test indicated statistically significant heteroscedasticity ($p<0.01$) and Model 1 residuals were not normally distributed which was suggested by the result of Jarque-Bera test ($p<0.01$). These results indicated that Model 1 might perform well in some areas in Massachusetts but did not work at all in other areas. The bias of Model 1 was also confirmed by spatial diagnostics (see Table 7): At a 99% confidence level ($p<0.01$), the value of Moran's I of 0.1321 ($p<0.01$) indicated high spatial autocorrelation.

From these spatial autocorrelation patterns presented in Model 1 residuals (see Figure 6), at a 95% confidence level ($p<0.05$), we found "underpredicted areas" expressed by "High-High" clusters (39 zip codes included) that represented the zip codes with high positive-residuals (log-transformed observations larger than log-transformed estimates) were also surrounded by high positive-residuals. We found "overpredicted areas" expressed by "Low-Low" clusters (30 zip codes included) that represented the zip codes with low negative-residuals (log-transformed observations smaller than log-transformed estimates) were also surrounded by low negative-residuals.

We also found "outlier areas" expressed by "High-Low" clusters (11 zip codes included) or "Low-High" clusters (13 zip codes included) that indicated that zip codes with high positive-residuals were surrounded by zip codes with low negative-residuals or the zip codes with low positive-residuals were surrounded by zip codes with high negative-residuals, respectively. All spatial clusters of Model 1 residuals indicate the presence of spatial heterogeneity and the inconsistence of Model 1 performance.

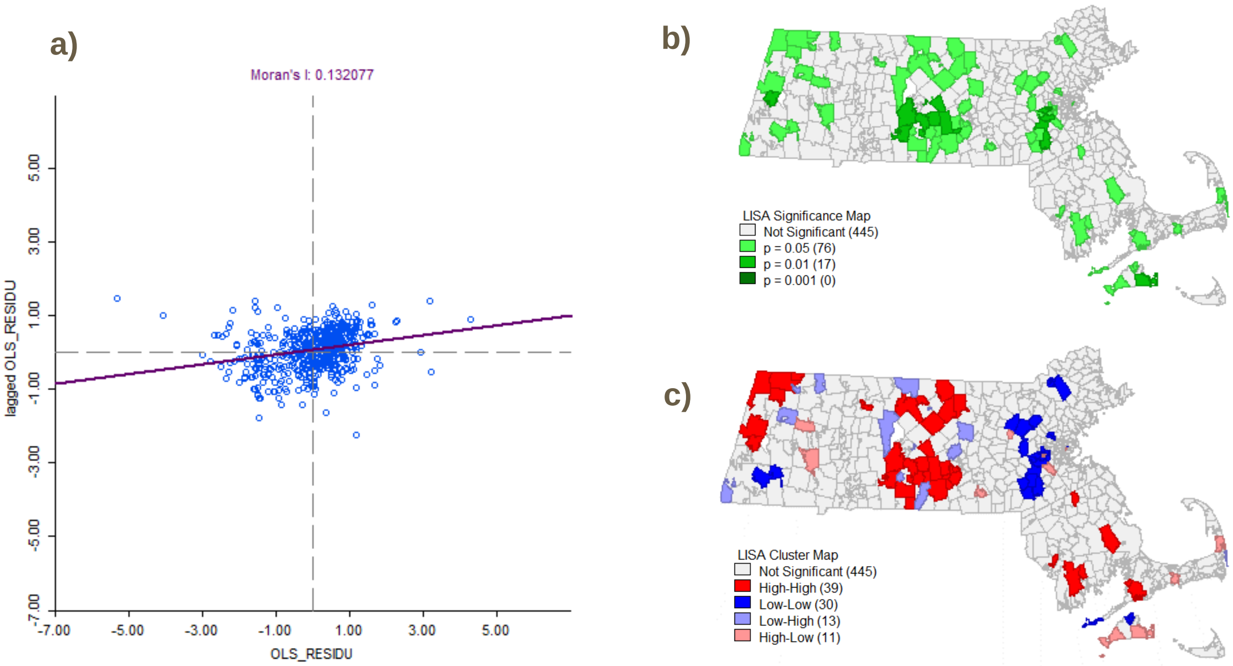


Figure 6. Mapping Spatial Clusters of OLS Regression Residuals with Log-transformed 2014 Child Maltreatment Rates as the Dependent Variable (Massachusetts Zip Codes, n=538): **a)** Moran's I scatter plot; **b)** significance map; and c) cluster map.

