Geographic Information System Methodologies and Spatial Analysis in Health and Environmental Disparity

BY

KRISTIN M. OSIECKI
BS, University of Illinois at Urbana, 1989
MS, University of Illinois at Chicago, 2011

DISSERTATION

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Public Health Sciences
(Environmental and Occupational Health Sciences)
in the Graduate College of the
University of Illinois at Chicago, 2013

Chicago, Illinois

Defense Committee:

Seijeoung Kim, Chair and Advisor
Michael Cailas, Environmental and Occupational Health Sciences
John Canar, Environmental and Occupational Health Sciences
Sherry Brandt-Rauf, Environmental and Occupational Health Sciences
Elizabeth Calhoun, Health Policy and Administration

This thesis is dedicated to my husband, Mark Osiecki, and my four children: Casandra, Mark Jr., Chessa, and Victoria. With their incredible support, I am able to fulfill my lifelong dream.

ACKNOWLEDGMENTS

I would like to thank my dissertation committee—Dr. Seijeoung Kim, Dr. Michael Cailas, Dr. John Canar, Dr. Sherry Brandt-Rauf, and Dr. Elizabeth Calhoun—for their incredible support and guidance. Their passion, dedication, and encouragement inspire me to make a difference.

KMO

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF ABBREVIATIONS

CCA	Chicago Community Area
EDA	Exploratory Data Analysis
ESDA	Exploratory Spatial Data Analysis
GIS	Geographic Information Systems
HUD	United States Department of Housing and Urban Development
IDPH	Illinois Department of Public Health
ISCR	Illinois State Cancer Registry
KNN	K-Nearest Neighbor
LISA	Local Indicators of Spatial Association
NATA	Nationa-Scalel Air Toxic Assessment
NEI	National Emissions Inventory
SES	Socioeconomic Status
TRI	Toxic Release Inventory
URE	Unit Risk Estimate
USEPA	United States Environmental Protection Agency

SUMMARY

Geographic information system methodologies were used to visualize and investigate possible spatial relationships between health and environmental disparities indicators using demographic variables, cancer rates and cancer risk in Cook County, IL including the City of Chicago. Data were obtained from the US Census Bureau, the United States Environmental Protection Agency and the Illinois State Cancer Registry. Statistical and spatial analysis was conducted to find possible relationships between these factors in disadvantaged communities to determine social vulnerability and environmental burden.

After looking at numerous variables from each database, there were no linear relationships found, however; chloropleth maps looking at percent African American, percent poverty and non-point cancer risk showed possible clustering. Spatial global autocorrelation results confirmed a spatial component, therefore; local spatial autocorrelation was completed. This analysis provided evidence of clusters on the West and South sides of Chicago between these three variables.

A conceptual methodological framework model was created to better understand the role of spatial analysis in health and environmental disparity research. This three-step approach addressed limitations and challenges associated with spatial analysis with public health data to ensure appropriate interpretations of the outcomes for spatial regression.

I. INTRODUCTION

Health disparity research examines numerous socioeconomic and demographic variables to better understand poor health outcomes of racial minorities living in poverty. There is a new emphasis placed on the built environment to understand surroundings of a neighborhood that may contribute to this disparity. Studies reported that racial/ethnic minorities living in disadvantaged neighborhoods experienced a greater rate of exposure to environmental hazards. Although neighborhood characteristics and the concept of the built environment have been shown to affect individual health, measuring the effects of environmental risks on health has been a less developed area of disparities research.

Early environmental justice research explored a number of hazardous sites in disadvantaged communities. The premise was that more hazardous sites translated to higher exposure to environmental risk factors. The US Environmental Protection Agency (USEPA) acknowledged that past methods analyzed the proximity to sources of environmental hazards. But mapping sites to evaluate spatial clusters lacked evidence between increased exposure and risk. In addition, mapping environmental injustice did not measure the correspondence between the location of potential environmental burdens, exposures, and health effects (1).

Past research presents mixed results when looking at proximity to environmental hazards and at-risk communities. To address this issue, the Symposium on Integrating the Science of Environmental Justice into Decision-Making at the Environmental Protection Agency: An

Overview listed seven areas of interest in environmental justice research to identify factors that contribute to environmental health disparities to better understand environmental burden and social vulnerability (2). One area that held promise was adopting methodologies for estimating population characteristics with emerging geostatistical techniques (e.g., geographically weighted regression) that had the ability to address limitations of conventional approaches (3).

The examination of environmental risk levels; human exposure and proximity to such risks in air, water, and soil; and availability of resources to mitigate the effects of these environmental risks influence health outcomes (4,5) Moving forward with social research, there are numerous considerations that require an understanding of limitations and assumptions of both the data and geographic information system (GIS) software to ensure correct interpretation of outcomes. Geographic information systems are a powerful tool to visualize and analyze spatial relationships especially when traditional statistical analysis does not provide conclusive results. Knowledge of environmental exposure risks, distributional patterns, and their effects on population health require a geographic perspective while investigating social injustices to better understand the causes of health disparities among different populations.

My first hypothesis is that the development of an integrative methodological framework when working with a geographic component is essential in addressing the complexities and challenges associated with environmental health risk and health disparities data and applications. The second is conducting exploratory data analysis and exploratory spatial data analysis which is necessary to determine variables of interest especially when utilizing several databases with varying units of analysis. Lastly, by using this process, GIS methodologies provide evidence that there is a spatial correlation between USEPA National-Scale Air Toxics Assessments (NATA) cancer risk data and health disparity indicators such as percent black and percent poverty in Cook County, Illinois. The goal of this research is to provide a comprehensive approach in analyzing social vulnerability and environmental burden with a spatial component.

II. LITERATURE REVIEW

Studies have documented that racial/ethnic minorities are more likely to live in areas with elevated levels environmental hazards (6-18), which results in increased risks of cancer, respiratory, and neurological diseases (19-31). The risk of exposure to air pollution is also known to be disproportionately elevated among minorities living in poor urban areas (32, 33, 34-36). The racial/ethnic differences in environmental risk exposure may be in part due to social and economic disadvantages among minority populations, which limit types of neighborhoods where they live. For example, Mohai and Saha found that toxic hazard sites were less likely to be located in neighborhoods with a greater proportion of residents who hold higher- ranking and professional occupations compared with neighborhoods with a high rate of residents who have labor and manufacturing jobs (37).

Other studies have reported inverse relationships between household income levels and the risk exposure to air pollutants (38-40). For example, Faber and Krieg argued that income is a strong predictor for the likelihood of living in the proximity of a waste site (41). In addition, they found that residents in lower-income areas are more likely to lack the political power to prevent waste sites from being placed in their neighborhoods (42). Data aggregated at higher levels of government unit (county or city) are less reliable as indicators of disproportionate burdens, and less accurate in identifying the affected populations than data aggregated by smaller units such as census tract (or even blocks). When using larger geographic units, homogeneity within the specific demographics unit cannot be assumed, in which average impact of environmental exposure within geographic boundaries is difficult to determine.

To investigate environmental injustice, GIS have great potential to better understand neighborhood dynamics because of the ability to conduct spatial analysis; however, there are many considerations when looking at data from a geographic perspective. An issue regarding data utilization is the availability of data at the community or neighborhood levels that can be compared across geographic locations (43,44). Also, it is difficult to reconcile data at different geographic units that do not spatially match (such as census tract versus zip code). Neighborhood sociodemographic data retrieved from the US census, which is commonly used in public health research, are aggregated at census tracts or blocks. While epidemiologic studies that examine the distribution and determinants of health can look at illness from larger geographical boundaries such as state or county levels. However, actual health outcome measures are at individual level, which are often not available. Environmental data can either be presented by larger geographical areas such as state or county levels or smaller parcels such as acreage at a manufacturing site. These various geographic boundaries introduce an unavoidable problem of linking, combining, and comparing data from multiple sources.

Also, depending on the level geographic unit of analysis, the distribution of factors, such as poverty or racial composition (45) is often under or overestimated. Unfortunately however, the availability of data often dictates the level of aggregations (46). A major hurdle in conducting our research is working within the constraints of these data sets. Traditional statistical results often fail to provide insights into complex relationships between geography, sociodemographic characteristics, and environmental exposures. For instance, multivariate regression captures the strength and the significance of statistical relationships between the dependent variable and other explanatory factors; however, important local variations between dependent and exploratory variables may not have been understood or examined at the high dimensionality of the datasets (47). Geostatistical modeling involves multivariate data that need to meet the assumption of joint multivariate distribution for valid inference (48). Traditional methods for multivariate visualization such as tables and scatter plots, commonly used to examine health disparities and environmental health data, have been found to be limited in their ability to represent very large datasets (49). Visualization methods have an advantage of assisting in identification and further exploration of certain patterns, thereby generating new analysis that can be created in an easily understood form (50). Various components of an analysis design coupled with the use of domain expertise through interactive exploration can develop into multivariate spatial patterns and the data can be allowed to show the obvious for hypotheses development (51).

Exploratory data analysis (EDA) look at data such as correlations and measures of fit but also needs to be carefully investigated because these results become invalid when there is spatial dependence (52). Exploratory spatial data analysis (ESDA) focuses on spatial aspects of the data to find possible spatial patterns and outliers (53). Spatial analysis software has a statistical pattern-recognition approach which uses a cluster or autocorrelation statistic to quantify a relevant aspect of a spatial pattern (54). However, the term “cluster” in health research is often generic in that it fails to describe spatial variation without a precise description of the statistical test, heterogeneous population sizes, spatial autocorrelation, and non-uniform risks in social science (55).

Exploratory data analysis (EDA) and ESDA provide informative visualizations for interpretable outcomes especially when working with multidisciplinary data with a geographic component. If clustering is identified with global spatial autocorrelation, we can now proceed to the last diagnostic tests of ESDA which involve local autocorrelation. Local autocorrelation is based on the Local Moran's I statistic (56), and can be visualized with significance and cluster maps with a corresponding scatterplot. A sensitivity analysis, which includes running permutations (to as many as 9,999 iterations) and changing the significance level, addresses problems associated with stability due to multiple comparisons with lower significance of the indicated clusters (57).

The use of the integrative methodological framework integrates traditional statistical analysis with spatial autocorrelation to examine the data for possible spatial regression. There are several challenges and limitations working with numerous databases and various statistical and GIS software applications. The objective is to discover underlying assumptions with this process for correct interpretations of outcomes while minimizing mixed results.

III. METHODS

A. Variables

1. National Air Toxic Assessment—Cancer Risk Measurement

The 2005 NATA Total Cancer Risk by census tract is a composite cancer risk score. The total cancer risk is the lifetime risk of developing cancer after exposed to air toxic compounds over a 70-year period. The 2005 NATA data include more than 80 air toxic substances, such as formaldehyde, that are known to be associated with cancer, and 110 air toxins related to non-cancer health outcomes. The total cancer risk score was calculated based on the unit risk estimate (URE) that indicates the probability of developing cancer when a person is exposed to a pollutant with concentration of one microgram per cubic meter of air (58). The risk of exposure to the concentration of a pollutant is then calculated by multiplying the particular concentration by the URE (59).

The total cancer risk was estimated with six subcategories. The subcategories were based on the USEPA emission types, including: point sources, non-point sources, on-road mobile, non-road mobile, secondary formation and decay, and background sources. Point sources are stationary sources, such as large waste incinerators and factories (60). Non-point sources are the stationary sources whose locations cannot be accurately documented. Non-point sources include dry cleaners, burns, and small manufacturers (61). On-road mobile sources include vehicles found on roads and highways. On the other hand, non-road mobile sources are sources not found on roads, such as “airport ground support equipment, trains, lawn mowers, construction vehicles, and farm machinery" (p.19) (62). Background sources include outdoor air toxins resulting from natural sources (63). Secondary formation and decay refers to "air toxics that are caused by the reaction in the environment of emitted primary air toxics" (p.21) (64). Of the six subcategories, the background and the secondary emissions are calculated outside of the National Emissions Inventory (NEI). Although the NEI and NATA may appear similar to each other, the EPA describes NEI as "a major emission inventory that was developed for a range of users and purposes" (65). On the other hand, the NATA emissions inventory was described as a dataset that is specifically compiled and configured for NATA modeling (using NEI as the starting point), and certain procedures are followed to develop the NATA emissions inventory (66).

2. Illinois State Cancer Registry Cancer Outcome Measures

We utilized the all-cancer, leukemia, breast, and lung cancer incidence rates at the census-tract level (zip code level). The incidence data are available from the Illinois Department of Public Health, as part of the Illinois State Cancer Registry (ISCR) (67). From the ISCR zip code file, we retrieved data on sex, cancer site, age, stage at diagnosis, and zip code latitude and longitude. Age at diagnosis was grouped into: 0–14 years, 15–44 years, 45–64 years, and 65 years and older. Stage at diagnosis included: in situ, localized, regional, distant metastases, and unknown or unstaged. The ISCR reports five-year cumulative incidence rates: periods from 1990–1994, 1995–1999, 2000–2004, and 2005–2009.

3. U.S. Census—Demographic Variables

We retrieved sociodemographic data from the 2000 US Census. All variables were measured at the census-tract level. Variables included in the analysis were: population density, proportional age distribution, median household income, percent of residents living below federal poverty line, percent of residents with less than high school education, percent of African Americans, percent of Hispanics, percent of vacant buildings, percent home ownership, and median housing value. To control for the effect of age on cancer outcomes, we used percent of individuals 55 years or older. Changes between the census of 2000 and that of 2010 hinders accurate comparison and follow-up of changes in health disparities at the census-tract level or Chicago Community Area level. There are services that reconcile these boundaries but researchers need to be sensitive to potential bias within the data when using the modified data set.

B. Analysis

We anticipated data issues upfront so we could formulate a strategy to approach the analysis with a clear grasp of challenges and limitations associated with each dataset. We concluded that Illinois urban counties such as Cook County where the city of Chicago resides, there were four major challenges with databases that examined data vertically (over time) as well as horizontally (multi-level):

1. There was a considerable lag between data collection and dissemination.
2. The US census tracts had significantly changed between 2000 and 2010.
3. Cancer incidence, mortality data, and risk factors were reported in varying units (census tract, zip code, and Chicago Community Area levels).
4. There were numerous limitations and assumptions associated with environmental health risk assessment and point source data.

We began with EDA and descriptive statistics to examine the distributions of dependent and independent variables. Correlations between continuous demographic variables, total cancer risk, number of hazardous site, and cancer incidence rates were examined. Second, we used linear regression models to explore relationships between the outcome measures and demographic and environmental risk factors. Additional EDA, including scatterplots and parallel coordinate plots visualized multivariate and bivariate relationships as a precursor to investigating spatial randomness. The ESDA involving global spatial autocorrelation searched for possible clustering to decide if a spatial relationship exists and if so, to justify further research with local spatial autocorrelation and spatial regression.

First, we performed EDA using variables of interest that best captured vulnerable populations and environmental burden within Cook County, Illinois. This list was extensive, utilizing health disparity indicators such as race/ethnicity and socioeconomic status (SES) and NATA environmental cancer risk data divided into six categories. Stata software was used for descriptive statistics to examine the distributions of dependent and independent variables, correlations between continuous demographic variables, the total cancer risk derived from NATA and cancer incidence rates, bivariate models to explore relationships between the outcome measures, and demographic and environmental risk factors.

We then used ArcGIS and OpenGeoDaA for EDA visualization in identifying variables and areas warranting further examination. Based on the results of EDA and parallel coordinate plots, we tailored the list to key variables to model a series of global spatial autocorrelations. We wanted to determine whether clustering existed and if so, to move forward with future local autocorrelation and spatial regression. Census tracts that showed possible correlation were then brushed and linked to box plot maps for visual comparison. In OpenGeoDda, brushing and linking was a technique that allowed us to highlight points of interest on a graph and then linked or identified these points on corresponding maps and charts.

We created chloropleth maps in ArcGIS software based on the variables of interest outlined above to visualize descriptive statistics. This included scatterplots and parallel coordinate plots, looked at multivariate and bivariate relationships as a precursor to investigating spatial randomness in OpenGeoDda software. Although these tools provided additional insight into the data, there were no definitive relationships between variables of interest when we looked at the entire geographical area. Therefore, we decided to narrow the scope of our geography by examining census tracts located in the West and South regions of the city of Chicago.

We used spatial weights based on a queen matrix rook to take into consideration contiguity issues. Spatial autocorrelation, Global Moran's I with permutation inference, examined negative and positive spatial correlation with a standardized z-value. The goal of positive and negative spatial autocorrelation was to investigate similar and dissimilar values in relation to location. The spatial autocorrelation statistic captured both attribute and location similarity but it was important that the z-statistic was not interpreted as statistical significance. This method investigated clustering to decide if a spatial relationship existed and if so, to justify further research with local spatial autocorrelation and spatial regression.

We completed a sensitivity analysis to ensure robustness within these three parameters. We then created a weight neighbor histogram for a queen weights matrix and found that we had no islands (isolated locations with no neighboring areas), binomial distribution, or a high number of census tracts with very large number of neighbors. Our maps were based on 999 permutations to avoid too much sensitivity on the particular randomization that was run multiple times until our results stabilized. The other diagnostic test we performed was permutation reference histogram to ensure that our data are non-random. Our observed distribution was significantly different from the expected distribution, thus we concluded that our data were not random.

We conducted univariate and bivariate LISA on 1,343 census tracts located within Cook County,comprising of the city of Chicago. With the queen weight matrix, we then randomized the data with 999 permutations. Initial sensitivity maps contained pseudo p-values ranging from 0.05 to 0.0001; however, results for p-value=0.05 could be unreliable. Our sensitivity maps were then adjusted to only include p-values 0.01 to 0.0001. With the bivariate local indicators of spatial association (LISA) scatterplot, we calculated the descriptive statistics, and ran the Chow test, which looks at the slope and the intercept of one group to see if they are different from those of another group. This test utilizes standard errors from each line and degrees of freedom to calculate apartial f-test. Positive results confirmed that the data were conducive for possible future research using geographic weighted regression.

IV. MANUSCRIPTS

A. Spatial Ecology of Neighborhood Sociodemographic Characteristics and Environmental Risk

Health disparity research examines numerous socioeconomic and demographic variables to better understand poor health outcomes of racial minorities living in poverty. There is a new emphasis placed on the built environment to understand surroundings of a neighborhood that may contribute to this disparity. Past research presents mixed results when looking at proximity to environmental hazards and at-risk communities. The role of EDA and ESDA needs to be investigated to determine variables of interest especially when utilizing several databases with varying units of analysis. The findings confirm that GIS is a powerful tool to visualize and analyze spatial relationships especially when traditional statistical analysis does not provide conclusive results. This paper examines data sources, methodology, and analysis to understand the complexities and challenges associated with data and applications when incorporating environmental health risk into health disparity research. Careful assessment of the data, software, and analytic methods is required to properly interpret the findings in understanding the associations between sociodemographic characteristics and environmental burden.

1. Introduction

Studies have documented that racial/ethnic minorities are more likely to live in areas with elevated levels environmental hazards (68-81), which results in increased risks of cancer, respiratory, and neurological diseases (82-94). For example, Schulz and others (95) reported that minority neighborhoods in Detroit are more likely to be located close to highways. Pastor and colleagues found that Hispanics were more likely compared with their White counterparts to live close to the Toxics Release Inventory (TRI) sites in California (96).

Similarly, in New York City, the re-zoning process between 1961 and 1998 disproportionately increased manufacturing zones in minority neighborhoods (97,98). The risk of exposure to air pollution is also known to be disproportionately elevated among minorities living in poor urban areas (99-101).

The racial/ethnic differences in environmental risk exposure may be in part due to social and economic disadvantages among minority populations, which limit types of neighborhoods where they live. For example, Mohai and Saha found that toxic hazard sites were less likely to be located in neighborhoods with a greater proportion of residents who hold higher-ranking and professional occupations, compared with neighborhoods with a high rate of residents who have labor and manufacturing jobs (102). Other studies have reported inverse relationships between household income levels and the risk exposure to air pollutants (103-105). For example, Faber and Krieg argued that income is a strong predictor for the likelihood of living in the proximity of a waste site (106). In addition, Faber and Krieg found that residents in lower-income areas are more likely to lack the political power to prevent waste sites from being placed in their neighborhoods (107). In fact, even after the initial exposure, the clean-up of hazardous sites has been known to be quicker in predominantly White neighborhoods compared with minority neighborhoods (108). Historically, industrial development and discriminatory redlining practices have resulted in systematic social exclusion of minorities for many years (109-115).

Political decisions on land use, zoning, and industrialization are not value-free (116, 117), resulting in unequal adverse effects on racial/ethnic minorities living in poverty areas (118-121).

Earlier studies (122.123) showed that that race/ethnicity was a consistent significant predictor for the exposure to environmental risks, controlling for income/poverty. Neighborhoods with low SES and high rate of racial residential segregation tend to experience low social integration and greater financial difficulties, which thereby affects a community's collective resources and capacity to deal with their neighborhood issues (124-128). Such collective capacity, often described as social capital, enables residents to achieve resources that are necessary for collective action to solve shared issues (129-134). This suggests that disadvantaged minority neighborhoods with low social capital are less likely to avoid environmental hazards in their neighborhood. Furthermore, when neighborhoods become less environmentally safe, individuals who have means to relocate may leave the neighborhood (135). It may be that disadvantaged minority neighborhoods lacking social capital would be more likely to be exposed to environmental hazards. And, increased environmental risks in neighborhoods would then drive out middle-class residents, resulting in even worse social capital in recent years: a vicious cycle of environmental disparities and neighborhood disadvantage.

Neighborhood social and physical environmental contexts often determine one's exposure to health risks and health outcomes (136-141). Epidemiologic studies support the association between geographic locations of pollution sources and the occurrence of various cancers including leukemia, lymphoma, breast, lung, brain, thyroid, endocrine, and skin cancers (142-146), which point out the persistent existence of inequalities in environmental health among poor neighborhoods (147). Studies have shown that this type of systematic and structural disadvantage among racial/ethnic minority communities exacerbates the existing racial disparities in health (148-151). Poor communities tend to lack access to parks, grocery stores, and health care facilities (152), and suffer from substandard quality housing and overcrowded living conditions (153). The built environment, whether it is the physical structures of communities, land use, access to healthcare, social services, grocery stores, parks, or proximity to air, water, or soil contaminants, profoundly affects the health status of individuals living within (154,155).

However, the spatial relationship between sociodemographic characteristics and environmental exposure has not fully been explored (156-158). Conflicting findings have been reported regarding the relationships between race/ethnicity, SES, and environmental exposure (159-161). These studies suggest that either mediating factors that determine the underlying mechanisms as to how race/ethnicity and SES are associated with environmental exposure; or the effects of race/ethnicity and SES on the risk exposure may differ by other factors. To explore such complex relationships between environmental risks and neighborhood sociodemographic characteristics, we examined the level of environmental hazards, racial/ethnic composition, SES, and health outcomes; and identified potential mediators between environmental risks and health outcomes.

In this paper, we utilize a health disparities conceptual framework to better describe the nature of the relationship between environmental health risk and health outcomes and its mediating factors. We then introduce a methodological framework to understand the role of EDA and ESDA when using multiple data sets. Such corresponding conceptual and methodological models offers a comprehensive analytic framework to examine multiple factors and the challenges associated with the selection of databases, the identification of variables of interest, implementation of methodologies, and interpretations of outcomes.

We focused on health disparities and the built environment in Cook County, Illinois which includes the city of Chicago. Chicago ranked as the 4th most racially segregated city in the United States(162), and Black and White disparities in Chicago were worse on seven of eleven health status indicators, including breast and lung cancer mortality rates(163,164). Orsi et al. examined the progress of Chicago in meeting the Healthy People 2010 goal of eliminating health disparities that concluded that racial differences in eleven out of the fifteen measures were widening (165). We are concerned of the backward progress of Chicago requiring further study at the local level, which offers the possibility of promoting action (166).

2. Health and Environmental Disparity Conceptual Model

Figure 1 summarizes the preliminary conceptual framework, with which we hypothesize that mechanisms of differential distributions of physical and social environmental factors mediate effects of social determinants on health outcomes. Environmental health studies have explored associations between exposure to environmental hazards and health outcomes, and previous studies have documented that racial/ethnic minorities living in poverty areas are more likely to be exposed to hazardous environments and to exhibit negative health outcomes. And yet, potential mechanisms that explain how predominantly poor minority neighborhoods may suffer from higher risks of exposure to environmental hazards, lack health care facilities, and poor health outcomes require further research.

This preliminary framework adopts a fundamental cause perspective: we consider that environmental health disparities, risk exposure, and lack of access to care facilities are determined by more upstream social determinants, such as racial residential segregation and neighborhood disadvantage. In addition to direct effects of environmental hazards on health outcomes, the indirect effects of environmental hazards on health exist and may interact with and be mediated by other social and physical environmental factors. Drawing from multi-disciplinary neighborhood research findings, this model is designed to investigate the complex relationships between social determinants, environmental conditions, and health outcomes.

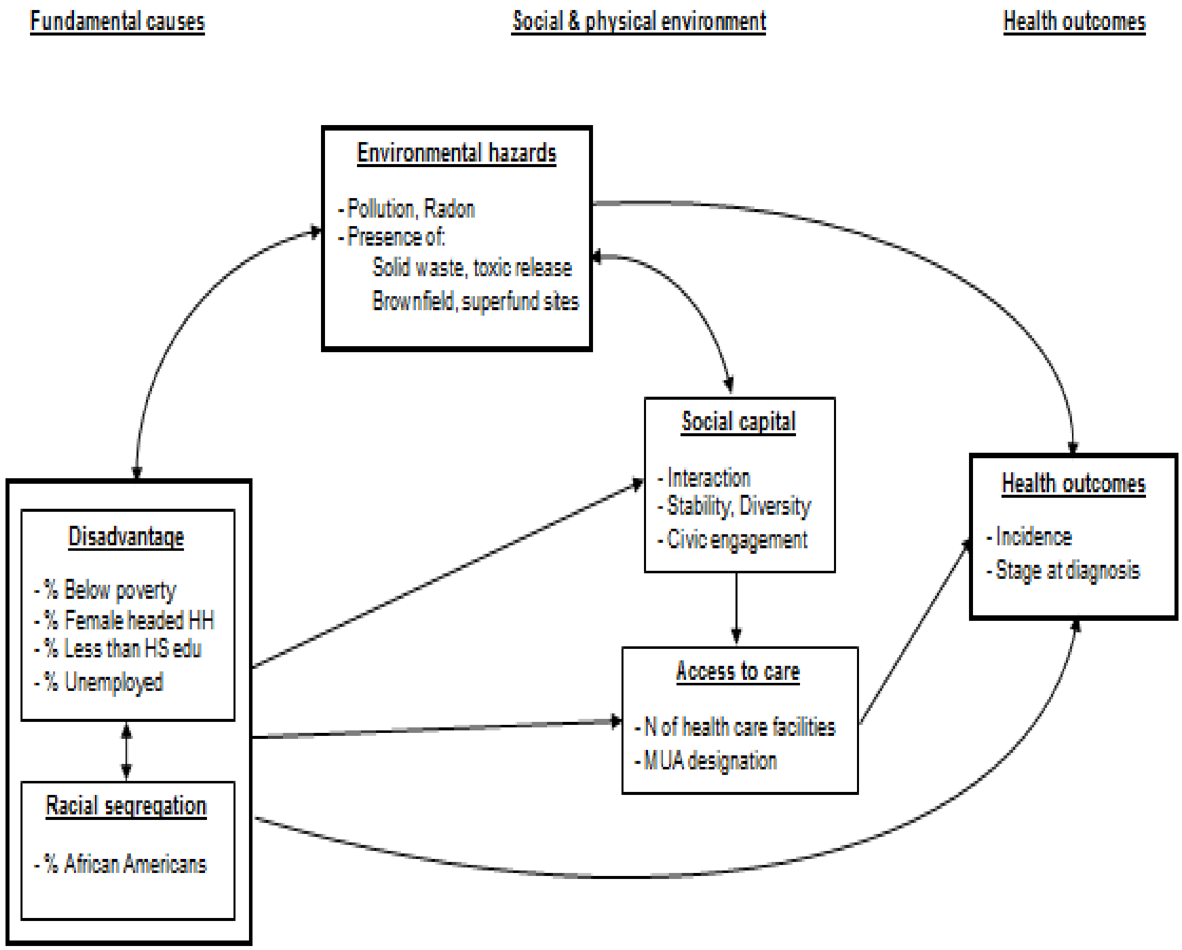

Figure 1. Conceptual and analytic framework: interactions between social determinants, environmental conditions, and health outcomes.

3. <u>Data Sources</u>

Before testing the conceptual model, we needed to make decisions about types of data and level of geographical analysis. Generally speaking, data aggregated at higher levels of governmental unit (county or city) are less reliable as indicators of disproportionate burdens, and less accurate in identifying the affected populations than data aggregated by smaller units such as census tract (or even blocks). When using larger geographic units, homogeneity within the specific demographics unit cannot be assumed, in which average impact of environmental exposure within geographic boundaries is difficult to determine. Also,the level geographic unit of analysis and the distribution of factors, such as poverty or racial composition (167) are often under or overestimated. Unfortunately however, the availability of data often dictates the level of aggregations (168). A major hurdle in conducting our research is working within the constraints of these data sets.

We synthesized perspectives from various disciplines, including epidemiology, environmental health, demography, and social science. An issue regarding data utilization is the availability of data at the community or neighborhood levels that can be compared across geographic locations (169,170). Also, it is difficult to reconcile data at different geographic units that do not spatially match (such as census tract versus zip code). Neighborhood sociodemographic data retrieved from the US census, which is commonly used in public health research, are aggregated at census tracts or blocks.Epidemiologic studies examine the distribution and determinants of health and can look at illness from larger geographical boundaries such as state or county levels. However, actual health outcome measures are at individual level, which are often not available. Environmental data can either be presented by larger geographical areas such as state or county levels or smaller parcels such as acreage at a manufacturing site. These various geographic boundaries introduce an unavoidable problem of linking, combining, and comparing data from multiple sources.

Linking data in a consistent and temporal manner can help determine whether the patterns of disease distribution or trends overtime are true or spurious. This requires that the data be consistent over time (171). However, the critical issue regarding data utilization is that the geographic units change over time. In fact, the census tracts (tract number as well as boundaries) have significantly changed from the US 2000 census to the 2010 census, which introduces introduces another underlying problem.

With all these considerations, our research study obtained data from the USEPA, the US Census Bureau, and the Illinois Department of Public Health (IDPH), as part of the ISCR. We decided to analyze two databases from the USEPA that examined air quality: environmental risk measures from NATA that used models to determine increased cancer risk due to inhalation exposure; and TRI locations that were point sources that look at proximity to hazard. The following discussion describes the types of information provided from each source but also the challenges associated with them.

4. Variables

a. National Air Toxic Assessment—Cancer Risk Measurement

The 2005 NATA Total Cancer Risk by census tract is a composite cancer risk score. The total cancer risk is the lifetime risk of developing cancer after exposed to air toxic compounds over a 70-year period. The 2005 NATA data include more than 80 air toxic substances, such as formaldehyde, that are known to be associated with cancer, and 110 air toxins related to non-cancer health outcomes. The total cancer risk score was calculated based on the URE that indicates the probability of developing cancer when a person is exposed to a pollutant with concentration of one microgram per cubic meter of air (172). The risk of exposure to the concentration of a pollutant is then calculated by multiplying the particular concentration by the URE (173).

The total cancer risk was estimated with six subcategories. The subcategories were based on the USEPA emission types, including: point sources, non-point sources, on-road mobile, non-road mobile, secondary formation and decay, and background sources. Point sources are stationary sources, such as large waste incinerators and factories (174). Non-point sources are the stationary sources whose locations cannot be accurately documented. Non-point sources include dry cleaners, burns, and small manufacturers (175). On-road mobile sources include vehicles found on-roads and highways. On the other hand, non-road mobile sources are sources not found on-roads, such as "airport ground support equipment, trains, lawn mowers, construction vehicles, and farm machinery" (p. 19) (176). The background sources include outdoor air toxins resulting from natural sources (p.20) (177). Secondary formation and decay refers to "air toxics that are caused by the reaction in the environment of emitted primary air toxics" (p.21) (178).

Of the six subcategories, the background and the secondary emissions are calculated outside of the NEI. Although the NEI and NATA may appear similar to each other, the EPA describes the NEI as "a major emission inventory that was developed for a range of users and purposes" (179). On the other hand, the NATA emissions inventory was described as a dataset that is specifically compiled and configured for NATA modeling (using NEI as the starting point), and certain procedures are followed to develop the NATA emissions inventory (180). The EPA document entitled "The Framework for Cumulative Risk Assessment" provides a generalized conceptual model that includes sources, stressors, exposure routes, receptors, and endpoints to determine possible health risks. With this generalized model, NATA incorporates measures from this health risk assessment to quantify cancer risk using three different air dispersion models. This document emphasizes that the data should be used with caution, fully understanding the EPA's processes, with which the risk assessment data was compiled (181).

The NATA data are provided at the census-tract level for those who wish to conduct technical analysis and comparison. However, the EPA also stipulates that NATA assessments should not be used as a definitive means to identify specific risk values within a census tract, or that these results are considered more meaningful at the state or national level (182).

Another limitation is that NATA data across the four time points are not comparable: EPA has changed their methodology for each assessment due to advancing techniques for more accurate modeling. The NATA is ultimately chosen as our environmental health factor because it is one of the few available datasets that quantifies exposure cancer risk and at the census-tract level. However, time and space analysis is impossible because changes cannot be discerned from the improvements or actual changes in emissions or source characterization (183).

b. Toxics Release Inventory—Proximity to Hazards

The TRI database is compiled and updated by the EPA, which includes information on quantities in pounds of toxic chemicals released from TRI facilities to the environment (184). The TRI database provides data on the location of TRI facilities, the level of toxicity of the toxic chemicals, and total amount of release. The TRI sites and cleanup efforts have continued to evolve since 1987, and the EPA updates the data regularly. Data files are available from 1987 to 2010. We geocoded the addresses using the Arc GIS version 10, and assigned a census-tract number to each site. We then calculated the total number of TRI sites per census tract, and normalized the number by the acreage of each census tract.

Early environmental justice research explored a number of hazardous sites in at-risk neighborhoods with the premise that more hazardous sites translate to higher exposure to environmental risk factors (185). The goal of TRI was to provide public information about the release of more than 650 toxic chemicals from more than 20,000 US industrial facilities into the air, surface water, and land. However, a major limitation was that TRI data did not reflect risks to human health and the environment (186). We compiled TRI sites from 1990 to 2010 into a master database with duplicate sites removed. The mapping of TRI sites was helpful to identify manufacturing areas especially in disadvantaged communities; however; it did not provide detailed information on the type and quantity of emissions that varied by each site.

c. Illinois State Cancer Registry—Cancer Outcome Measures

We utilized the all-cancer, leukemia, breast, and lung cancer incidence rates at the census-tract level (zip code level). The incidence data are available from the IDPH, as part of the ISCR (187). From the ISCR zip code file, we retrieved data on sex, cancer site, age, stage at diagnosis, and zip code latitude and longitude. Age at diagnosis was grouped into: 0–14 years; 15–44 years; 45–64 years; and 65 years or older. Stage at diagnosis included: in situ, localized, regional, distant metastases, and unknown or unstaged. The ISCR reports five-year cumulative incidence rates: periods from 1990–1994; 1995–1999; 2000–2004; and 2005–2009.

Publicly available cancer outcomes data are often three- to five-years behind. Researchers request individual-level cancer cases and mortality with address (researchers then geocode data to examine census tract or community-level cancer incidence and mortality) to the state cancer registry. This data application is at least a 12-month-long process due to limited state health department resources and personnel. Moreover, researchers inadvertently duplicate data requests so the state repeats data reviewing and preparing efforts.

Census tracts and zip codes do not match geographically: many key cancer data are reported at zip code, while demographic data are reported by census tract by the US census bureau. In addition, 77 Chicago Community Areas are used in many studies and health care agencies providing a clear sense of neighborhood. Chicago Community Areas are made up of several adjacent census tracts and thus do not match with the geographic division of zip codes. Due to the fact that census tract and zip code areas do not match—and currently many health outcomes measures are only available at zip code level—we reconciled the two area-level measures by using the United States Housing and Urban Development (HUD) crosswalk files (188), which allow the conversion between census tract and zip code areas.

d. US Census—Demographic Variables

We retrieved sociodemographic data from the 2000 US Census. All variables were measured at the census-tract level. Variables included in the analysis were: population density, proportional age distribution, median household income, percent of residents living below federal poverty line, percent of residents with less than high school education, percent of African Americans, percent of Hispanics, percent of vacant buildings, percent home ownership, and median housing value. To control for the effect of age on cancer outcomes, we used percent of individuals 55 years or older. Changes between the 2000 census and that of 2010 hinders accurate comparison and follow-up of changes in health disparities at the census-tract level or Chicago Community Area level. There are services that reconcile these boundaries but researchers need to be sensitive to potential bias within the data when using the modified data set.

5. Analysis

We anticipated data issues upfront so we could formulate a strategy to approach the analysis with a clear grasp of challenges and limitations associated with each dataset. We concluded that Illinois urban counties such as Cook County there were four major challenges with databases that examined data vertically (over time) as well as horizontally (multi-level):

1. There was a considerable lag between data collection and dissemination.
2. The US census tracts had significantly changed between 2000 and 2010,
3. Cancer incidence, mortality data, and risk factors were reported in varying units (census tract, zip code, and Chicago Community Area levels).