To confirm conclusions based on these spatial autocorrelation patterns, we selected alternative specifications based on the results of Lagrange Multiplier tests. At a 99% confidence level($p<0.01$), both standard versions of the LM-Lag and LM-Error tests were highly significant ($p<0.01$), but of the Robust versions, only Robust LM-Lag statistic was significant ($p<0.01$), while the Robust LM-Error statistic was not ($p=0.73395$). Therefore, we selected the spatial lag regression model for spatial regression analysis.

Results of Spatial Lag Regression (Model 2)

At a 99% confidence level($p<0.01$), we found that the spatial autoregressive coefficient was highly significant ($p<0.01$) and had a positive relationship with the dependent variable (coefficient=0.2812) (see Table 7). That means values of the log transformed outcome variable for a zip code is highly associated with the values of the log transformed outcome variable for its neighboring zip codes. Moreover, 32.47% increase in child maltreatment rates for a zip code would relate to one unit increase in mean child maltreatment rates for its neighboring zip codes since exp (0.2812) equals 1.3247.

Like Model 1, Model 2 also presented three regression coefficients that were statistically significant at a 99% confidence level (p<0.01) and had a positive relationship with the dependent variable: poverty level (p<0.01), family headship (p<0.01), and residential stability (p<0.01). The three regression coefficients had slight differences between Model 1 and Model 2, and the coefficients for the three regression coefficients decreased in value from Model 1 to Model 2.

By incorporating spatial effects, Model 2 performed better than Model 1. The higher log-likelihood is associated with a better model (Kennedy 1998, 21). We found that the log-likelihood increased from -745.211 for Model 1 to -731.444 for Model 2 (see Table 7), which meant an improved goodness of fit of a statistical model. We also found a decrease in Akaike Information Criterion (AIC) from Model 1 (1510.42) to Model 2 (1484.89), which suggested an improvement of fit (L. A. Anselin 2017). Furthermore, including the spatially lagged dependent variable helped Model 2 eliminate the spatial effects, which were indicated by the value of Moran's I of -0.0131 for the residuals from Model 2 compared to Moran's I of 0.1321 for the residuals from Model 1 (see Figure 7).