4. There were numerous limitations and assumptions associated with environmental health risk assessment and point source data.

We began with EDA and descriptive statistics to examine the distributions of dependent and independent variables. Correlations between continuous demographic variables, total cancer risk, number of hazardous site, and cancer incidence rates were examined. Second, we used linear regression models to explore relationships between the outcome measures and demographic and environmental risk factors. Additional EDA, including scatterplots and parallel coordinate plots visualized multivariate and bivariate relationships as a precursor to investigating spatial randomness. Global spatial autocorrelation searched for possible clustering to decide if a spatial relationship exists and if so, to justify further research with local spatial autocorrelation and spatial regression.

6. Results

Initially, we approached the data with traditional statistical methods including univariate and bivariate analysis. Using several sociodemographic factors such as income, race, and education; along with risk variables including cancer rates, NATA cancer risk, and hazards per square mile, we produced several outcomes including the Pearson Correlation coefficient (Table I). As expected, we saw relationships between the sociodemographic variables, such as percent Black and percent poverty. However, there were no significant associations between explanatory variables and cancer rates or cancer risk.

TABLE I.

PEARSON CORRELATIONS OF SELECTED VARIABLES, COOK COUNTY, IL

	NATA cancer risk	Point source	% Black	% White	% Poverty	Total cancer rate
NATA cancer risk	1.00					
Point source	0.35**	1.00				
% Black	-0.07*	0.80	1.00			
% White	-0.05	-0.07*	-0.90**	1.00		
% Poverty	0.17**	0.12**	0.61**	-0.70**	1.00	
Total cancer rate	-0.31**	-0.25**	-0.23**	0.36**	-0.41**	1.00

We then narrowed the scope of the data to include only the select sociodemographic variables (race and income), proximity to hazard, and cancer risk. Demographic variables were then divided into percent quantile. Table II described the mean hazard point sources per square mile, and the mean NATA total cancer risk for each demographic percentile. Overall, neighborhoods falling into the lowest percentile of the proportion of African American residents had a significantly lower number of hazard sources and a lower average cancer risk compared with the highest percentile of African Americans. Similarly, the mean hazard sources per square mile and the total cancer risk were lower for the census tracts with the lowest percentile of Hispanic residents compared with the highest percentile of Hispanics. Both risk measures were also lower in the lowest poverty quartile compared with the highest poverty quartile.

The opposite pattern was observed for the proportion of Whites and the median household income. Neighborhoods in the lowest quartile of percent Whites and the median income had the highest hazard sources per square mile and the highest cancer risk. However, these patterns did not hold for the middle two quartiles (25th to 75th percentile).

Although these results showed some promise, they still did not provide the conclusive evidence that environmental cancer risk had an impact on health disparity outcomes. Therefore, we created chloropleth maps to investigate possible visual patterns within the data. Although this would not give conclusive analytical results, it was an opportunity to look at the data from another viewpoint.

TABLE II.

CENSUS TRACT LEVEL DEMOGRAPHIC CHARACTERISTICS, COOK COUNTY, ILLINOIS

	2010	Mean hazard per sqml	Mean NATA total cancer risk
Mean percent African American	33.0	-	-
Median	6.1	-	-
Percentile			
0–25th	0.6	76.7	5.87
25th–50th	3.1	117.0	6.58
50th–75th	32.0	125.6	6.36
75th–100th	96.4	102.5	6.12
Mean percent Hispanic	18.9	-	-
Median	6.1	-	-
Percentile			
0–25th	1.0	92.0	5.99
25th–50th	3.9	77.5	6.01
50th–75th	13.2	111.0	6.38
75th–100th	57.6	141.3	6.54
Mean percent White	50.1	-	-
Median	57.4	-	-
Percentile			
0–25th	15.9	107.4	6.16
25th–50th	38.0	135.1	6.44
50th–75th	71.5	114.3	6.44
75th–100th	91.1	65.0	5.89
Mean percent Poverty	16.3	-	-
Median	10.6	-	-
Percentile			
0–25th	2.8%	63.0	5.65
25th–50th	7.6%	103.2	6.30
50th–75th	16.4%	123.6	6.55
75th–100th	38.2%	131.9	6.42
Mean HH Income ($)	45,153	-	-
Median	42,154	-	-
Percentile			
0–25th	20,922	128.9	6.38
25th–50th	37,279	120.8	6.35
50th–75th	47,885	84.7	6.15
75th–100th	74,521	87.7	6.04

Hazard: all point sources per square mile (*100,000)

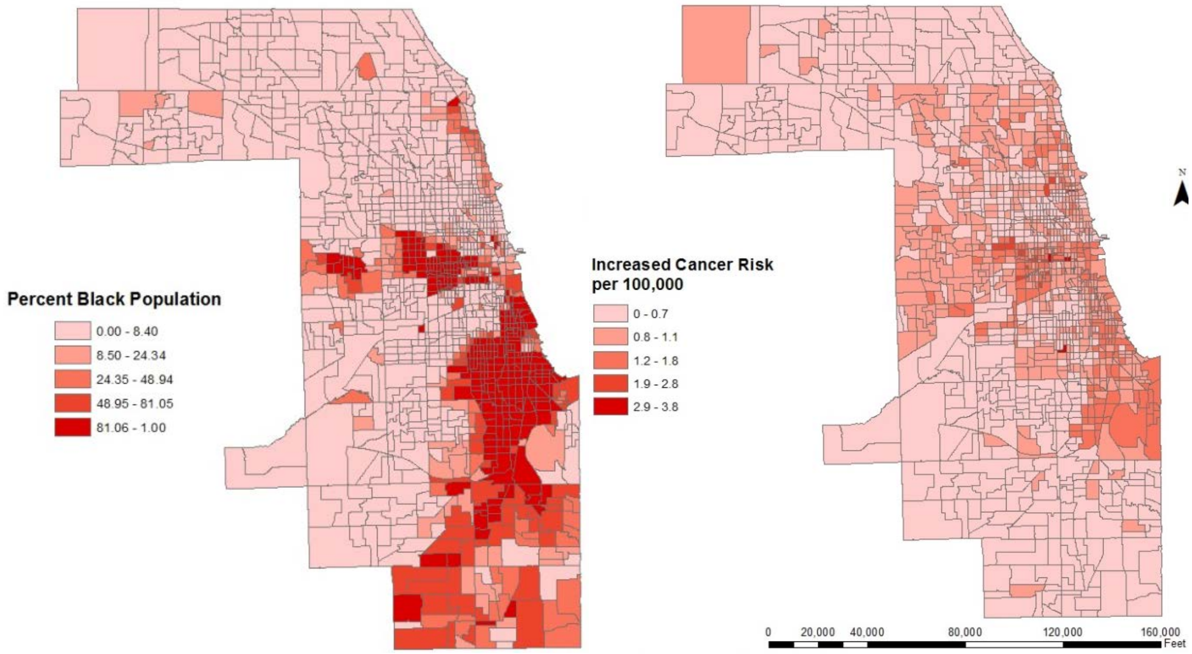

Figure 2. Cook county maps comparing non-point source cancer risk and percent Black population.

In Figure 2, there was a distinct geographic pattern of higher cancer risk from nonpoint sources and higher-percent Black on the West side and South side of Chicago and Cook County. Moreover, the areas with higher cancer risk and higher-percent Blacks appeared to overlap. It contradicts the findings of our traditional statistical results, but it also confirmed a common problem of finding mixed results in health and environmental disparity research. The GIS software applications allowed us to perform analysis on multiple combinations of variables; however, we discovered that this approach was not optimal in spatial analysis. These preliminary findings suggested that the identification of analytical steps in the process was critical in order to overcome data challenges.

7. Discussion

The environmental justice movement has sparked contentious debates among researchers, policy makers, activists, and industry as to whether environmental discrimination actually exists and how broader social and structural factors contribute to such disparities (189). Traditional statistical results may provide little insight into the complex relationships between geography, sociodemographic characteristics, and environmental exposures. Multivariate regression models look at the strength and the significance of relationships between the dependent and independent variables but do not take into consideration locality or the spatial relationship between factors. The non-parametric indicator approaches, such as Poisson, binomial, and generalized linear model, can also introduce biases (190,191).

We did not find significant relationships between demographics, cancer rates, proximity to hazards, and environmental cancer risk when we employed EDA and ESDA separately. However, the results confirmed that geographic pattern existed in chloropleth maps. We concluded that traditional ordinary least squares based statistical analyses were not conducive when exploring non-linear geographical components. We speculate that this may be one of the reasons for conflicting results demonstrated in previous studies. The geographical patterns and associations began to emerge when we combined EDA and ESDA with brushing and linking technics. An integration of both traditional and spatial statistical methodologies was needed to test the health disparities conceptual framework.

8. Methodological Conceptual Model

To integrate traditional statistical methodologies and spatial analysis, we developed a sequential methodological framework (Figure 3). It became apparent that our research was divided into three areas of analysis: EDA, ESDA, and spatial statistics. The process begins with EDA, including descriptive univariate, bivariate, and multivariate analysis. This allows us to investigate the data without a spatial component to determine relationships of significance and narrow our list of variables of interest.

With these variables, we then conducted ESDA to visually examine the data with a geographical component. According to the GeoDa Center, "At the core of ESDA techniques are measures of spatial autocorrelation with other techniques that allow for the detection of outliers, spatial trends, and spatial regimes. ESDA is exploratory in the sense that it cannot explain the patterns it reveals (192)." Lastly, to confirm non-random spatial patterns, spatial statistics, including spatial regression, are necessary to confirm relationships of significance

Initially, we believed this was a chronological three-step model but this was only partially true. Overall, the three papers that were produced from our research were partitioned into these three areas but there was an overlap between EDA and ESDA when it came to mapping data. We found that EDA and ESDA were not mutually exclusive methods but required concurrent analysis while determining possible visual patterns with variable of interest. The EDA is emphasized in the model to show the analysis conducted in this paper.

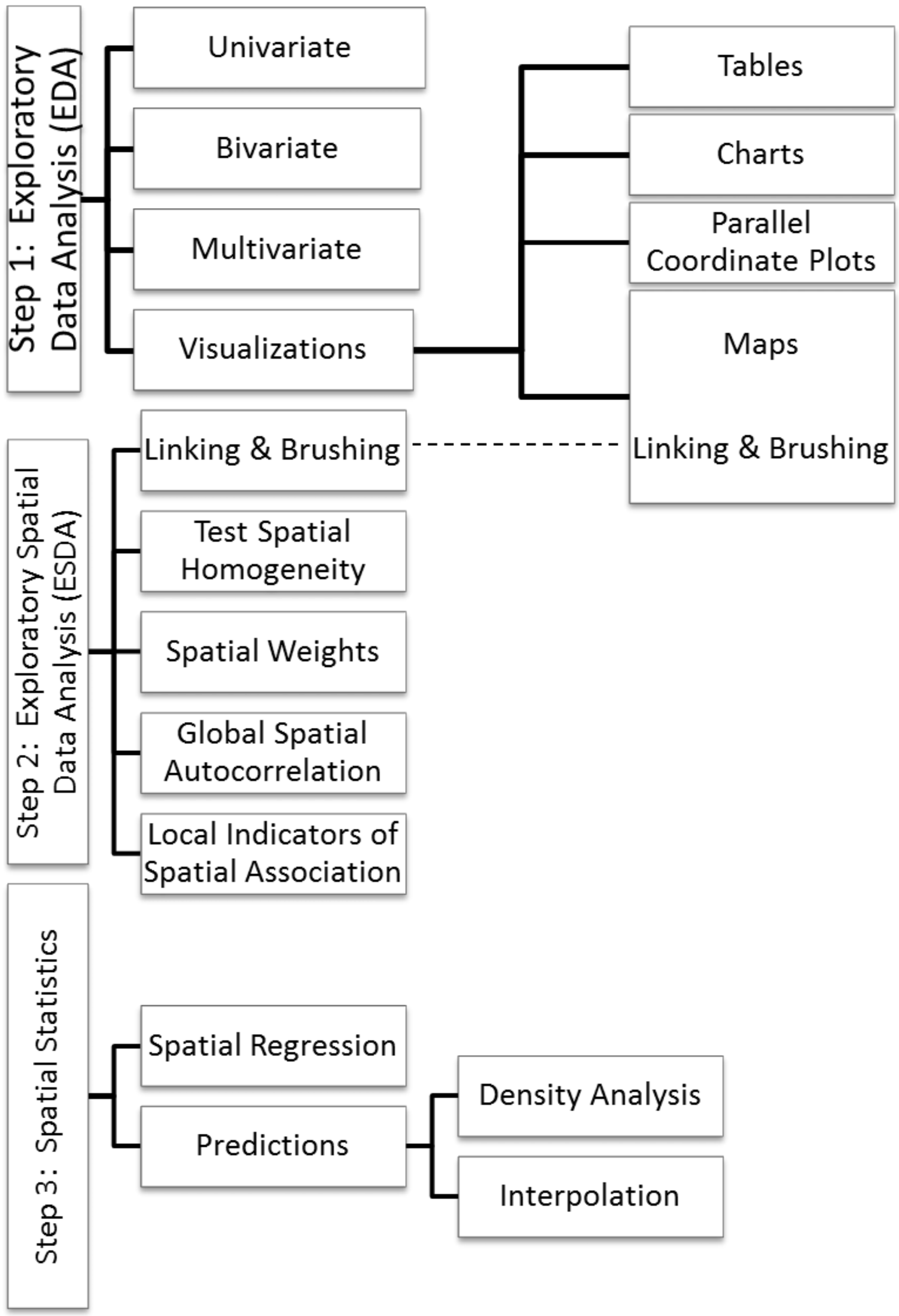

Figure 3. Integrative methodological framework with a modified three-step approach connecting eda and esda with linking and brushing.

Our results in this paper confirmed that mapping was a key component to discovering indicators; however, originally narrowing the list did not produce significant outcomes when using these variables to perform spatial autocorrelation. Our second paper (193), highlighted brushing and linking techniques that were most effective when combined with EDA mapping. Parallel coordinate plots and scatterplots were also helpful in identifying factors of concern to move onto the second step of ESDA and the successful implementation of global spatial autocorrelation. At the end of this stage, LISA are only conducted if global spatial autocorrelation finds clustering, and then this analysis confirms the presence or absence of significant spatial clusters or outliers for each location. Although LISA is an ESDA function, we start our third paper with univariate and multivariate LISA interpreting our findings before we move onto spatial regression (194).

9. Conclusion

Geographic visualization and spatial analysis are a promising approach to identify areas with various characterizations of interest in the field of public health. Moving forward with social research, there are numerous considerations that require an understanding of limitations and assumptions of both the data and GIS software to ensure the correct interpretation of outcomes. Spatial autocorrelation and regression methods can be implemented to substantiate claims of a possible spatial relationship for further spatial analysis. To analyze data with a spatial component, it is important to understand the role of EDA and ESDA in the process to ensure interpretable outcomes. This is especially true when working with multiple data sets across multiple disciplines. Future research needs to further investigate the effectiveness of the integrated analysis methods in explaining varying health outcomes.

B.. <u>Utilizing Exploratory Spatial Data Analysis to Examine Health and Environmental Disparities in Disadvantaged Neighborhoods</u>

Health disparities research has focused primarily on racial and socioeconomic differences in health outcomes. Although neighborhood characteristics and the concept of built environment have been shown to affect individual health, measuring the effects of environmental risks on health has been a less-developed area of disparities research. To examine spatial associations and the distribution of geographic patterns of sociodemographic characteristics, environmental cancer risk, and cancer rates, we utilized existing data from multiple sources. The findings from our initial analysis, which was concerned with proximity to environmental hazards and at-risk communities, were consistent with results of previous studies, which often reported mixed relationships between health disparity indicators and environmental burden. However, further analysis with refined models showed that several key demographic and subdomains of cancer risk measures were shown to have spatial components. With the application of exploratory spatial data analysis, we were able to identify areas with both high rates of poverty and racial minorities to further examine for possible associations to environmental cancer risk. Global spatial autocorrelation found spatial clustering with percent Black, percent poverty, point and non-point cancer risks requiring further spatial analysis to determine relationship of significance based on geography. This methodology was based upon particular assumptions associated with data and applications, which needed to be met. We conclude that careful assessment of the data and applications were required to properly interpret the findings in understanding the relationship between vulnerable populations and environmental burden.

1. Introduction

Health disparity research examines numerous factors such as socioeconomic and demographic variables to understand poor health outcomes including racial minorities living in poverty. Recently, a focus has been placed on the built environment, taking into consideration the surroundings of a neighborhood that may contribute to disparity. Environmental health factors are of great interest because certain cancers had associations to environmental cancer risk. Epidemiological studies confirm a relationship between the location of pollution sources and incidences of multiple cancer types(195-199). The examination of environmental risk levels, human exposure, and proximity to such risks in air, water, and soil, and availability of resources to mitigate the effects of these environmental risks health influence health outcomes (200,201).Moving forward with social research, there are numerous considerations that require an understanding of limitations and assumptions of both the data and GIS software to ensure the correct interpretation of outcomes.

Traditional statistical results often fail to provide insights into complex relationships between geography, socio-demographic characteristics, and environmental exposures. For instance, multivariate regression captures the strength and the significance of statistical relationships between the dependent variable and other explanatory factors, however; important local variations between dependent and exploratory variables may not have been understood or looked at the high dimensionality of the datasets(202). Therefore, there is a need for other methods to describe these interactions of interest (203).

Early environmental justice research explored a number of hazardous sites in disadvantaged communities. The premise was that more hazardous sites translated to higher exposure to environmental risk factors. The U.S. Environmental Protection Agency (EPA) acknowledged that past methods analyzed the proximity to sources of environmental hazards. But, mapping sites to evaluate spatial clusters lacked evidence between increased exposure and risk. In addition, mapping environmental injustice did not measure the correspondence between the location of potential environmental burdens, exposures, and health effects (204). NATA cancer risk data was an opportunity to understand exposure risk in relation to health outcomes despite limitations within this assessment.

Geostatistical modeling involves multivariate data which needs an underlying joint multivariate distribution for valid inference (205). Traditional data analysis methods for multivariate visualization such as tables and scatter plots, commonly used to examine health disparities and environmental health data, have been found to be limited in their ability to represent very large datasets (206). GIS visualization methods have an advantage of assisting in identification and further exploration of certain patterns, thereby generating new analysis that can be created in an easily understood form (207). Various components of an analysis design coupled with the use of domain expertise through interactive exploration can develop into multivariate spatial patterns and the data can be allowed to show the obvious for hypotheses development (208).

Exploratory data analysis (EDA) look at data such as correlations and measures of fit but also needs to be carefully investigated because these results become invalid when there is spatial dependence (209). Exploratory spatial data analysis (ESDA) focuses on spatial aspects of the data to find possible spatial patterns and outliers (210). Spatial methodologies find apparent spatial relationship, however; there are several problems inherent with social data that need to be considered. For example, spatial health research uses data that is collected for a purpose not specific for spatial analysis and is sometimes sampled in a systematic way from a spatially distributed population (211). Because there are limitations regarding inference from analyzing spatial patterns, researchers need to understand spatial systems, the selection of and specification of spatial weights and the subjectivity of the methods themselves (212).

Visualization methods have an advantage of identifying and exploring certain patterns, thereby generating new analysis that could be designed in an easily understood form (213). ESDA tools such as maps, scatterplots, and parallel coordinate plots present information in a seeable manner to discover these patterns. Various components of an analysis design coupled with the use of domain expertise through interactive exploration could develop into multivariate spatial patterns and the data could show the obvious for hypotheses development (214). Spatial analysis software has a statistical pattern recognition approach and is implemented in which a spatial cluster statistic or autocorrelation statistic is used to quantify a relevant aspect of a spatial pattern (215). However, the term “cluster” in health research is often generic that it fails to describe spatial variation without a precise description of the statistical test, heterogeneous population sizes, spatial autocorrelation, and non-uniform risks in social science (216).

In our study, limitations and assumptions associated with the EPA National-Scale Air Toxics Assessment (NATA) environmental health risk assessment data posed challenges in reviewing the relationship between cancer risk and health disparity measures. The EPA provided data at the census tract level, yet, stipulated that NATA assessments were not a definitive means to identify specific risk values within a census tract and that these results were more meaningful at the State or national level (217). Current research suggested that smaller units such as tract or block group measures were: 1) most attuned to capturing economic deprivation, 2) meaningful across regions and over time, and 3) easily understood, and hence based on readily interpretable variables (218).

2. Methods

First, we performed EDA using variables of interest that best captured vulnerable populations and environmental burden within Cook County, IL. This list was extensive utilizing health disparity indicators such as race/ethnicity and socioeconomic status (SES) and NATA environmental cancer risk data divided into six categories. STATA software was used for descriptive statistics to examine the distributions of dependent and independent variables, correlations between continuous demographic variables, the total cancer risk derived from NATA, and cancer incidence rates, and bivariate models to explore relationships between the outcome measures and demographic and environmental risk factors.

We then used ArcGIS and OpenGeoDA for EDA visualization in identifying variables and areas warranting further examination. Based on the results of EDA and parallel coordinate plots, we tailored the list to key variables to model a series of global spatial autocorrelations to determine whether clustering existed and if so, to move forward with future local autocorrelation and spatial regression. Census tracts that showed possible correlation were then brushed and linked to box plot maps for visual comparison. In OpenGeoda, brushing and linking was a technique that allowed us to highlight points of interest on a graph and then links or identifies these points on corresponding maps and charts.

We created chloropleth maps in ArcGIS software based on the variables of interest outlined above to visualize descriptive statistics. ESDA including scatterplots and parallel coordinate plots looked at multivariate and bivariate relationships as a precursor to investigating spatial randomness in OpenGeoDa software. Although these tools provided additional insight into the data, there were no definitive relationships between variables of interest when we looked at the entire geographical area. Therefore, we decided to narrow the scope of our geography by examining census tracts located in the West and South regions of the City of Chicago.

We used spatial weights based on a queen matrix rook to take into consideration contiguity issues. Spatial autocorrelation, Global Moran's I with permutation inference, examined negative and positive spatial correlation with a standardized z-value. The goal of positive and negative spatial autocorrelation was to investigate similar and dissimilar values in relation to location. The spatial autocorrelation statistic captured both attribute and location similarity but it was important that the z- statistic was not interpreted as statistical significance.

This method investigated clustering to decide if a spatial relationship exists and if so, to justify further research with local spatial autocorrelation and spatial regression.

3. Results

We explored the data based upon the variables of interest that were found to be correlated with multiple socio-economic and demographic attributes. This included percent of Hispanic and Black populations, percent of poverty and median household income as well as percent rented housing units and percent of population without a high school diploma. These factors were then examined in relation to the six categories of NATA cancer risk and cancer rates at the census tract level. Initially, it appeared that there was no association between health disparity indicators and environmental burden. This corresponded to past research that had mixed results when looking at proximity to environmental hazards and at-risk communities.

Instead of looking at the data from a broad perspective to narrow down the list, we decided to start with three key variables: percent poverty, percent black residents, and NATA total cancer risk. By linking the upper outlier of percent poverty gave us an opportunity to identify these census tracts which were all located within the City of Chicago.

The chloropleth maps (Figure 4) zoomed into these areas to see if these tracts were concentrated in certain Chicago Community Areas (CCA). The seventy-seven CCAs had unique neighborhood characteristics and have defined census data to correspond with these boundaries.

We identified 18 of 77 CCAs predominantly on the south and west sides that contain at least one census tract within their boundary. There was a visual pattern concentrated on the west and south sides with a few tracts scattered toward the north and southwest.

A parallel coordinate plot was then generated using the seventy census tracts in the highest outlier category with high percent poverty and high percent black population to look at the different types of cancers and cancer risk exposures from point and non-point sources. Figure 5 showed total cancer incidence rate, and the rates of breast and lung cancer, point source cancer risk and non-point source cancer risk. These two cancer risk categories were based on the premise that at-risk neighborhoods have higher environmental burden due to the proximity to hazard. Figure 4 showed that point source cancer risk appeared to be higher in eight of the CCAs, which were all known disadvantaged neighborhoods including North Lawndale. Also, there were census tracts of higher risk adjacent to these CCAs that needed to be further examined to determine if they are also areas with health disparity. The non-point source cancer risk map yielded a more prominent pattern within the majority of the eighteen CCAs and with neighborhoods adjacent to CCAs on the west side of Chicago.

The global spatial autocorrelation looked at a pattern as whole or “clustering” and a general approach to similarity and dissimilarity. In Figure 6, the positive spatial autocorrealtion in the upper right quadrant looked at similarity of neighbors while negative spatial correlation in the lower left corner looked at the dissimilarity of neighbors. This was not a definitive outcome showing a spatial relationship but an indicator that we could move onto the next diagnostic test of local autocorrelation.

We conducted global spatial autocorrelation on percent poverty for census tracts in Cook County, IL to determine if there was clustering and if so, did the upper outliers correspond to the seventy census tracts identified through ESDA. With a Moran's I score of 0.66, we observed clustering and with the linking technique we saw that the census tracts from the west and south side were in the upper right quadrant of the scatterplot (Figure 6). This indicated a positive and significant clustering of like values. We then examined global spatial autocorrelation with non-point source cancer risk for census tracts in Cook County, and found a weaker possibility of clustering with a Moran's I score of 0.49. There was no clustering of the lung cancer incidence rates and the linking of the high-high quadrant showed no patterns.

4. <u>Discussion</u>

The use of EDA/ESDA with health disparities and the built environment may provide additional insights into identifying at-risk neighborhoods with vulnerable populations and increased environmental burden. Our findings showed that some of demographic and subdomains of cancer risk measures had possible spatial components. With the application of ESDA, we were able to identify seventy census tracts with both the high rates of poverty and racial minorities. The Global Moran's I results further showed clustering for percent poverty and non-point cancer risk. These areas were predominantly poor neighborhoods in Chicago.

Additional investigation will be required as to the reason for the high rate and risk in theses census tracts when compared to other census tracts with lower rates and risks.

GIS methodologies have become popular in health disparities research because of the ability to conduct spatial analysis, however, a challenge with geographical aspects of data is the determination of appropriate boundaries of study. For the use of GIS to generate a reliable data for testing hypotheses in population heath, however, there needs to be an understanding of the GIS methods used in and providing justification for the geographic level of study chosen (219). Choosing levels and geospatial units to analyze depends on several factors, including the research objective, the causal model selected, the exposures and health outcomes of interest, and the extent to which data are available (220). The availability of data is often the determining factor in decisions about geospatial issues especially with a driving force being socio-demographic variables from the census. In deciding the scale of analysis or the level of aggregation in a study, there tends to be a trade-off of the specificity of the study and the precision of the study (221). The trend seems to be that the larger an area is, the less the specificity and the relevance of the findings of the study to the local populations but the higher is the precision and the reduction of bias (222).

The reconciliation of various geographical boundaries presented numerous concerns. In the Chicago land area, data were collected by various geographical boundaries including: county, census tract, zip code, and CCAs. There were techniques within GIS programs to reconcile these boundaries such as clipping, intersection and dissolving, however;

there was an assumption of homogeneity when manipulating these boundaries. We utilized crosswalk files from the U.S. Department of Housing and Urban Development which provided ratios to distribute various land categories. This proportionately estimated the type of land used within the geographic area but does not discern by locality.

Spatial patterns on GIS maps helped formulate hypotheses for explaining geographic patterns, but such approaches are hardly sufficient to explain complex interactions among spatial data (223). We identified several issues in mapping environmental health disparities including the lack of comprehensive hazard databases; inadequacy of exposure indices, risk assessment methodologies, and insufficient health effects data (224). When a disproportionate environmental burden based on race and/or income was found, it was critical to demonstrate the disproportionate effects of pollution rather than just the disproportionate distribution of pollution sources (225).

GIS was based on spatial models that apply to static spatial systems (226), which made it difficult to represent human mobility and temporal change in cancer, environmental and socioeconomic data (227). Environmental health factors added another layer of complexity in determining exposure and risk especially when arbitrary boundaries could not take into effect the migration of contaminants. GIS had great promise in health disparities research as we investigated the physical environment to determine if hazardous exposures impact health in at-risk populations, however; exercising caution was needed especially with defining assumptions and limitations in interpreting outcomes.

5. Conclusion

ESDA identified seventy census tracts in the upper outlier for both percent poverty and percent blacks in eighteen CCAs in the City of Chicago. The Global Moran's I results further showed clustering for percent poverty and non-point cancer risk. These areas were predominantly poor neighborhoods in the west and south side of Chicago with non-point cancer risk located on the west, north and south sides of Chicago. The North Lawndale neighborhood and the adjacent area had the most census tracts with high non -point cancer risk and cancer incidence. The next step was to continue with ESDA to look at local spatial autocorrelation to confirm our potential variables for spatial regression.

As we move forward with spatial analysis in health disparities research, it is important to understand the applicability of GIS software with social science data. In addition, researchers need to discern parametric approaches in traditional statistical methods and nonparametric approaches in spatial analysis; and, the limitations and assumptions associated with both. Methodologies to address these issues will ensure appropriate interpretations of outcomes.

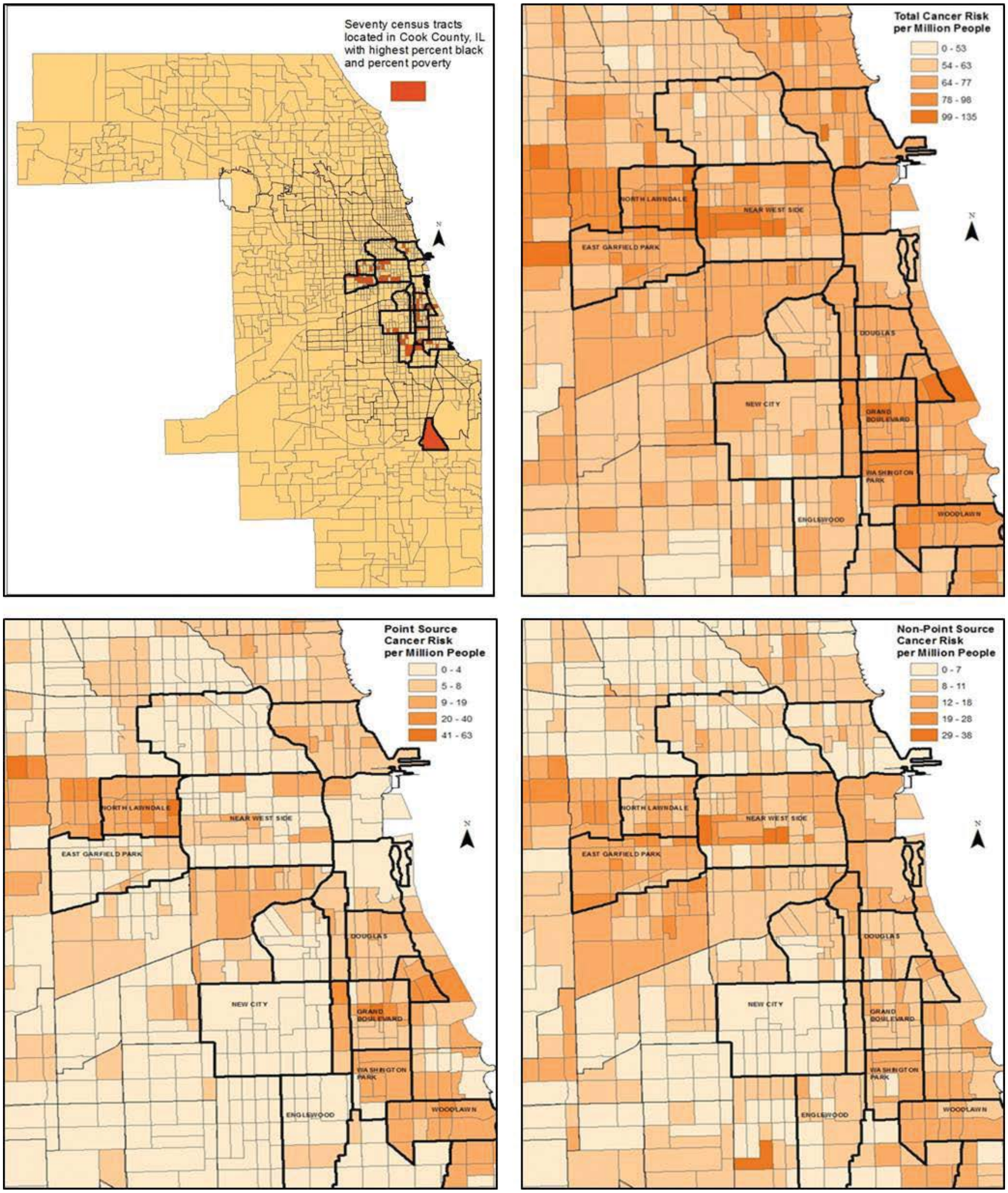

Figure 4. Chloropleth maps for visualization of patterns.

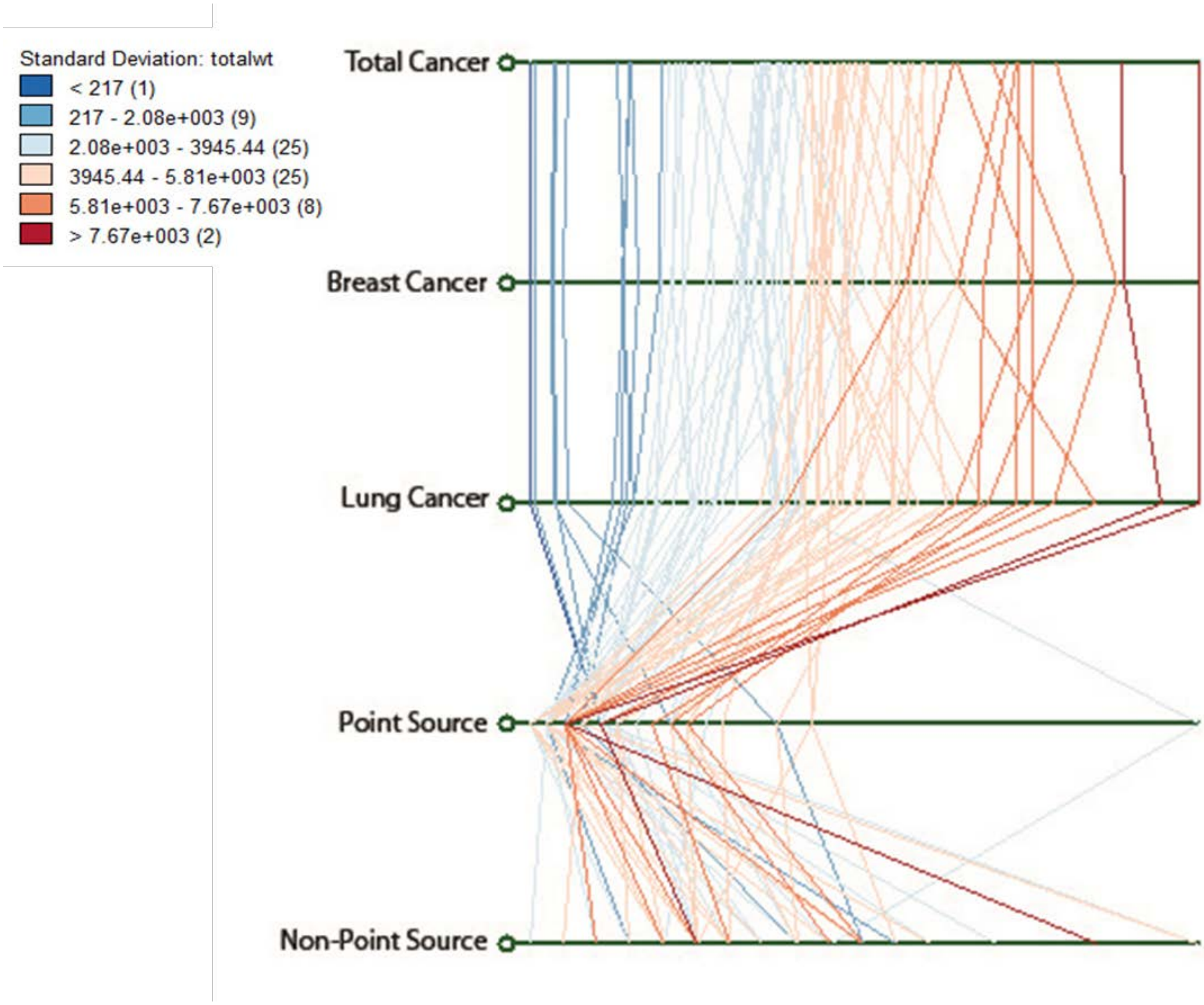

Figure 5. Parallel coordinate plot for the seventy census tracts with highest percent poverty and highest percent black population in relation to total cancer incidence rate, breast cancer incidence rate, lung cancer incidence rate, point source cancer risk and non-point source cancer risk.