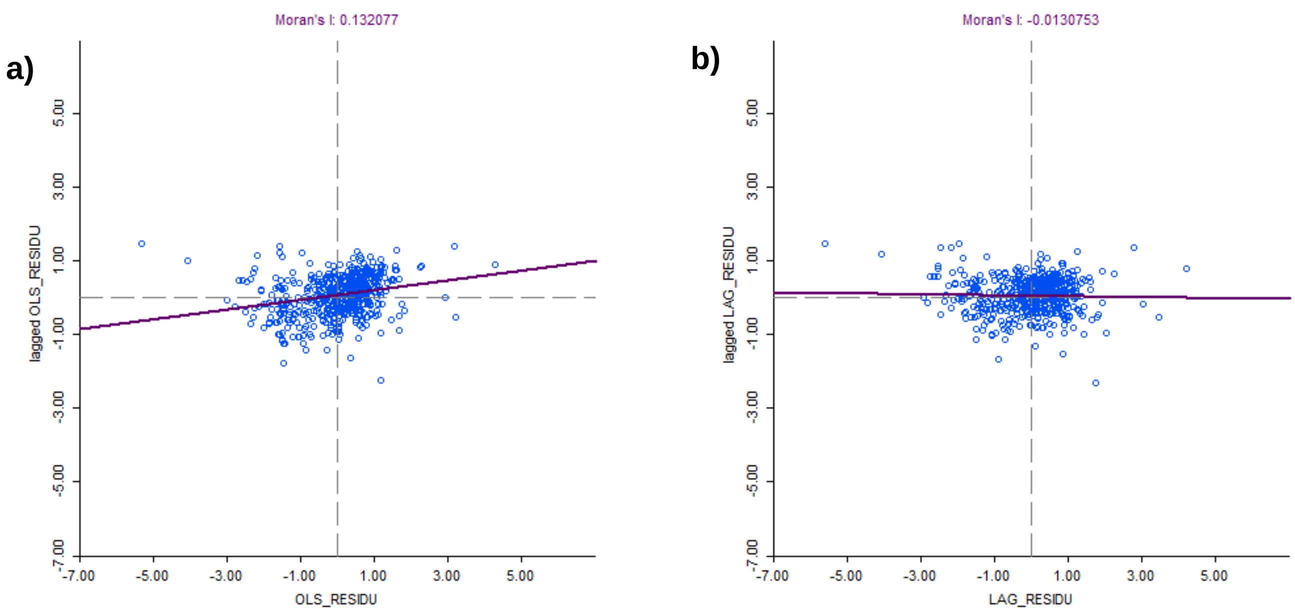


Figure 7. Moran Scatter Plot for Regression Residuals with Log-transformed 2014 Child Maltreatment Rates as the Dependent Variable: a) Classic OLS regression (Model 1); b) Spatial Lag Regression (Model 2).

Table 7. Comparative Regression Results of Model 1 and Model 2

	Model 1: Classic OLS Regression		Model 2: Spatial Lag Regression	
	Coefficient	**Probability**	**Coefficient**	**Probability**
Regression Constant	-0.04007	(0.91998) NS	-0.34315	(0.37980) NS
Spatial Autoregressive Coefficient	NA	NA	0.28122	(0.00000) ***
Independent Variables				
Poverty Level	0.01924	(0.00245) ***	0.01805	(0.00287) ***
Sex Ratio	-0.00036	(0.26523) NS	-0.00035	(0.24812) NS
Child Dependency Ratio	0.00100	(0.79925) NS	0.00395	(0.29581) NS
Family Headship	0.02972	(0.00000) ***	0.02597	(0.00000) ***
Tenure Ratio	0.00001	(0.87383) NS	-0.00001	(0.87514) NS
Residential Stability	0.01561	(0.00071) ***	0.01237	(0.00486) ***
Population Density of Hispanic or Latino Americans	-0.00060	(0.92315) NS	-0.00096	(0.87301) NS
Population Density of African Americans	0.00566	(0.28743) NS	0.00453	(0.37697) NS
R^2	0.3195		NA	
Adjusted R^2	0.3079		NA	
Log Likelihood	-745.211		-731.444	
Akaike Information Criterion	1510.42		1484.89	
Moran's I Value of Residuals	0.1321 (p=0.0000)***		-0.0131 (p=0.0000)***	
Lagrange Multiplier (lag)	28.8462 (p=0.0000)***		NA	
Lagrange Multiplier (error)	0.1155 (p=0.0000)***		NA	
Robust LM (lag)	7.9488 (p=0.00481)***		NA	
Robust LM (error)	0.1155 (p=0.73395) NS		NA	

Note: ***= Significance (p<0.01); NS=Not Significant; NA= Not Applicable.

To confirm these eliminated the spatial effects by Model 2, we also found less zip codes in the "underpredicted areas" from Model 2 (n=25) than from Model 1 (n=39) (see Figure 8). These "underpredicted areas" are expressed by "High-High" clusters that represented the zip codes with high positive-residuals (log-transformed observations larger than log-transformed estimates) were also surrounded by high positive-residuals from spatial autocorrelation patterns presented in residuals at a 95% confidence level ($p<0.05$), including a number of zip codes in Worcester and Berkshire Counties.

However, we also found the significant value of 168.71 (p<0.01) from the results of Breusch-Pagan test for Model 2, which suggested heteroskedasticity was still a problem. This heteroskedasticity can also be indicated by remaining "underpredicted areas" expressed by "High-High" clusters of residuals from Model 2 (see Figure 8, b)), including a number of zip codes in Berkshire, Franklin, Hampshire, Hampden, Worcester, Bristol, Plymouth, and Barnstable. These results indicate that Model 2 still doesn't work well for the "underpredicted areas", so further analyses are warranted for these areas when allocating resources or making policy for preventing child maltreatment at the state level in the future.