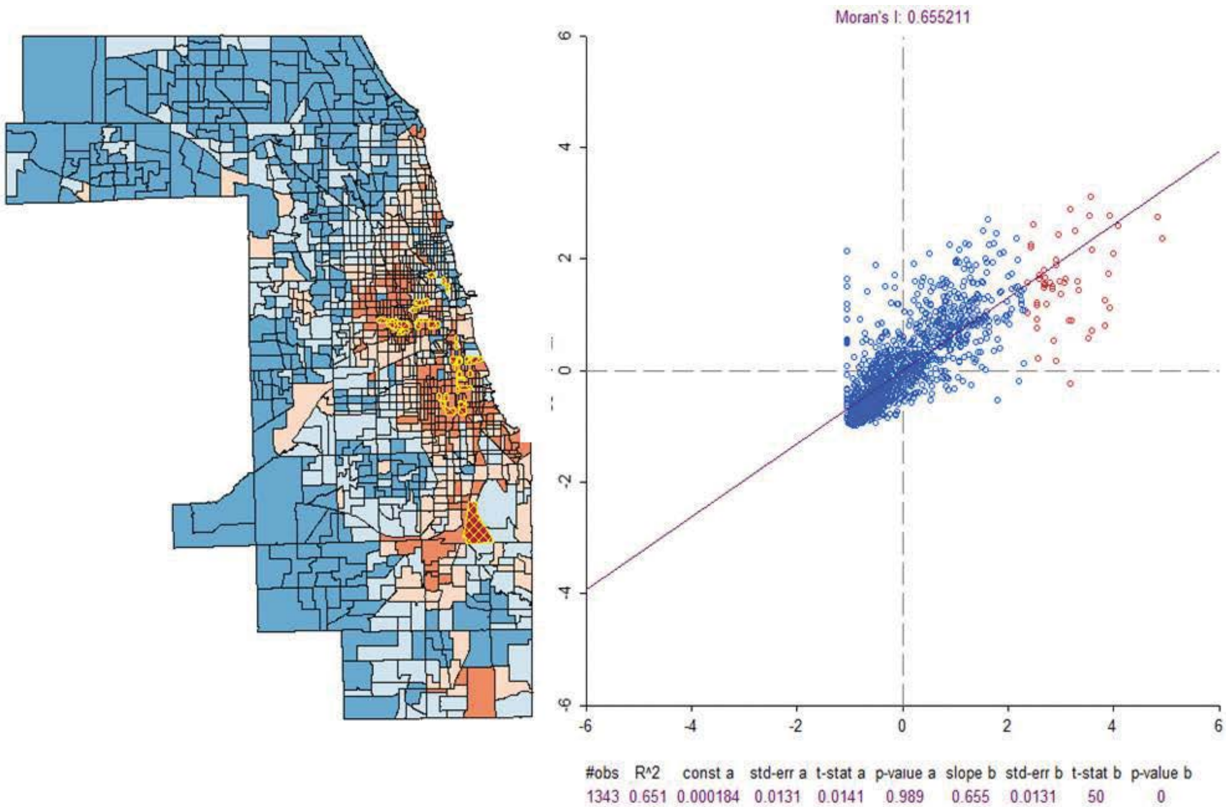

Figure 6. Box plot (hinge 1.5) percent poverty in Global Moran's I with upper outliers in the high-high quandrant highlighted with the corresponding census tract location on the Cook County map.

.

C. **Local Indicators of Spatial Association and Identifying Clusters in Health and Environmental Disparity Research**

Studies reported that racial/ethnic minorities living in disadvantaged neighborhoods experienced a greater rate of exposure to environmental hazards. Knowledge of environmental exposure risks, distributional patterns and their effects on population health require a geographic perspective while investigating social injustices to better understand the causes of health disparities among different populations. However, previous studies often fail to recognize processes and assumptions of spatial analyses. In this paper, we demonstrated the importance of such processes. We used exploratory spatial data analysis methods to examine potential spatial patterns of demographic and cancer risk distributions in Chicago. First, we examined the presence of overall spatial clustering using Moran's I statistic. Our Global Moran's I statistic showed clustering for percent poverty, percent black and non-point cancer risk in predominantly poor neighborhoods in Chicago. Local autocorrelation was conducted to identify spatial clusters and spatial outliers. Local indicators of spatial association provided univariate significant maps, cluster maps and scatterplots which identified spatial clusters for percent poverty, percent black and non-point cancer risk in Chicago. We then conducted bivariate analysis which showed that standardized high percent poverty was significantly correlated with a standardized high neighboring non-point source cancer risk. These findings were conclusive evidence that indicated the presence of spatial clusters, while the strengths of the associations cannot be determined. The findings warrant further analysis with spatial regression methods.

1. Introduction

Geographic spatial patterns of health disparities can assist in recognizing two primary requirements in disease prevention and intervention efforts: 1) linking a wide range of factors at varying geographic units; and 2) monitoring changes overtime (2). The built environment, such as lack of adequate housing, parks, sidewalks, and access to health care facilities have been investigated but it is also important to incorporate environmental risks, such as exposure to air pollution (228-231). Neighborhood social and physical environmental contexts often determine one's exposure to health risks and health outcomes (232-238). Due to the fact that most of social factors are not randomly distributed, not all neighborhoods are equally affected by multiple socioeconomic, demographic and environmental burden (239-241)).

Studies report that racial/ethnic minorities living in disadvantaged neighborhoods experience a greater rate of exposure to environmental hazards (242-250). Knowledge of environmental exposure risks, distributional patterns and their effects on population health requires a geographic perspective while investigating social injustices to better understand the causes of disparities among different populations (251). Exploratory data analysis (EDA) and exploratory spatial data analysis (ESDA) provide informative visualizations for interpretable outcomes especially when working with multidisciplinary data with a geographic component (252). The Integrative Methodological Framework (Figure 7) provides a three step process for approaching spatial analysis. Our first manuscript explains the concept of the model and investigates step 1 and EDA . In this paper, we focus on Step II and ESDA examining the role of global spatial autocorrelation. If clustering is identified with global spatial autocorrelation, we can now proceed to the last diagnostic tests of ESDA which is local autocorrelation.

Local autocorrelation is based on the Local Moran's I statistic (253), and can be visualized with significance and cluster maps with a corresponding scatterplot. A sensitivity analysis, which includes running permutations (to as many as 9,999 iterations) and changing the significance level, addresses problems associated with stability due to multiple comparisons with lower significance of the indicated clusters (254).

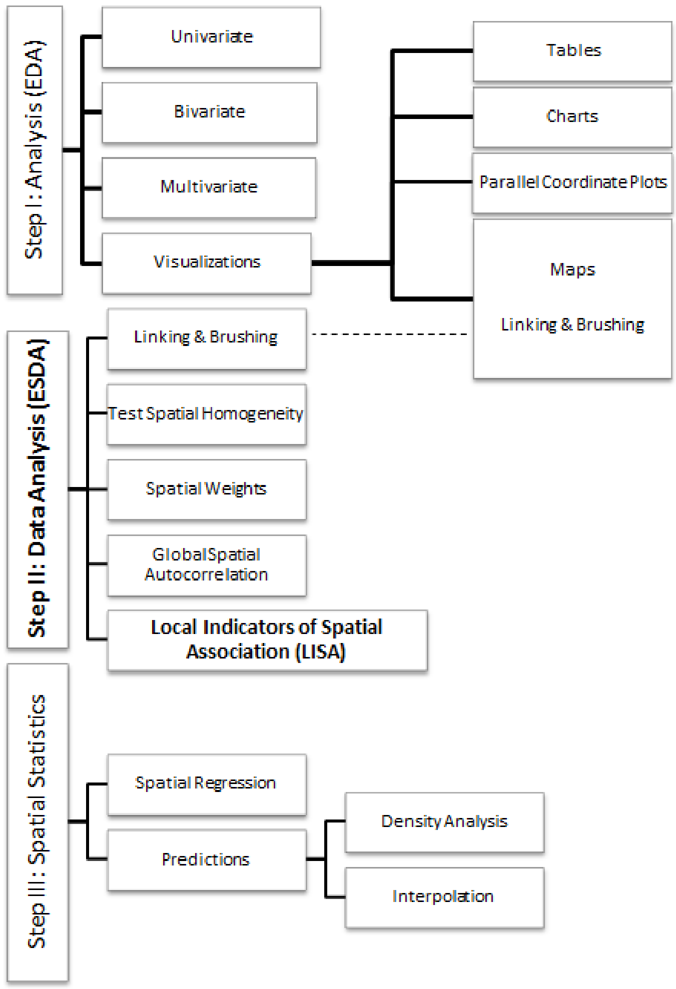

Figure 7. Integrative methodological framework looking at step 2 exploratory spatial data analysis and local indicators of spatial association.

Our Global Moran's I statistic shows clustering for percent poverty, percent black and non-point cancer risk in predominantly poor neighborhoods in the west and south side of Chicago with non-point cancer risk located on the west, north and south sides of the city (255). The next step with ESDA is to look at local indicators of spatial association (LISA). The objective of local autocorrelation is to identify spatial clusters, also called hot spots and cold spots, and spatial outliers. There are two common diagnostic tests which can be performed in both ArcGIS and GeoDa software: Anselin Local Moran's I statistic and Getis-Ord Gi* statistic. The difference between these two tests is that the Getis-Ord Gi* statistic does not detect spatial outliers.

For the purpose of our study, we utilize the Anselin statistic. The theory behind Anselin's (256) LISA is that a statistic is computed for each spatial unit. In this equation z_i is the standardized rate at location *i* that is under investigation and z_j represents the observed rates for the neighboring locations *j* that share a common border with *i* (257). The model estimates the local auto correction for a local in relation to neighboring locations as:

$$L_i = z_i \sum_{j=1}^{n} w_{ij}\, z_j$$

The values are standardized with a mean of zero and a unit standard deviation creating a z-score (258). This is computationally based on conditional permutations of an observed value at *i* which is repeated to obtain a reference distribution (259). The sum of LISA is proportional to the corresponding global statistic and it assesses significance of the local statistic at each location (32). In the presence of global autocorrelation, LISA will represent the cases that have more than the average amount of spatial autocorrelation. High-high (hot spot) represents locations with a high score on a measure that are surrounded by neighboring locations that also score high on the particular measure. Low-low (cold spot) represents locations with a low score on a measure with low scoring neighboring locations. High-high and low-low are considered to be clusters that are referred to as spatial clusters (260). On the other hand, high-low and low-high areas can be determined as spatial outliers (261). A group of areal units is classified as a cluster when the value at a location is more similar to its neighbors than would be under spatial randomness. Global spatial autocorrelation discovers clustering and now LISA identifies these clusters and outliers, however, it does not indicate conclusive relationships or reasons for such relationships.

In both univariate and bivariate LISA cases, you are testing whether local correlations between values at location i and those of its neighbors are significantly different from what you would observe under conditions of spatial randomness (262). The bivariate LISA involves the cross product of the standardized values of one variable at location i (poverty) with those of the average neighboring values of another variable (non-point cancer risk).

These results are sensitive to spatial weights, the number of permutations and level of significance. During sensitivity analysis, it is important to understand these parameters to adjust them accordingly to ensure robustness and stability. In our study, we use OpenGeoDa software and the following discussion is based upon the definitions and assumptions within this program. This is an exploratory exercise that examines both spatial weights and permutations providing a local spatial statistic for each location.

Spatial lag in GeoDa is defined as, “a variable that averages neighboring values of location in which the value of each neighboring location is multiplied by the spatial weight and then the products are summed (263).” There are two types of spatial weights: distance weights and contiguity weight. Distance weights include: distance bands which draw a radius around each point and counts every point within the radius as a neighbor, and k-nearest neighbors (KNN) which measure the distance between the central point of a polygon and the number (k) of nearest neighbor points (264). Contiguity weights include: the rook weight matrix with neighbors that share borders (North-South and East-West) and the queen weight matrix with neighbors that share borders and vertices (North-South, East-West, Northeast, Northwest, Southeast & Southwest) (265).

KNN is a quick way to find patterns and can be used in the social sciences to look at intervening factors in relation to distance (266). It is also useful when looking at census tracts especially when you have a larger census tract area with a small population in relation to a smaller census tract area with a large population. Unfortunately, this approach assumes geographical symmetry, which cannot be met because geographic units and boundaries are not standardized (267). Therefore, KNN cannot determine potential variables for spatial analysis because you cannot estimate spatial lag or error models (268). Social research commonly uses census tracts, therefore; the queen or rook weight matrix is the only option if spatial regression will be conducted.

Permutations are used to determine how likely it would be to observe the Moran's I value of an actual distribution under conditions of spatial randomness. A complicating factor in the assessment of significance of LISA is that the statistics for individual locations tend to be correlated, therefore, the usual interpretation of significance is flawed (269). It is important to note that these are numeric results based on a pseudo significant level (known as pseudo p-value) which is different than the traditional statistic definition of a normal data distribution p-value. According to Anselin, “for instance, if an observed Moran's I value is higher than any of the randomly generated Moran's I values, the pseudo p-value would be 1/100=0.01 for 99 permutations or 1/1,000=0.001 for 999 permutations (270).” Initial LISA maps are based on 99 permutations and a pseudo-significance level of p=0.05.

2. **Methods**

We completed a sensitivity analysis to ensure robustness within these three parameters. We then created a weight neighbor histogram for a queen weight matrix and found that we had no islands (isolated locations with no neighboring areas), binomial distribution or a high number of census tracts with very large number of neighbors. Figure 8 shows the frequency of neighbors by census tract: 371 census tracts had seven bordering areas and one census tract had 14 bordering areas.

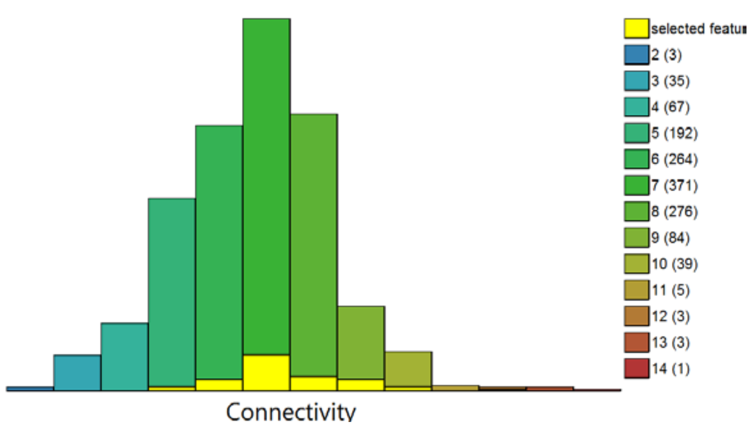

Figure 8. Queen weight matrix histogram.

Our maps were based on 999 permutations to avoid too much sensitivity on the particular randomization which were run multiple times until our results stabilized.

The other diagnostic test we performed was permutation reference histogram to ensure whether our data are non-random. Figure 9 indicates expected (red) and observed (yellow) spatial distributions. As shown, our observed distribution was significantly different from the expected distribution, thus we concluded that our data are not random.

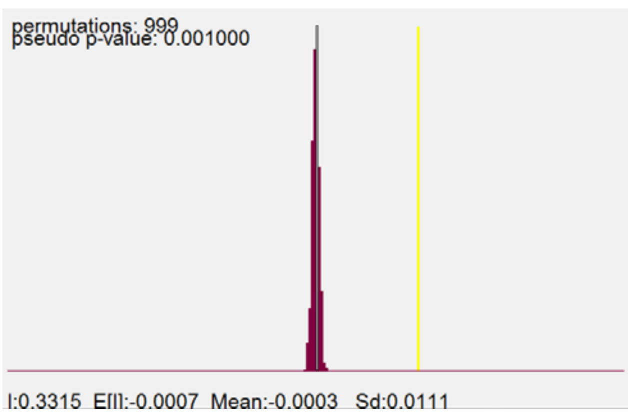

Figure 9. Permutation reference histogram – the lack of overlap between the red and yellow lines indicate nonrandomization..

We conducted univariate and bivariate LISA on 1,343 census tracts located within Cook County, IL comprising of the City of Chicago. With the queen weight matrix, we then randomized the data with 999 permutations. Initial sensitivity maps contained pseudo p-values ranging from 0.05 to 0.0001, however, results for p-value=0.05 could be unreliable. Our sensitivity maps were then adjusted to only include p-values 0.01 to 0.0001.

With the bivariate LISA scatterplot, we calculated the descriptive statistics, and ran the Chow test which looks at the slope and the intercept of one group to see if they are different from those of another group. This test utilizes standard errors from each line and degrees of freedom to calculate a partial f-test.

3. Results

In Figure 10, the LISA significance map showed the locations with a significant local Moran statistic in different shades of green with pseudo p-values greater than 0.01. We observed that there were clusters of significance with pseudo p-values of 0.01 and 0.001 located on the west and south sides of Chicago and in northern and southwestern areas of Cook County, IL. In Figure 11, the LISA cluster map provided the same information as the significance map; however, the significant locations were color coded by type of spatial autocorrelation. This map showed 157 census tracts in red that are high-high (hot spots) which confirmed that there were clusters of high percent poverty on the west and south sides of Chicago. The 272 census tracts on the north and southwest sides of Cook County, IL were low-low (cold spots) which means these were clusters of low poverty. There were only 14 census tracts that were spatial outliers with most being a census tract of low poverty neighboring clusters of high poverty.

We use the brushing and linking technique to select the 157 high-high census tracts in Figure 11. These census tracts are highlighted in red in the upper right (high-high) quadrant of the Moran's I scatter plot (Figure 12). The Moran's I statistic of 0.655211 determines a linear relationship between poverty at a given location on the x-axis with values of poverty in neighboring locations on the y-axis. The red circles represent the red high-high census tracts in the Figure 11 cluster map and the blue circles represent all other census tracts. This highlights how the brushing and linking technique allows us to understand high-high locations on a cluster map and their respective location in a quadrant in the scatterplot.

The significance map (Figure 13) and the cluster map (Figure 14) for percent black produced the same type of outcomes as stated above. In Figure 14, there were 287 census tracts that were clustered with high percent black on the west and south side of Chicago. We also observed this extend into the southern suburbs of Cook County, IL.

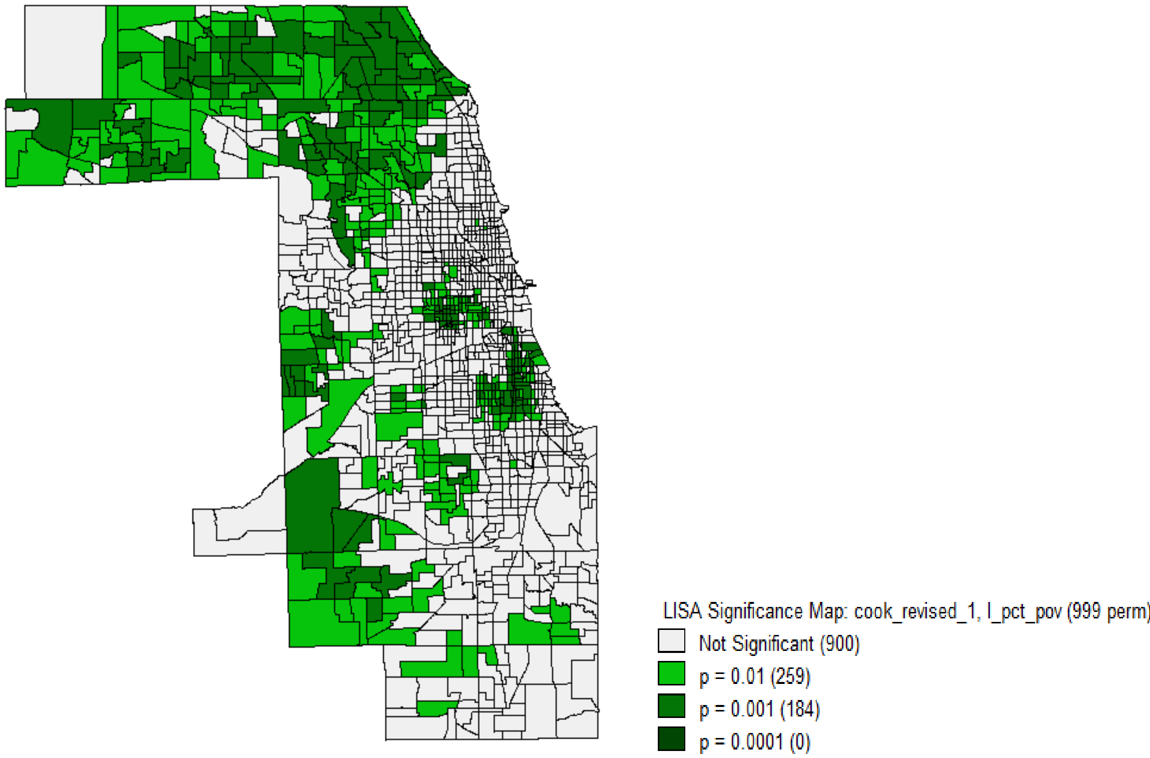

Figure 10. Univariate lisa significance map for percent poverty, cook county.

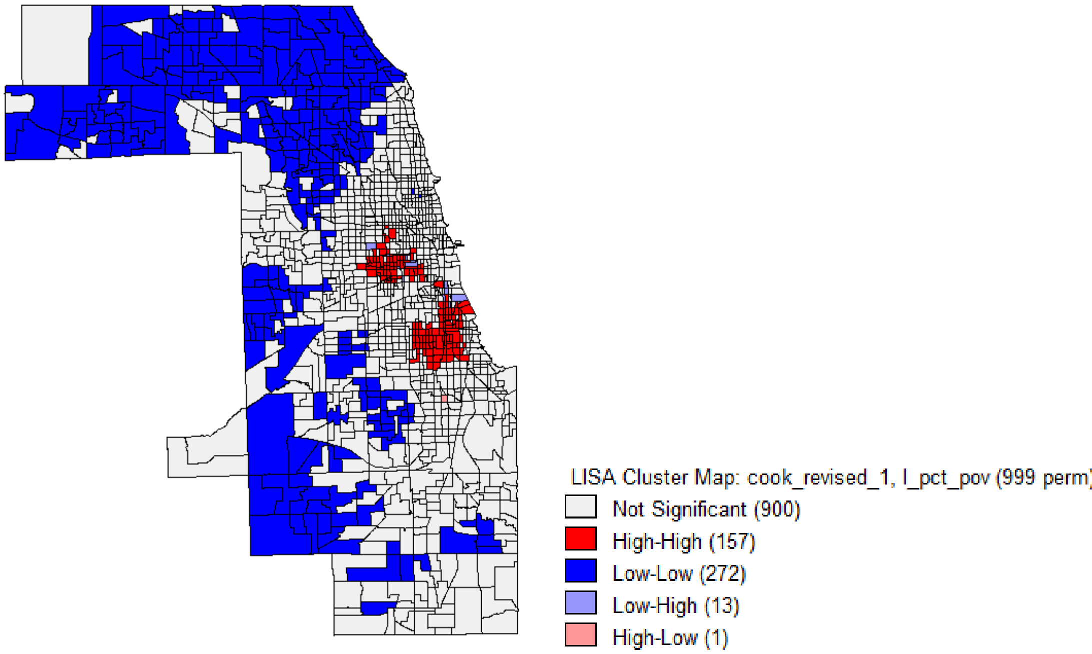

Figure 11. Univariate lisa cluster map for percent poverty, cook county.

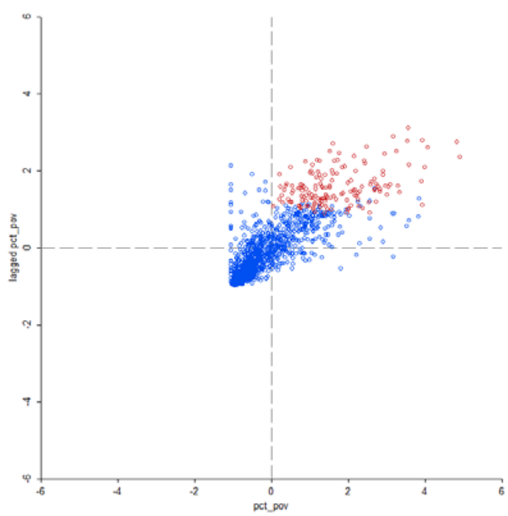

Figure 12. Moran's I scatterplot with a moran's I statistic of 0.655211 for percent poverty.

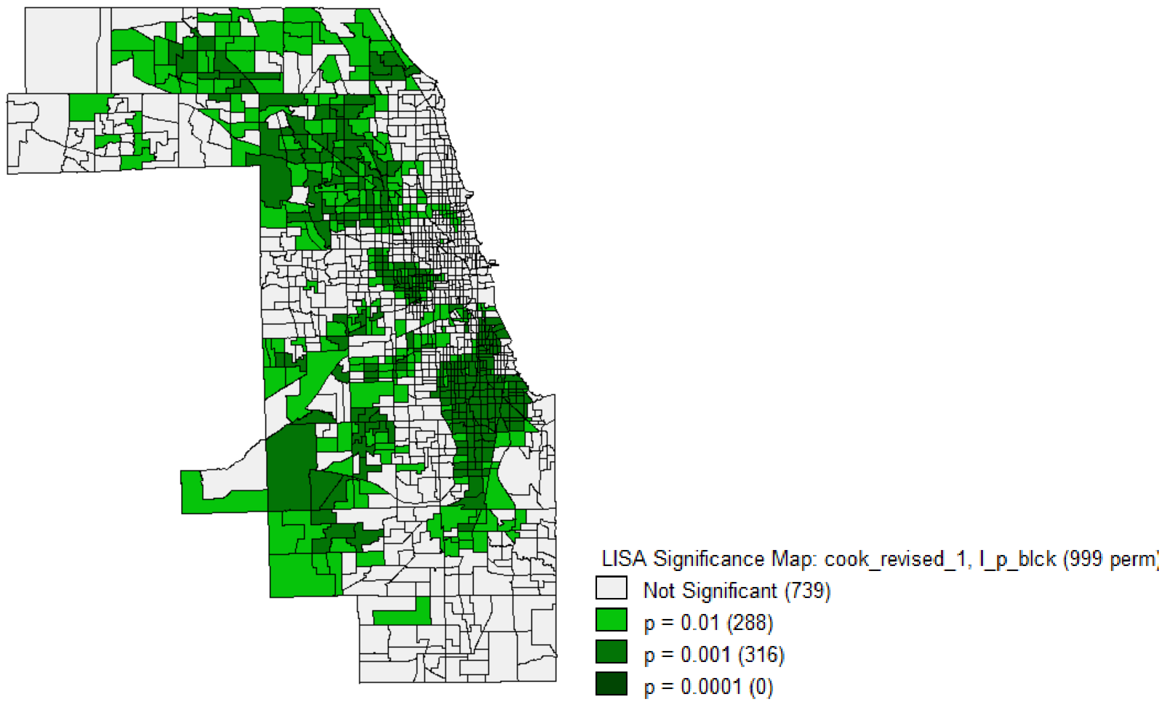

Figure 13. Univariate lisa significance map for percent black, cook county.

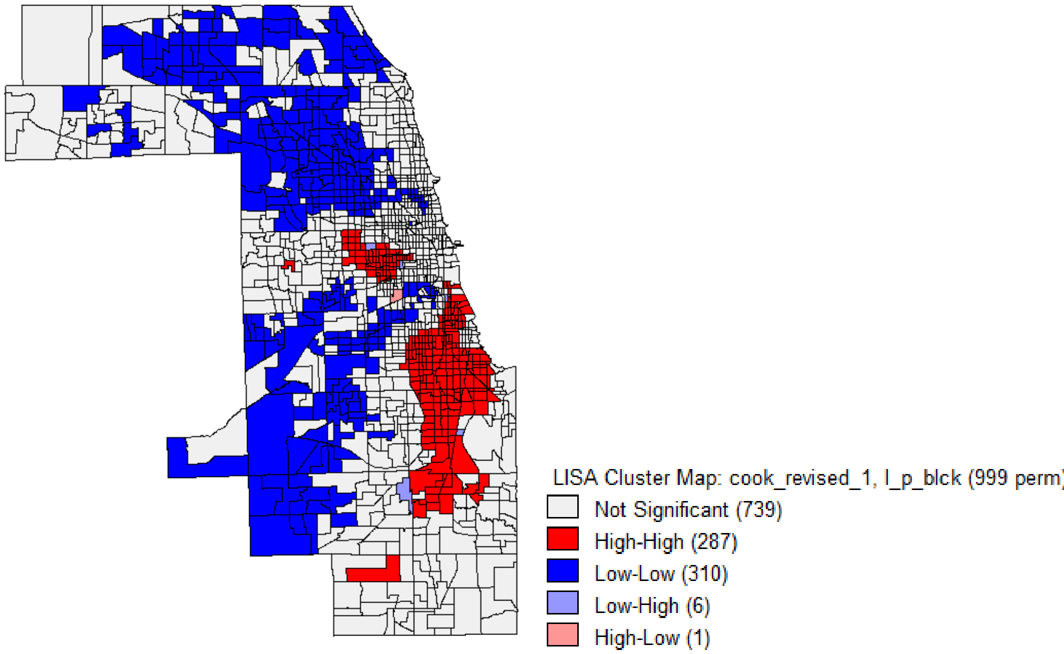

Figure 14. Univariate lisa cluster map for percent black, cook county.

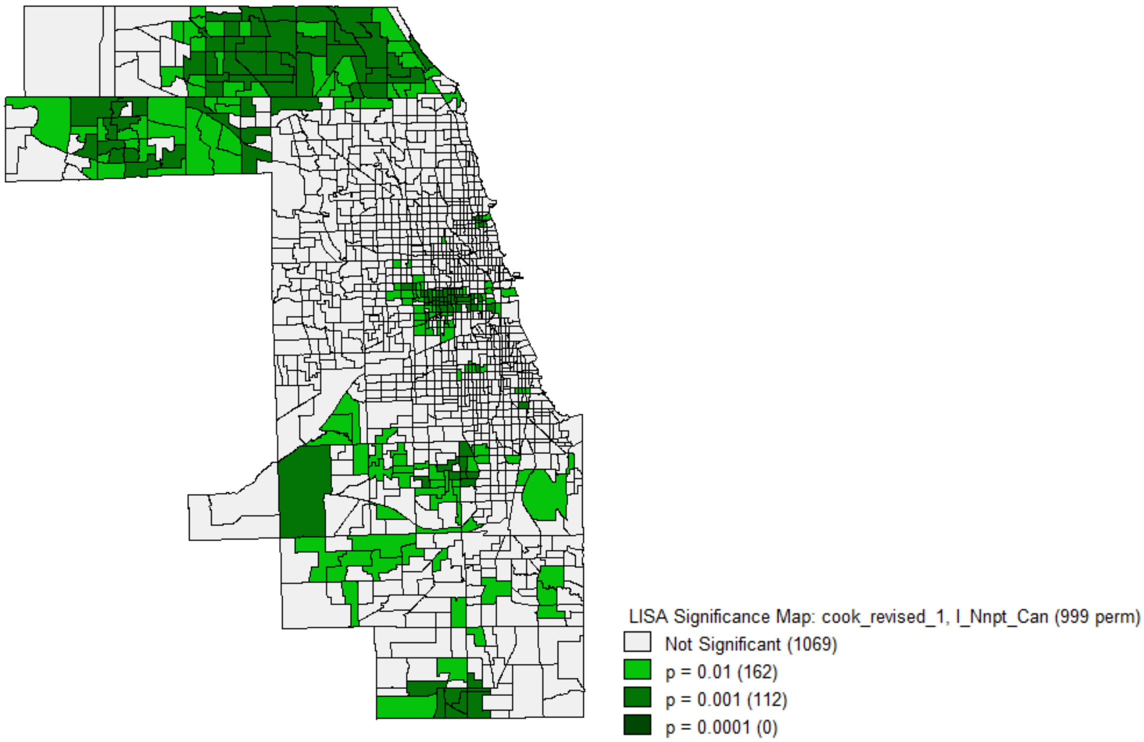

Figure 15. Univariate lisa significance map for non-point cancer risk, cook county.

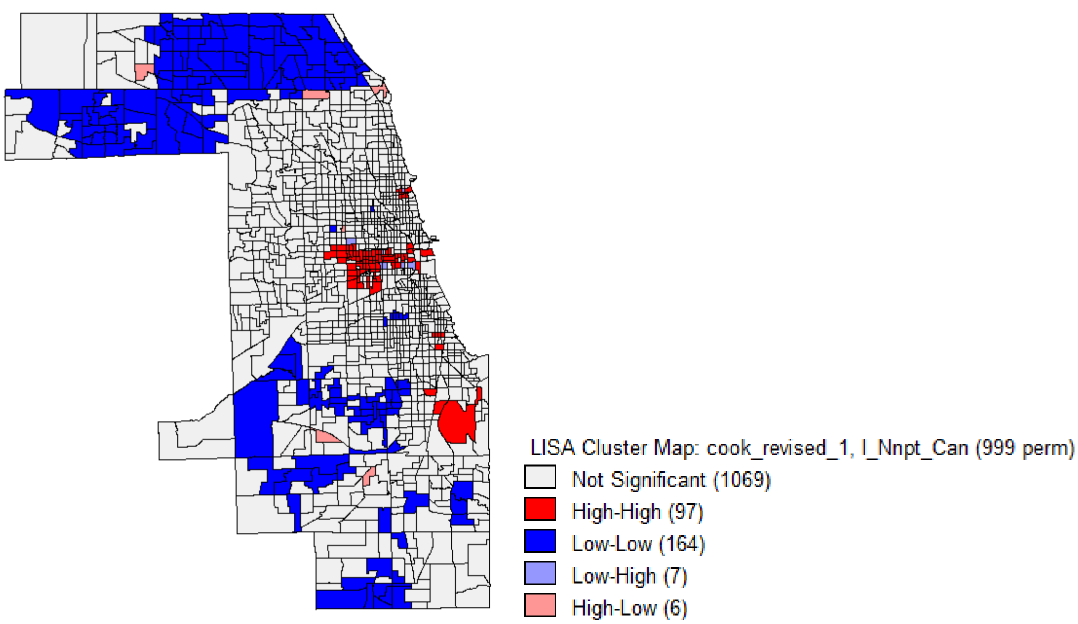

Figure 16. Univariate lisa significance map for non-point cancer risk, cook county.

In Figure 15, there were 274 census tracts of significance for non-point cancer risk on the west side of Chicago and the northern and southwest sections of Cook County, IL. The cluster map in Figure 16 showed clusters for high-high (hot spots) non-point cancer risk on the west side of Chicago with low-low (cold spots) non-point cancer risk mostly on the northern portion of Cook County, IL.

With these results, now that we know there are clusters of non-point source cancer risk, bivariate LISA analysis may help us determine if there are hot spots with non-point source cancer risk in relation to percent poverty and percent black. The bivariate LISA map is the correlation between the spatially lagged Y variable distribution and the non-spatially lagged X variable distribution (271). In this instance, our spatially lagged Y variable is non-point cancer risk with non-spatially lagged X demographic variables (percent black and percent poverty). The maps refer to the local patterns of spatial correlation at a location between percent poverty or percent black and the average nonpoint cancer risk for its neighbors. Again, we look at the parameters with the bivariate LISA maps and we use 999 permutations, a queen weight matrix, with p-values of 0.01 – 0.0001.

The first variable, X, is measured at a specific location and the second variable, Y, is an average of its neighbors' values at that location. The cluster is classified as such when the value at a location (either high or low) is more similar to its neighbors (as summarized by the weighted average of the neighboring values which is the spatial lag) than would be the case under spatial randomness (272). Any location for which this is the case is labeled on the cluster map. However, the cluster itself likely extends to the neighbors of this location as well.

A significant high-high spatial cluster means that high percent poverty is significantly correlated with a high neighboring non-point source cancer risk. A significant high-low spatial outlier means that low percent poverty is significantly correlated with high neighboring non-point cancer risk. There are 80 high-high census tracts with a cluster located on the west side of Chicago. The bivariate Moran's I scatterplot (Figure 18) shows percent poverty on the x-axis and non-point source cancer risk of neighboring areas on the y-axis. Each quadrant is labeled to interpret high-high, low-low, high-low and low-high in relation between these two variables.