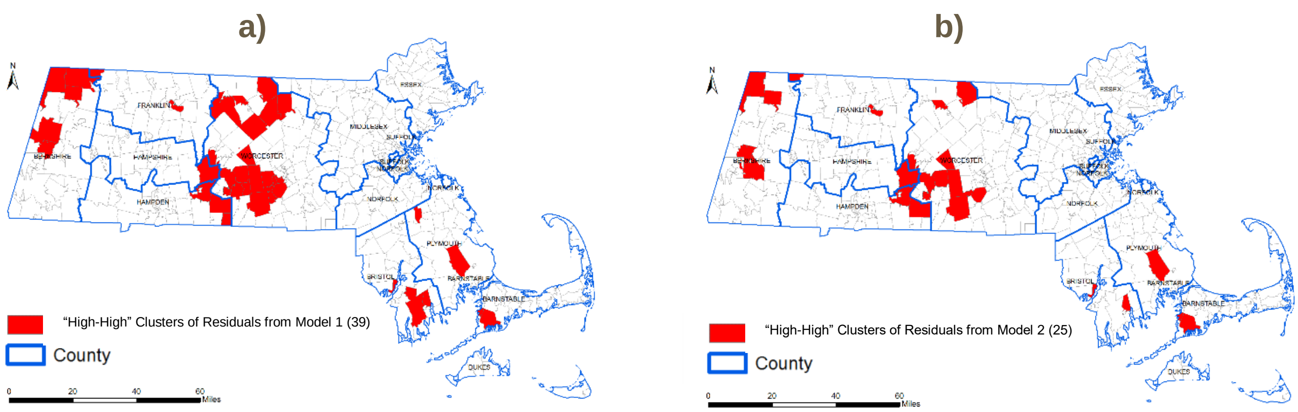


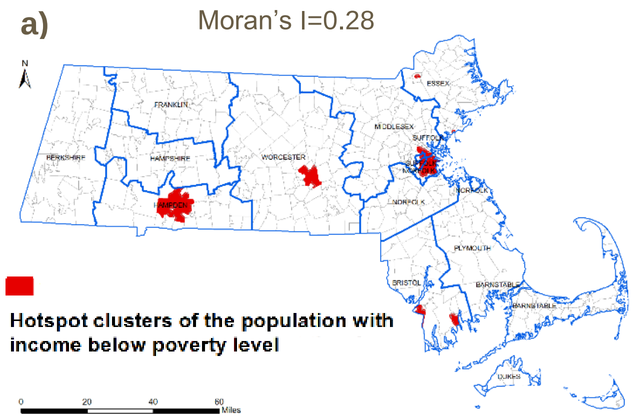
Figure 8. "High-High" Clusters of Regression Residuals with Log-transformed 2014 Child Maltreatment Rates as the Dependent Variable (Massachusetts Zip Codes, n=538): a) "High-High" Clusters of Residuals from Model 1; b) "High-High" Clusters of Residuals from Model 2

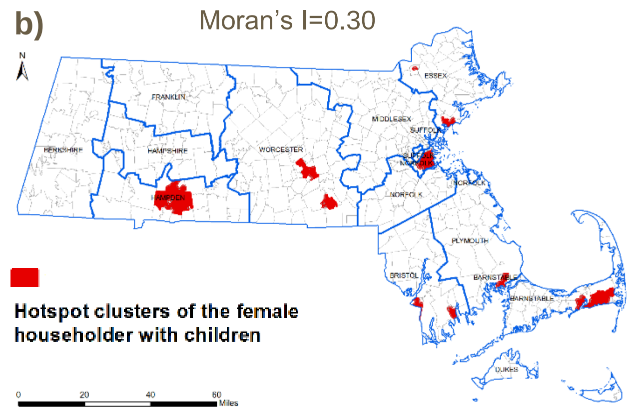
Note: "High-High" clusters of residuals represent the zip codes with high positive-residuals (log-transformed observations larger than log-transformed estimates) are also surrounded by high positive-residuals at a 95% confidence level (p<0.05).

Analysis of covariates that significantly relate to child maltreatment risk

We found three covariates that significantly relate to the outcome variable: poverty, family householder with children, and population living in the same house more than one year, clustered in space across Massachusetts (see Figure 9). We found that hotspot clusters of poverty tended to concentrate in Hampden, Worcester, Essex, Bristol, and neighboring areas between Middlesex, Suffolk, and Norfolk (see Figure 9, a)), while hotspot clusters of female householders tended to concentrate in Hampden, Worcester, Essex, Bristol, Barnstable, and neighboring areas between Middlesex and Suffolk (see Figure 9, b)). Generally, the distributions of the hotspot clusters for poverty and female-householder are similar. On the other hand, the distribution of residential hotspot clusters is different from them (see Figure 9, c)): no residential hotspot clusters in Essex, Middlesex, Suffolk, a number of residential hotspot clusters in Berkshire, Franklin, Hampshire, and Dukes, and more residential hotspot clusters than poverty or female-householder hotspot clusters in Hampden, Worcester, Bristol, and Barnstable.

However, we found that a couple of zip codes in Barnstable and Worcester were located in both hotspot clusters of female householders and hotspot clusters of population living in the same house more than one year. These spatial patterns suggested different findings than past studies that didn't consider spatial effects: areas with a high risk of child maltreatment could relate to a high density of both single mothers with children and families with residential stability.





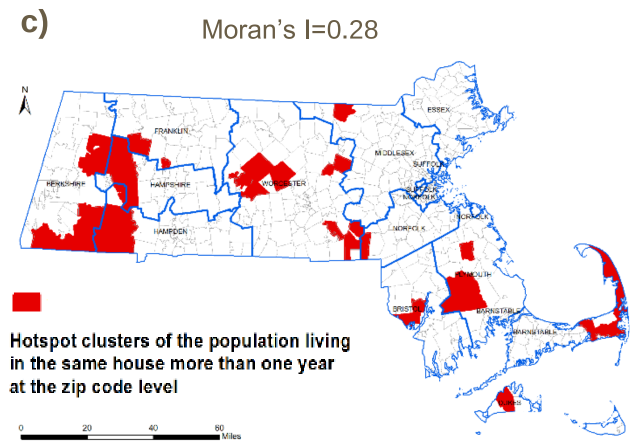


Figure 9. Moran's I and Hotspot Clusters of Three Covariates that Significantly Relate with 2014 Child Maltreatment Rates by Massachusetts Zip Codes (n=538)

a) Population with income below poverty level; **b)** Female house-holder with children; **c)** Population living in the same house more than one year.

Note: The hotspot results are based on a 95% confidence level ($p<0.05$).

Discussion

We confirmed the presence of the spatial autocorrelation in regression residuals from the classic OLS regression model (Model 1), which meant biased estimates based on Model 1 due to the violation of the OLS assumption that no dependence would exist in the test results of OLS regression residuals. We improved the estimates by taking spatial effects into account in Model 2 of a spatial lag regression model and reducing these spatial effects on regression residuals and obtained an apparent improvement in model fit. Furthermore, we identified "unusual locations" (i.e., "underpredicted areas") to warrant further study on localities to allocate resources or make policy for preventing child maltreatment more effectively by considering local needs.

We concluded that child maltreatment rates are not homogeneously distributed at the zip code level across Massachusetts and the risk of child maltreatment for a zip code positively relates to the risk in its neighboring zip codes. Similar to the clustering tendency in space of low-income families, these confirmed spatial effects may relate to grouping forces, which means that individuals or families with common characteristics tend to cluster by choice or due to constrictions to relocate when they are subjected to socioeconomic forces including labor market or local housing (Voss et al. 2006). In other words, families at a high risk of child maltreatment tend to cluster if they share some characteristics such as disadvantaged socioeconomic status. This argument can be reinforced by the relationship between child maltreatment and some community-based factors as follows.