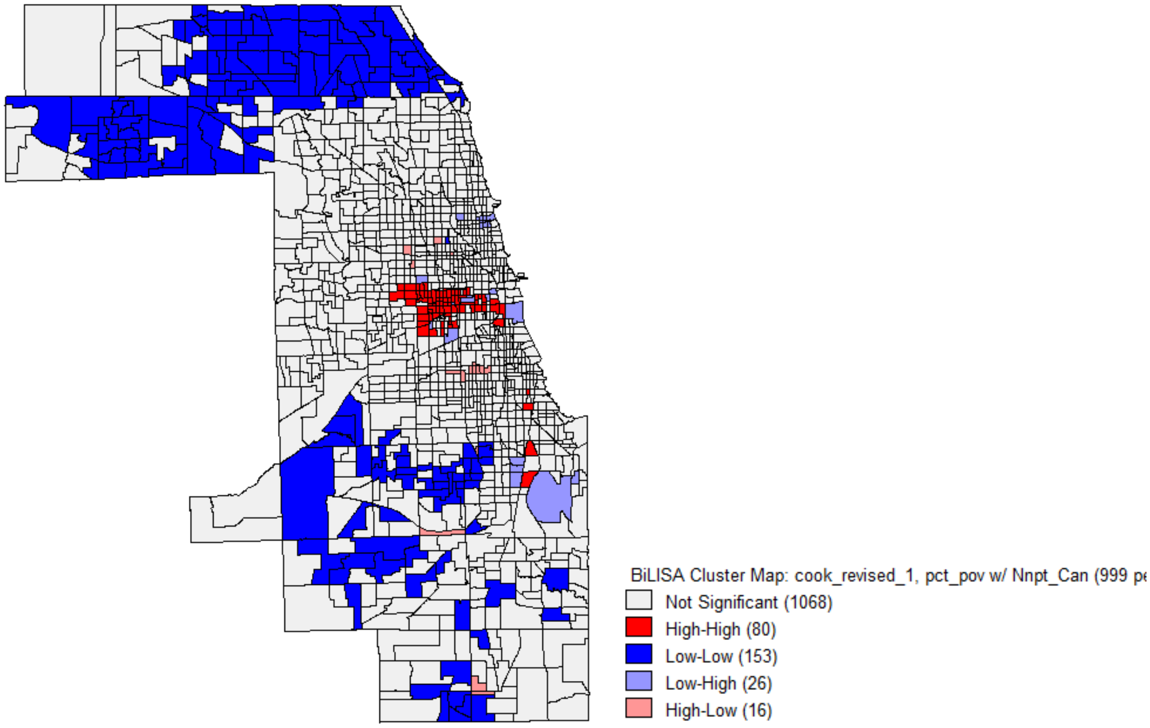

Figure 17. Bivariate lisa cluster map percent poverty lagged with non-point cancer, cook county,

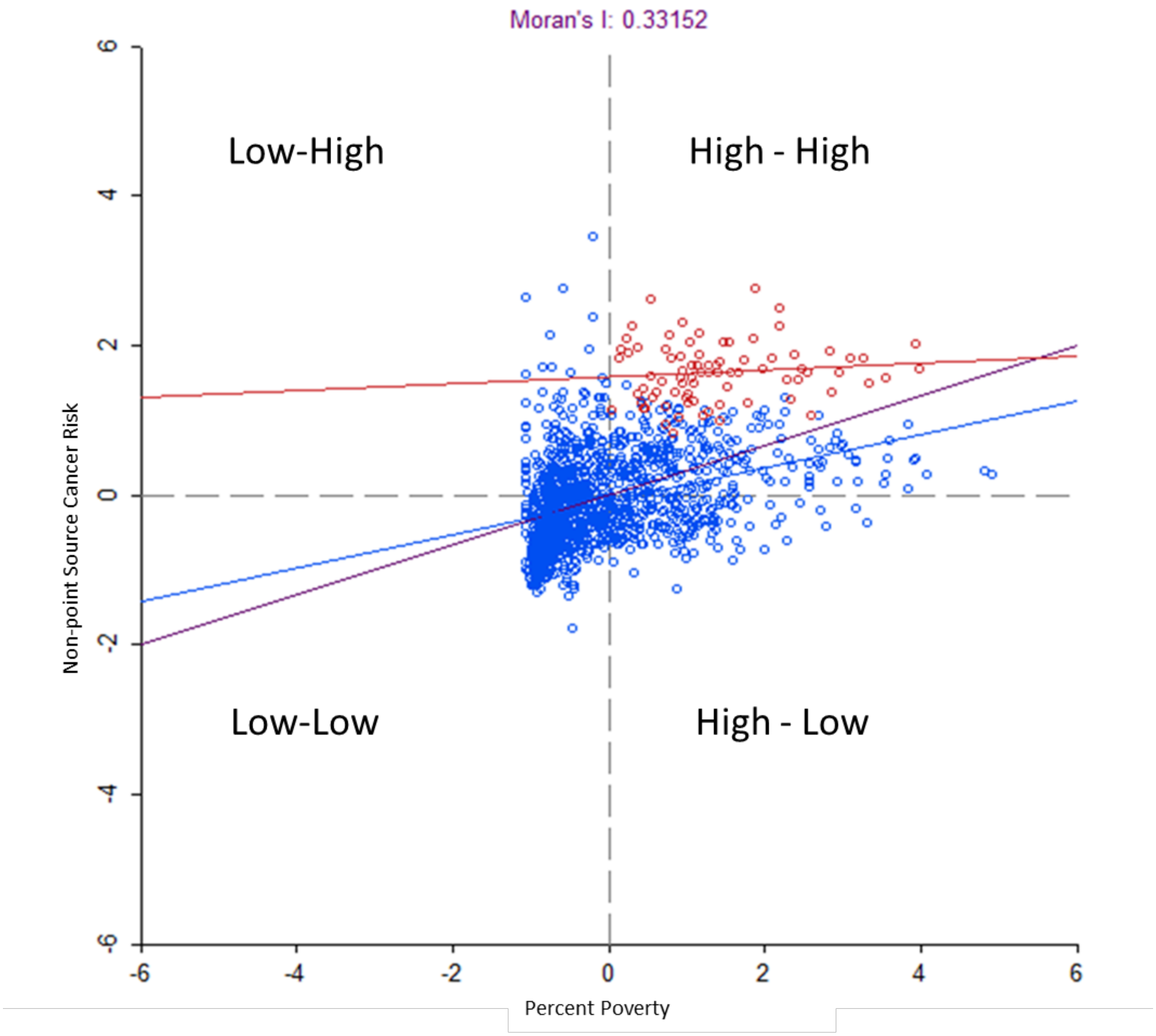

Figure 18. Bivariate lisa scatterplot percent poverty lagged with non-point cancer, cook county.

TABLE III.

DESCRIPTIVE STATISTICS FOR BIVARAITE LISA FOR PERCENT POVERTY AND NON-POINT CANCER RISK

		INTERCEPT				SLOPE			
# of Observations	R^2	Constant	Std Error	t-statistic	p-value	Slope	Std Error	t-statistic	p-value
1343	0.209	0.00442	0.0176	0.251	0.802	0.332	0.0176	18.8	0
80	0.1116	1.58	0.0797	19.8	0	0.045	0.0475	0.957	0.342
1263	0.118	-0.0794	0.0161	-4.92	0	0.223	0.0172	13	0

Chow test for selected/unselected regression subsets distribution F(2,1339) ratio=214.6 p-value=0

In Figure 18, the Bivariate LISA scatterplot with descriptive statistics (Table III) also provides another diagnostic tool to better understand the relationship of high percent poverty and high neighboring non-point cancer risk. We can investigate whether one regression model applies to the entire dataset or if it is more appropriate to look at subsets of the data. For example, high-high is selected on the cluster map in Figure 17 and with the linking feature; we see these census tracts highlighted in red in the scatterplot. A regression line is added to this selected subset (red line) and then another regression line is added to the remaining data which includes data points in the other three quadrants, thus, low-low, low-high, and high-low (blue line). The null hypothesis is that the red and the blue lines have a same slope and intercept are the same (273). The Chow test results are significant and we can reject the null hypothesis. The rejection of the null hypothesis means that the regression line of high percent poverty and high neighboring non-point cancer risk is different than the remaining subset of the data. We then look at the descriptive statistics for the high-high red subset shows significance for the intercept with a p-value=0 and a t-statistic=19.8.

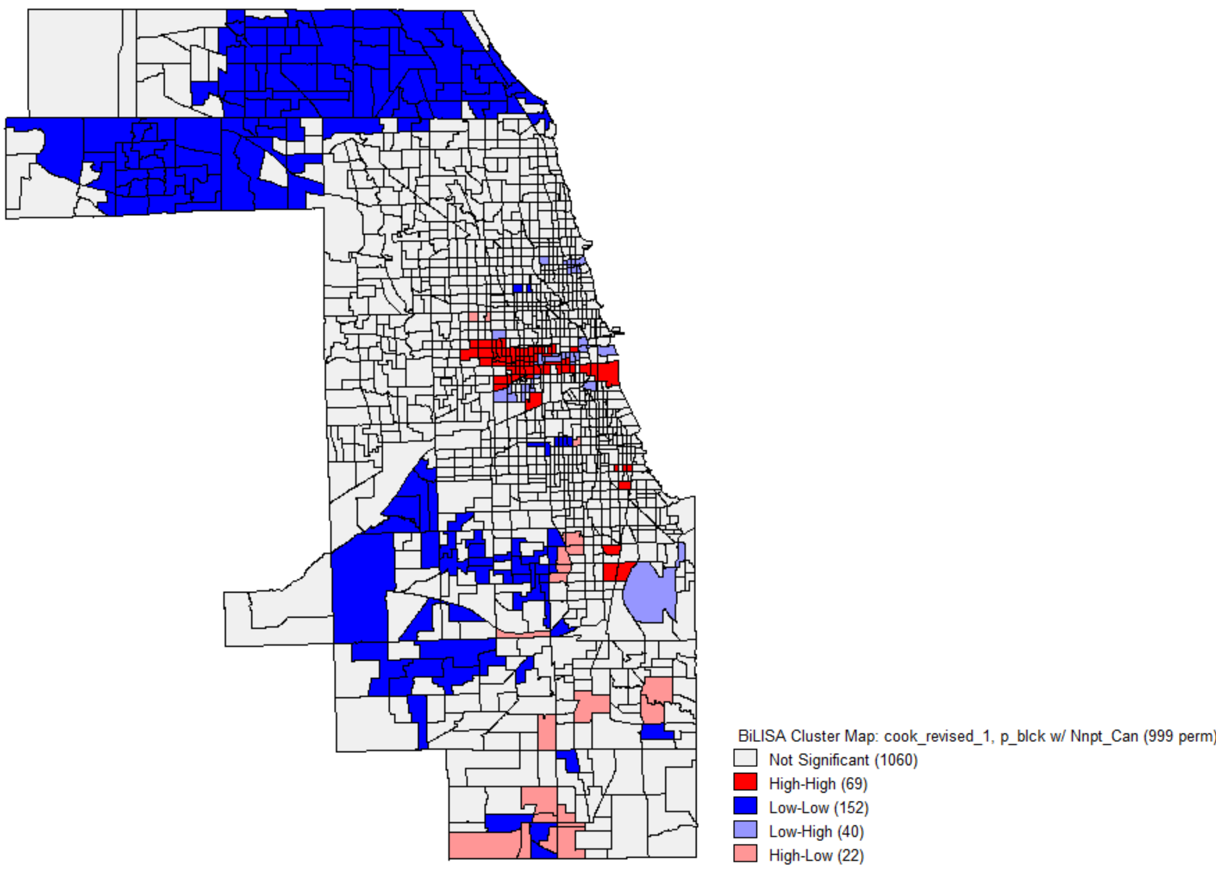

Figure 19. Bivariate lisa cluster map percent black lagged with non-point cancer, cook county.

Lastly, In Figure 19, the Bivariate LISA cluster map shows standardized high percent black is significantly correlated with a standardized high neighboring non-point source cancer risk. There are 69 high-high census tracts clustered again on the west side of Chicago.

The 80 high-high census tracts in Figure 17 were mapped in ArcGIS version 10 software to visualize their location within Cook County, IL. All 80 census tracts are located within the City of Chicago predominantly on the west side with some clusters on the south side with only one census tract outside the city limits. Zooming into this area in Figure 20, we see that the these tracts are in the Chicago Community Areas of Austin, West Garfield Park, East Garfield Park, North Lawndale, Near West Side, South Garfield Park, Washington Park and Woodlawn.

4. Discussion

After global autocorrelation confirms clustering, Anselin Moran's I Local statistic investigates spatial clusters and spatial outliers. Prior to conducting univariate and bivariate LISA diagnostic tests, a sensitivity analysis needs to be completed which includes choosing a weight matrix, number of permutations and significant levels to ensure stability and robustness of the data. Once this is complete, we can go ahead with the tests to discover spatial clusters and spatial outliers and possible spatial correlation. This due diligence is important to ensure that our positive findings are due to possible spatial relationships and not errors created during the analysis. Also, exploratory spatial data analysis has assumptions that need to be met for spatial regression.

The City of Chicago has seventy-seven Chicago Community Areas (CCAs) that are neighborhoods with unique characteristics and defined census data (274). With global spatial autocorrelation, we identified 18 CCAs on the south and west side of Chicago that showed clustering for percent black, percent poverty and non-point cancer risk (275). With the continuation of ESDA in Step II of the integrative methodological framework, we conducted LISA to explore spatial clusters and outliers looking at percent poverty, percent black and non-point cancer risk. Univariate LISA allows us to examine rates at a particular location in relation to neighboring locations. Bivariate LISA results show that standardized high percent poverty is significantly correlated with a standardized high neighboring non-point source cancer risk. We discovered that there are two significant clusters: six CCAs on the west side of Chicago and two CCAs on the south side. Interestingly, these neighborhoods are known disadvantaged areas characterized with high residential segregation, poverty, lack of access to care, food desert, and crime (Figure 20).

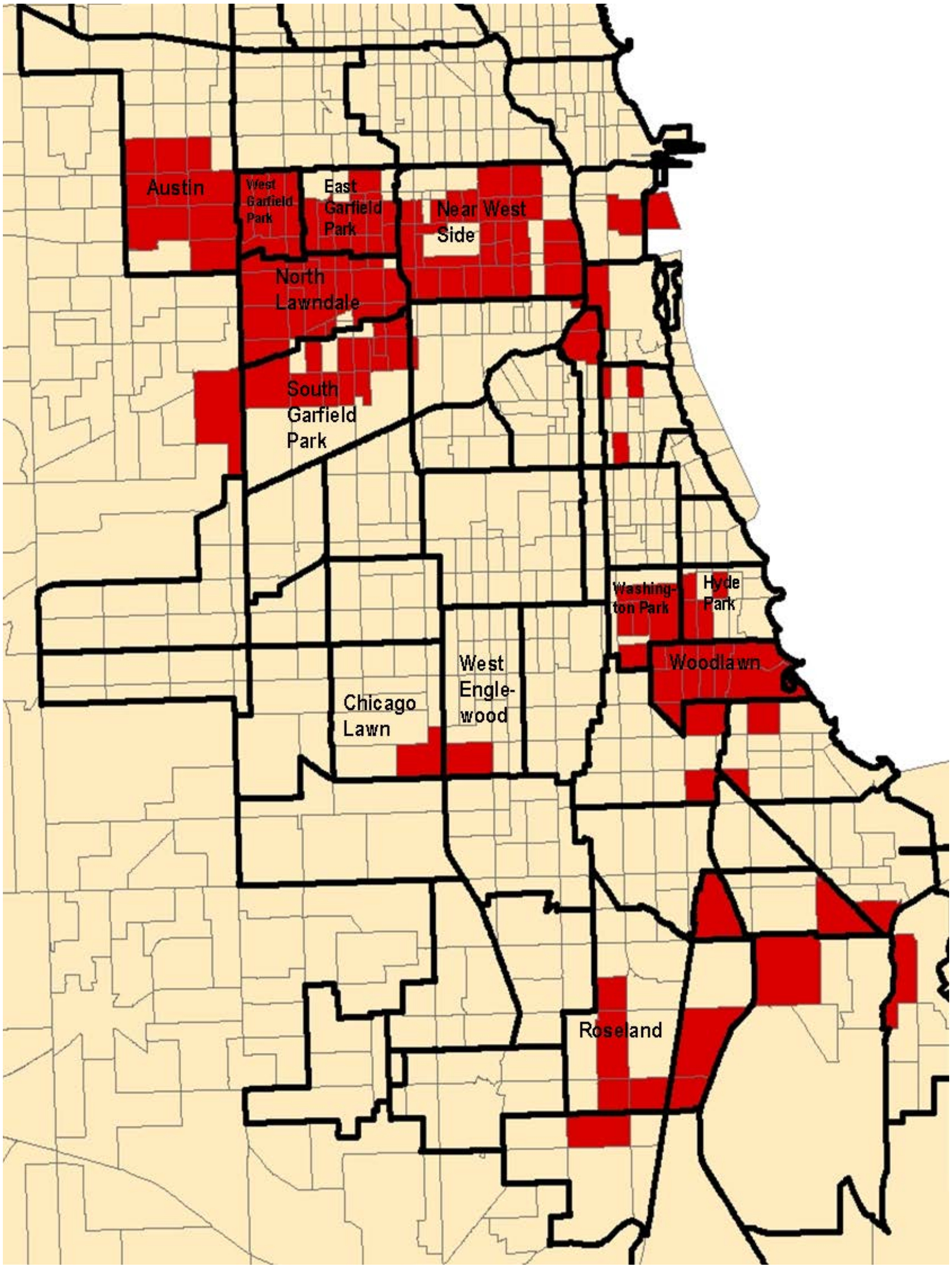

Figure 20. Map zooming into the high percent poverty and high non-point cancer risk census tracts to identify clusters in chicago community areas.

Early environmental justice research explores a number of hazardous sites in at-risk neighborhoods with the premise that more hazardous sites translate to higher exposure to environmental risk factors (276). These hazardous sites are often identified through the USEPA Toxic Release Inventory (TRI) which only includes larger emitters of pollution. The criteria for a facility to report is that it employs 10 or more full-time employees processing over 25,000 pounds or uses over 10,000 pounds of a TRI listed chemical, or it is a TRI-covered industry (i.e. mining, paper mills, etc.) (277). Non-point sources, on the other hand, are stationary locations that produce air pollution but are not accurately documented such as dry cleaners, gas stations or small manufacturers.

Our findings suggest that the built environment in poorer communities may not have easily identifiable exposure sources and further investigation of non-point sources is needed to better understand the type sources in these areas. We will proceed to Step III in the integrative methodological framework with spatial statistics with a focus on spatial regression. In addition, GIS will be integral in identifying and mapping non-point sources to learn more about the types of local business and if there are inherent qualities associated with them in disadvantaged communities.

There are several limitations to our study. There are many choices that can affect outcomes such as the selection of one of the multiple tests for cluster detection, the parameters of sensitivity analysis or the software program producing the results. Also, there are questions of effectiveness of these tests when dealing with numerous databases and determining the reliability, the validity and the completeness of the data (278). For example, the non-point source cancer risk data from the USEPA National Air Toxic Assessment (NATA) incorporates measures from health risk assessments to quantify cancer risk using three different air dispersion models (279). Knowing the challenges and limitations associated with data applications is essential to interpret findings and moving forward with spatial analysis.

5. Conclusion

Understanding the built environment provided invaluable insight into the relationship between vulnerable populations and environmental burden in at-risk communities. Through GIS applications we were able to complete EDA and ESDA which resulted in the discovery of spatial clusters in disadvantaged neighborhoods on the west side of the Chicago. LISA produced significant maps, cluster maps and scatterplots to identify spatial clusters and spatial outliers which identified CCAs for high poverty and high non-point cancer risk. Future analysis to examine spatial regression will confirm the presence of spatial relationships to better understand the role of the built environment with health outcomes.

V. CONCLUSION

Geographic visualization and spatial analysis is a promising approach to identify areas with various characterizations of interest in the field of public health. Moving forward with social research, there are numerous considerations that require an understanding of limitations and assumptions of both the data and the geographic information system (GIS) software to ensure the correct interpretation of outcomes. Spatial autocorrelation and regression methods can be implemented to substantiate claims of a possible spatial relationship for further spatial analysis. To analyze data with a spatial component, it is important to understand the role of EDA and ESDA in the process to ensure interpretable outcomes. This is especially true when working with multiple data sets across multiple disciplines.

ESDA identified seventy census tracts in the upper outlier for both percent poverty and percent blacks in eighteen CCAs in the City of Chicago. The Global Moran's I results further showed clustering for percent poverty and non-point cancer risk. These areas were predominantly poor neighborhoods in the west and south side of Chicago with non-point cancer risk located on the west, north and south sides of Chicago. The North Lawndale neighborhood and the adjacent area had the most census tracts with high non -point cancer risk and cancer incidence. The next step was to continue with ESDA to look at local spatial autocorrelation to confirm our potential variables for spatial regression.