As hypothesized, we found that independent variables of poverty level and family structure in both models have a statistically significant positive relationship with child maltreatment rates. These findings are consistent with existing studies that suggest child maltreatment outcomes have a close tie with poverty (Coulton, Korbin, and Su 1999; Coulton et al. 1995; Garbarino and Sherman 1980; Drake and Pandey 1996; Ernst 2000; Garbarino and Crouter 1978; Garbarino and Kostelny 1992; Kim 2004) and a large proportion of female-headed families that indicate a high child-care burden (Ha, Collins, and Martino 2015; Coulton et al. 1995). Earlier studies didn't take spatial effects into account, so our findings strengthen the argument that higher concentration of low income or single-mother families link to a higher risk of child maltreatment.

However, we didn't find the statistically significant relationship between child maltreatment rates and unemployment rates, child dependency ratio, sex ratio, tenure ratio, population density of Hispanic or Latino Americans, or population density of African Americans. Unexpectedly, we found the independent variable of residential stability had a statistically negative relationship with child maltreatment rates. Child center staff are mandated to report suspected abuse and neglect, so neighborhoods with a higher density of child care centers could have higher rates of early child maltreatment referrals (Stanley and Kovacs 2003). Therefore, families with residential stability might have a bigger chance to learn and build their network with community-based service providers than those with no residential stability and thus less knowledge for these services. If so, it could be difficult for child center staff to learn about or report child maltreatment in these families that didn’t reach the program at all. For future study, the joint effect of the availability of community-based services, the awareness to use these services, and residential stability should be examined for a more confirmative argument.

CHAPTER 4: CONCLUSION

The primary purpose of this study was to examine spatial distribution of child maltreatment across Massachusetts. We identified spatial clusters of child maltreatment counts and rates at the zip code level to provide policymakers with the state-wide landscape of child maltreatment risks. We identified the areas with a high risk of child maltreatment but lack of Family Center placement to inform the CT of candidate site selection for future Family Centers (i.e., Bristol and Worcester). To better understand the mechanisms underlying these spatial patterns, we also compared the estimates from classic OLS regression with the estimates from spatial lag regression analysis: Incorporation of spatial effects in the latter model allowed us to obtain improved estimates of the relationship between child maltreatment risk and community-based risk factors by taking spatial autocorrelation into account. Furthermore, we identified "unusual locations" (i.e., "underpredicted areas") that indicated a need for further research in local areas to improve effectiveness of service allocation and policymaking decisions.

In light of our findings, several limitations should be considered. First, we employed a cross-sectional ecological study design rather than a longitudinal design; we examined the statistical relationship between 2014 child maltreatment rates in Massachusetts and nine community-level risk factors rather than detecting changes in child maltreatment rates for the target population over an extended period of time. Therefore, future research with a longitudinal design is needed to examine whether there is a linear trend and a causal relationship among those community-based risk factors that didn't present a linear relationship with child maltreatment rates in this study. Secondly, we still found heteroskedasticity in the test results of the spatial lag regression analysis, so including new explanatory variables or incorporating different spatial weights will be considered for future research. Thirdly, this study doesn't assess the distribution of existing service resources for supporting families and children across Massachusetts. The geographic distribution of services was found to be related to Child maltreatment referrals in previous research (Freisthler 2011), so it could be possible that some events of child maltreatment were not reported due to a lack of service placement. If this is the case, these under-served areas with a high risk but low reports of child maltreatment should be given special attention for future research.

APPENDIX A: LIST OF “ISLAND POLYGONS” AND THEIR ZIP CODES

“Island Polygons”	Zip Codes
1	01003
2	01029
3	01063
4	01071
5	01073
6	01077
7	01258
8	01338
9	01346
10	01350
11	01718
12	01908
13	01966
14	02128
15	02152
16	02535
17	02543
18	02562
19	02564
20	02657
21	02748

BIBLIOGRAPHY

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Springer Science & Business Media.

Anselin, Luc Anselin. 2017. "Exploring Spatial Data with GeoDaTM : A Workbook." Accessed April 26. http://www.csiss.org/clearinghouse/GeoDa/.

Anselin, Luc, and Sergio J. Rey. 2014. *Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySAL*. Chicago, IL: GeoDa Press LLC.

Anselin, Sridharan, and Gholston, LUC. 2007. "USING EXPLORATORY SPATIAL DATA ANALYSIS TO LEVERAGE SOCIAL INDICATOR DATABASES: THE DISCOVERY OF INTERESTING PATTERNS." *Social Indicators Research* 82 (2): 287–309.

Berger, Lawrence M., Sarah K. Bruch, Sigrid James, Elizabeth I. Johnson, and David Rubin. 2009. "Estimating the 'Impact' of Out-of-Home Placement on Child Well-Being: Approaching the Problem of Selection Bias." *Child Development* 80 (6): 1856–76.

Cardazone, Gina, Angela U. Sy, Ivan Chik, and Laura Kate Corlew. 2014. "Mapping One Strong 'Ohana: Using Network Analysis and GIS to Enhance the Effectiveness of a Statewide Coalition to Prevent Child Abuse and Neglect." *American Journal of Community Psychology* 53 (3–4): 346–56. doi:10.1007/s10464-014-9641-7.

Coulton, Claudia J., David S. Crampton, Molly Irwin, James C. Spilsbury, and Jill E. Korbin. 2007. "How Neighborhoods Influence Child Maltreatment: A Review of the Literature and Alternative Pathways." *Child Abuse & Neglect* 31 (11–12): 1117–42. doi:10.1016/j.chiabu.2007.03.023.

Coulton, Claudia J., Jill E. Korbin, and Marilyn Su. 1999. "Neighborhoods and Child Maltreatment: A Multi-Level Study." *Child Abuse & Neglect* 23 (11): 1019–40. doi:10.1016/S0145-2134(99)00076-9.

Coulton, Claudia J., Jill E. Korbin, Marilyn Su, and Julian Chow. 1995. "Community Level Factors and Child Maltreatment Rates." *Child Development* 66 (5): 1262–76. doi:10.2307/1131646.

Dijken, M. W. van, G. J. J. M. Stams, and M. de Winter. 2016. "Can Community-Based Interventions Prevent Child Maltreatment?" *Children and Youth Services Review* 61 (February): 149–58. doi:10.1016/j.childyouth.2015.12.007.

Drake, Brett, and Shanta Pandey. 1996. "Understanding the Relationship between Neighborhood Poverty and Specific Types of Child Maltreatment." *Child Abuse & Neglect* 20 (11): 1003–18. doi:10.1016/0145-2134(96)00091-9.

Ernst, Joy Swanson. 2000. "Mapping Child Maltreatment: Looking at Neighborhoods in a Suburban County." *Child Welfare* 79 (5): 555–72.

Freisthler, Bridget. 2011. "Need for and Access to Supportive Services in the Child Welfare System." *GeoJournal* 78 (3): 429–41. doi:10.1007/s10708-011-9426-6.

Freisthler, Bridget, Emily Bruce, and Barbara Needell. 2007. "Understanding the Geospatial Relationship of Neighborhood Characteristics and Rates of Maltreatment for Black, Hispanic, and White Children." *Social Work* 52 (1): 7–16.

Garbarino and Sherman. 1980. "High-Risk Neighborhoods and High-Risk Families: The Human Ecology of Child Maltreatment." *Child Development* 51 (1): 188–98. doi:10.2307/1129606.

Garbarino, James, and Ann Crouter. 1978. “Defining the Community Context for Parent-Child Relations: The Correlates of Child Maltreatment.” *Child Development* 49 (3): 604–16. doi:10.2307/1128227.

Garbarino, James, and Kathleen Kostelny. 1992. “Child Maltreatment as a Community Problem.” *Child Abuse & Neglect* 16 (4): 455–64. doi:10.1016/0145-2134(92)90062-V.

Garbarino, James, and Deborah Sherman. 1980. “High-Risk Neighborhoods and High-Risk Families: The Human Ecology of Child Maltreatment.” *Child Development* 51 (1): 188–98. doi:10.2307/1129606.

Getis, Arthur, and J. K. Ord. 1992. “The Analysis of Spatial Association by Use of Distance Statistics.” *Geographical Analysis* 24 (3): 189–206. doi:10.1111/j.1538-4632.1992.tb00261.x.

Ha, Yoonsook, Mary Elizabeth Collins, and David Martino. 2015. “Child Care Burden and the Risk of Child Maltreatment among Low-Income Working Families.” *Children and Youth Services Review* 59 (December): 19–27. doi:10.1016/j.childyouth.2015.10.008.

Kennedy, Peter. 1998. *A Guide to Econometrics*. 4th ed. Cambridge, Mass: MIT Press.

Kim, Jin Seok. 2004. “Neighborhood Effects on the Etiology of Child Maltreatment: A Multilevel Study.” Ph.D., United States -- Texas: The University of Texas at Austin. http://search.proquest.com/docview/305127819/abstract/C0C705639F5C4A32PQ/1.

Marsh, Jeanne C., Thomas A. D’Aunno, and Brenda D. Smith. 2000. “Increasing Access and Providing Social Services to Improve Drug Abuse Treatment for Women with Children.” *Addiction* 95 (8): 1237–47. doi:10.1046/j.1360-0443.2000.958123710.x.

Massachusetts Children’s Trust. 2017. “Introduction of Family Centers.” Accessed April 2. http://childrenstrustma.org/our-programs/family-centers.

Mitchell, Andy. 1999. *The ESRI Guide to GIS Analysis*. Redlands, Calif: ESRI Press.

Onyskiw, Judee E, Margaret J Harrison, Don Spady, and Linda McConnan. 1999. “Formative Evaluation of a Collaborative Community-Based Child Abuse Prevention project1.” *Child Abuse & Neglect* 23 (11): 1069–81. doi:10.1016/S0145-2134(99)00080-0.

Ord, J. K., and Arthur Getis. 1995. “Local Spatial Autocorrelation Statistics: Distributional Issues and an Application.” *Geographical Analysis* 27 (4): 286–306. doi:10.1111/j.1538-4632.1995.tb00912.x.

Paulsen, Derek J. 2004. “No Safe Place.” *Journal of Aggression, Maltreatment & Trauma* 8 (1–2): 63–85. doi:10.1300/J146v08n01_03.

WHO. 2016. “Preventing Child Maltreatment: A Guide to Taking Action and Generating Evidence.” *World Health Organization*. September. http://www.who.int/violence_injury_prevention/publications/violence/child_maltreatment/en/.

Spearly, James L., and Michael Lauderdale. 1983. “Community Characteristics and Ethnicity in the Prediction of Child Maltreatment Rates.” *Child Abuse & Neglect* 7 (1): 91–105. doi:10.1016/0145-2134(83)90036-4.

Stanley and Kovacs, Stanley, Janet and Kovacs, Katie. 2003. “Accessibility Issues in Child Abuse Prevention Services.” *Australian Institute of Family Studies* 18. https://aifs.gov.au/cfca/publications/accessibility-issues-child-abuse-prevention-services.

Stopka, Thomas J., Christopher Krawczyk, Pat Gradziel, and Estella M. Geraghty. 2014. “Use of Spatial Epidemiology and Hot Spot Analysis to Target Women Eligible for Prenatal Women, Infants, and Children Services.” *American Journal of Public Health* 104 (1): S183–89.

UCLA Institute for Digital Research and Education. 2017. “Interpret a Regression Model.” *IDRE Stats*. Accessed May 22. https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faqhow-do-i-interpret-a-regression-model-when-some-variables-are-log-transformed/.

US Census Bureau. 2017. “U.S. Census Bureau.” Accessed May 22. https://www.census.gov/en.html.

U.S. Department of Health & Human Services. 2017. “Child Maltreatment 2015.” https://www.acf.hhs.gov/cb/resource/child-maltreatment-2015.

Voss, Paul R., David D. Long, Roger B. Hammer, and Samantha Friedman. 2006. “County Child Poverty Rates in the US: A Spatial Regression Approach.” *Population Research and Policy Review* 25 (4): 369–91.