As we move forward with spatial analysis in health disparities research, it is important to understand the applicability of GIS software with social science data. In addition, researchers need to discern parametric approaches in traditional statistical methods and nonparametric approaches in spatial analysis; and, the limitations and assumptions associated with both.methodologies to address these issues will ensure appropriate interpretations of outcomes.

Understanding the built environment provided invaluable insight into the relationship between vulnerable populations and environmental burden in at-risk communities. Through GIS applications we were able to complete EDA and ESDA which resulted in the discovery of spatial clusters in disadvantaged neighborhoods on the west side of the Chicago. LISA produced significant maps, cluster maps and scatterplots to identify spatial clusters and spatial outliers which identified CCAs for high poverty and high non-point cancer risk. Future analysis to examine spatial regression will confirm the presence of spatial relationships to better understand the role of the built environment with health outcomes. Future research needs to further investigate the effectiveness of the integrated analysis methods in explaining varying health outcomes.

This research has a great impact on understanding the role of geographic information systems methodologies in environmental and health disparity research. The integrative methodological framework provides a structured approach for researchers to look at several databases simultaneously to better understand the complexities of the built environment. Also, this study utilized smaller geographic units (census tracts) which definitively showed a correlation between environmental health cancer risk and health disparity indicators such as race and poverty. In addition, this research acknowledges the challenges associated with GIS technologies and the application with social research data to highlight the importance of following a hierarchal process to ensure robustness and stability with the data. Moving forward, it is objective to expand upon these finding to further investigate the role of spatial analysis with social vulnerability and environmental burden.

CITED LITERATURE

1. Payne-Sturges, D., Garcia, L., Lee, C., Zenick, H., Grevatt, P., Sanders III, W., Case, H., and Dankwa-Mullan, I.: Symposium on integrating the science of environmental justice into decision-making at the Environmental Protection Agency: An overview. *Am.J.Public Health* 101: S19–S26, 2011.

2. Payne-Sturges, D., Garcia, L., Lee, C., Zenick, H., Grevatt, P., Sanders III, W., Case, H., and Dankwa-Mullan, I.: Symposium on integrating the science of environmental justice into decision-making at the Environmental Protection Agency: An overview. *Am.J.Public Health* 101: S19–S26, 2011.

3. Payne-Sturges, D., Garcia, L., Lee, C., Zenick, H., Grevatt, P., Sanders III, W., Case, H., and Dankwa-Mullan, I.: Symposium on integrating the science of environmental justice into decision-making at the Environmental Protection Agency: An overview. *Am.J.Public Health* 101: S19–S26, 2011.

4. Cohen, L., Chehimi, S., and Chavez, V.: *Prevention is Primary: Strategies for Community Wellbeing*. Jossey-Bass, 2010.

5. Kawachi, I., and Berkman, L.: *Neighborhoods and Health*. Oxford University Press Inc., 2003.

6. Bullard, R., Mohai, P., Saha, R., and Wright, B.: Toxic Wastes and Race at Twenty: 1987-2007: Grassroots Struggles to Dismantle Environmental Racism in the United States. The United Church of Christ Justice and Witness Ministries. Ohio, 2007.

7. Wilson, S., Fraser-Rahim, H., Williams, E., Zhang, H., Rice, R., and Svendsen, E., et al.: Assessment of the distribution of toxic release inventory facilities in metropolitan charleston: An environmental justice case study. *Am.J.Public Health* 16:e1–e7, 2012.

8. Adeola, F.O.: Environmental hazards, health, and racial inequity in hazardous waste distribution. *Environment and Behavior* 26(1):99–126, 1994.

9. Chavis, B., and Lee, C.: Toxic Wastes and Race in the United States: A National Report on the Racial and Socioeconomic Characteristics of Communities Surrounding Hazardous Waste Sites. United Church of Christ Commission for Racial Justice. New York, 1987.

10. Goldman, B.A., and Fitton, L.: Toxics Wastes and Race Revisited: An Update of the 1987 Report on the Racial and Socioeconomic Characteristics of Communities with Hazardous Waste Sites.Center for Policy Alternatives, 1994.

11. Mohai, P., and Saha, R.: Reassessing racial and socioeconomic disparities in environmental justice research. *Demography* 43(2):383-399, 2006.

12. Sexton, K.: Socioeconomic and racial disparities in environmental health: Is risk assessment part of the problem or part of the solution? *Human and Ecological Risk Assessment: An International Journal* 6(4):561-574, 2000.

13. Young, G., Fox, M., Trush, M., Kanarek,N., Glass, T., and Curriero, F.: Differential exposure to hazardous air pollution in the United States: A multilevel analysis of urbanization and neighborhood socioeconomic deprivation. *International Journal of Environmental Research and Public Health* 9(6):2204-2225, 2012.

14. Gilbert, A.,and Chakraborty J.: Using geographically weighted regression for environmental justice analysis: Cumulative cancer risks from air toxics in Florida. *Social Science Research* 40(1):273-286, 2010.

15. Landrigan, P., Rauh, V., and Galvez, M.: Environmental justice and the health of children. *Mount Sinai Journal of Medicine* 77(2):178-187, 2010.

16. Crowder, K., and Downey, L.: Interneighborhood migration, race, and environmental hazards: modeling microlevel processes of environmental inequality. *American Journal of Sociology* 115(4):1110-1149, 2010.

17. Downey, L., and Hawkins, B.: Race, income, and environmental inequality in the United States. *Sociological Perspectives: SP: official publication of the Pacific Sociological Association* 51(4):759, 2008

18. Anderton, D.L., Anderson, A.B., Oakes, J.M., and Fraser, M.R.: Environmental equity: the demographics of dumping. *Demography* 31(2):229-248, 1994.

19. Bullard, R., Mohai, P., Saha, R., and Wright, B.: Toxic Wastes and Race at Twenty: 1987-2007: Grassroots Struggles to Dismantle Environmental Racism in the United States. The United Church of Christ Justice and Witness Ministries. Ohio, 2007.

20. Wilson, S., Fraser-Rahim, H., Williams, E., Zhang, H., Rice, R., and Svendsen, E., et al.: Assessment of the distribution of toxic release inventory facilities in metropolitan charleston: An environmental justice case study. *Am.J.Public Health* 16:e1-e7, 2012.

21. Mohai, P., and Saha, R.: Reassessing racial and socioeconomic disparities in environmental justice research. *Demography* 43(2):383–399, 2006.

22. Gilbert, A., and Chakraborty J.: Using geographically weighted regression for environmental justice analysis: Cumulative cancer risks from air toxics in Florida. *Social Science Research* 40(1):273–286, 2010.

23. Landrigan, P., Rauh, V.,and Galvez, M.: Environmental justice and the health of children. *Mount Sinai Journal of Medicine* 77(2):178–187, 2010.

24. Luo, J., Hendryx, M.: Environmental carcinogen releases and lung cancer mortality in rural-urban areas of the United States. *Journal of Rural Health* 27(4):342–349, 2011.

25. Downey, L., Dubois, S., Hawkins, B., and Walker, M.: Environmental inequality in metropolitan America. *Organ Environ*. 21(3):270–294, 2008.

26. Payne-Sturges, D., and Gee, G.C.: National environmental health measures for minority and low-income populations: tracking social disparities in environmental health. *Environmental Research* 102(2):154–171, 2006.

27. Morello-Frosch, R., and Jesdale, B.M.: Separate and unequal: residential segregation and estimated cancer risks associated with ambient air toxics in US metropolitan areas. *Environmental Health Perspectives* 114(3):386, 2006.

28. Poehlmann, J.: Children's family environments and intellectual outcomes during maternal incarceration. *Journal of Marriage and Family* 5:1275–1285, 2005.

29. Gomez, S., Glaseer, S., McClure, L., Shema, S., Kealey, M., and Keegan, T., et al.: The California neighborhoods data system: a new resource for examining the impact of neighborhood characteristics on cancer incidence and outcomes in populations. *Cancer Causes Control* 22(4):631–647, 2011.

30. Tian, N., Goovaets, P., Zhan, F., Wilson, J.: Identification of racial disparities in breast cancer mortality: Does scale matter? *International Journal of Health Geographics* 9(35):1–14, 2010.

31. Brennan, P.K.: An intermediate sanction that fosters the mother-child bond: A process evaluation of summit house. *Women & Criminal Justice* 18(3):47–80, 2008.

32. Payne-Sturges, D., and Gee, G.C.: National environmental health measures for minority and low-income populations: tracking social disparities in environmental health. *Environmental Research* 102(2):154–171, 2006.

33. Maantay, J.: Zoning, equity, and public health. *Am.J.Public Health* 91(7):1033–1041, 2001.

34. Woodruff, T.J., Parker, J.D., Kyle, A.D., and Schoendorf, K.C.: Disparities in exposure to air pollution during pregnancy. *Environmental Health Perspectives* 111(7):942, 2003.

35. American Lung Association.: Urban air pollution and health inequities: A workshop report. *Health Perspect* 109 (Suppl 3):357–374, 2001.

36. Evans, G., Colome, S., and Shearer, D.: Psychological reactions to air pollution. *Environ. Res.* 45(1):1–15, 1988.

37. Mohai, P., and Saha, R.: Reassessing racial and socioeconomic disparities in environmental justice research. *Demography* 43(2):383–399, 2006.

38. Faber, D.R., and Krieg, E.J.: Unequal exposure to ecological hazards: environmental injustices in the Commonwealth of Massachusetts. *Environmental Health Perspectives* 110(Suppl 2):277, 2002.

39. Gelobter, M.: Toward a model of environmental discrimination. In: Race and the Incidence of Environmental Hazards: A Time for Discourse, eds. P. Mohai and B. Bryant, pp. 64–81. Colorado, Westview 1992.

40. Gianessi, L.P., Peskin, H.M., and Wolff, E.: The distributional effects of uniform air pollution policy in the United States. *The Quarterly Journal of Economics* 281-301, 1979.

41. Payne-Sturges, D., and Gee, G.C.: National environmental health measures for minority and low-income populations: tracking social disparities in environmental health. *Environmental Research* 102(2):154–171, 2006.

42. Payne-Sturges, D., and Gee, G.C.: National environmental health measures for minority and low-income populations: tracking social disparities in environmental health. *Environmental Research* 102(2):154–171, 2006.

43. Kawachi, I., and Berkman, L.: *Neighborhoods and Health*. Oxford University Press Inc., 2003.

44. Mohai, P., and Saha, R.: Reassessing racial and socioeconomic disparities in environmental justice research. *Demography* 43(2):383–399, 2006.

45. Young, G., Fox, M., Trush, M., Kanarek,N., Glass, T., and Curriero, F.: Differential exposure to hazardous air pollution in the United States: A multilevel analysis of urbanization and neighborhood socioeconomic deprivation. *International Journal of Environmental Research and Public Health* 9(6):2204–2225, 2012.

46. Gilbert, A.,and Chakraborty J.: Using geographically weighted regression for environmental justice analysis: Cumulative cancer risks from air toxics in Florida. *Social Science Research* 40(1):273–286, 2010.

47. Gilbert, A., and Chakraborty, J.: Using geographically weighted regression for environmental justice analysis: Cumulative cancer risks from air toxics in Florida. *Social Science Research* 40:273–286, 2011.

48. Young, L., and Gotway, C.: Using geostatistical methods in the analysis of public health data: The final frontier. *geoENV VII - Geostatistics for Environmental Applications* 16: 89-98, 2010.

49. Chen, J., MacEachren, A., and Guo, D.: Supporting the process of exploring and interpreting space-time multivariate patterns: The visual toolkit. *Cartography and Geographic Information Science* 35:33, 2009.

50. Chen, J., MacEachren, A., and Guo, D.: Supporting the process of exploring and interpreting space-time multivariate patterns: The visual toolkit. *Cartography and Geographic Information Science* 35:33, 2009.

51. Anselin, L.: The moran scatterplot as an ESDA tool to assess local instability in spatial association. In: *Analytical Perspective on GIS*: *GISDATA*, eds. M.M. Fischer, pp. 121–138. Pennsylvania, Taylor & Francis Ltd., 1996.

52. Anselin, L.: The moran scatterplot as an ESDA tool to assess local instability in spatial association. In: *Analytical Perspective on GIS*: *GISDATA*, eds. M.M. Fischer, pp. 121–138. Pennsylvania, Taylor & Francis Ltd., 1996.

53. Anselin, L.: The moran scatterplot as an ESDA tool to assess local instability in spatial association. In: *Analytical Perspective on GIS*: *GISDATA*, eds. M.M. Fischer, pp. 121–138. Pennsylvania, Taylor & Francis Ltd., 1996.

54. Goovaerts, P., and Jacquez G.: Detection of temporal changes in the spatial distribution of cancer rates using local moran's I and geostatistically simulated spatial neutral models. *Journal of Geographical Systems* 7:137–159, 2008.

55. Goovaerts, P., and Jacquez G.: Detection of temporal changes in the spatial distribution of cancer rates using local moran's I and geostatistically simulated spatial neutral models. *Journal of Geographical Systems* 7:137–159, 2008.

56. Arizona State University GeoDa Center for Geospatial Analysis and Computation, Glossary of Key Terms. https://geodacenter.asu.edu/node/390#e. Accessed on June 11, 2013.

57. Anselin, L., Syabri, I., and Youngihn, K.: GeoDa: An introduction to spatial data analysis. *Geographical Analysis* 38(1): 5-22, 2006.

58. USEPA: 2005 National-Scale Air Toxics Assessment: Data Tables State-Specific Emission by County. North Carolina, 2011.

59. USEPA: Background on Risk Characterization, 2002 National-Scale Air Toxics Assessment. North Carolina, 2010.

60. USEPA: An Overview of Methods for EPA's National-Scale Air Toxics Assessment. North Carolina, 2011.

61. USEPA: An Overview of Methods for EPA's National-Scale Air Toxics Assessment. North Carolina, 2011.

62. USEPA: An Overview of Methods for EPA's National-Scale Air Toxics Assessment. North Carolina, 2011.

63. Morello-Frosch R, Jesdale BM. Separate and unequal: residential segregation and estimated cancer risks associated with ambient air toxics in US metropolitan areas. Environmental Health Perspectives 2006;114(3):386.

64. USEPA: An Overview of Methods for EPA's National-Scale Air Toxics Assessment. North Carolina, 2011.

65. USEPA: An Overview of Methods for EPA's National-Scale Air Toxics Assessment. North Carolina, 2011.

66. USEPA: An Overview of Methods for EPA's National-Scale Air Toxics Assessment. North Carolina, 2011.

67. Illinois Department of Public Health: Cancer in Illinois. In: Illinois State Cancer Registry (ISCR). http://www.idph.state.il.us/cancer/statistics.htm. Accessed on June 10, 2013.

68. Bullard, R., Mohai, P., Saha, R., and Wright, B.: Toxic Wastes and Race at Twenty: 1987-2007: Grassroots Struggles to Dismantle Environmental Racism in the United States. The United Church of Christ Justice and Witness Ministries. Ohio, 2007.

69. Wilson, S., Fraser-Rahim, H., Williams, E., Zhang, H., Rice, R., and Svendsen, E., et al.: Assessment of the distribution of toxic release inventory facilities in metropolitan charleston: An environmental justice case study. *Am.J.Public Health* 16:e1–e7, 2012.

70. Adeola, F.O.: Environmental hazards, health, and racial inequity in hazardous waste distribution. *Environment and Behavior* 26(1):99–126, 1994.

71. Chavis, B., and Lee, C.: Toxic Wastes and Race in the United States: A National Report on the Racial and Socioeconomic Characteristics of Communities Surrounding Hazardous Waste Sites. United Church of Christ Commission for Racial Justice. New York, 1987.

72. Goldman, B.A., and Fitton, L.: Toxics Wastes and Race Revisited: An Update of the 1987 Report on the Racial and Socioeconomic Characteristics of Communities with Hazardous Waste Sites.Center for Policy Alternatives, 1994.

73. Mohai, P., and Saha, R.: Reassessing racial and socioeconomic disparities in environmental justice research. *Demography* 43(2):383-399, 2006.

74. Sexton, K.: Socioeconomic and racial disparities in environmental health: Is risk assessment part of the problem or part of the solution? *Human and Ecological Risk Assessment: An International Journal* 6(4):561-574, 2000.

75. Young, G., Fox, M., Trush, M., Kanarek,N., Glass, T., and Curriero, F.: Differential exposure to hazardous air pollution in the United States: A multilevel analysis of urbanization and neighborhood socioeconomic deprivation. *International Journal of Environmental Research and Public Health* 9(6):2204-2225, 2012.

76. Gilbert, A.,and Chakraborty J.: Using geographically weighted regression for environmental justice analysis: Cumulative cancer risks from air toxics in Florida. *Social Science Research* 40(1):273-286, 2010.

77. Landrigan, P., Rauh, V., and Galvez, M.: Environmental justice and the health of children. *Mount Sinai Journal of Medicine* 77(2):178-187, 2010.

78. Crowder, K., and Downey, L.: Interneighborhood migration, race, and environmental hazards: modeling microlevel processes of environmental inequality. *American Journal of Sociology* 115(4):1110-1149, 2010.

79. Downey, L., and Hawkins, B.: Race, income, and environmental inequality in the United States. *Sociological Perspectives: SP: official publication of the Pacific Sociological Association* 51(4):759, 2008.

80.. Anderton, D.L., Anderson, A.B., Oakes, J.M., and Fraser, M.R.: Environmental equity: the demographics of dumping. *Demography* 31(2):229-248, 1994.

81. Bullard, R., Mohai, P., Saha, R., and Wright, B.: Toxic Wastes and Race at Twenty: 1987-2007: Grassroots Struggles to Dismantle Environmental Racism in the United States. The United Church of Christ Justice and Witness Ministries. Ohio, 2007.

82. Wilson, S., Fraser-Rahim, H., Williams, E., Zhang, H., Rice, R., and Svendsen, E., et al.: Assessment of the distribution of toxic release inventory facilities in metropolitan charleston: An environmental justice case study. *Am.J.Public Health* 16:e1-e7, 2012.

83. Mohai, P., and Saha, R.: Reassessing racial and socioeconomic disparities in environmental justice research. *Demography* 43(2):383–399, 2006.

84. Gilbert, A., and Chakraborty J.: Using geographically weighted regression for environmental justice analysis: Cumulative cancer risks from air toxics in Florida. *Social Science Research* 40(1):273–286, 2010.

85. Landrigan, P., Rauh, V.,and Galvez, M.: Environmental justice and the health of children. *Mount Sinai Journal of Medicine* 77(2):178–187, 2010.

86. Luo, J., Hendryx, M.: Environmental carcinogen releases and lung cancer mortality in rural-urban areas of the United States. *Journal of Rural Health* 27(4):342–349, 2011.

87. Downey, L., Dubois, S., Hawkins, B., and Walker, M.: Environmental inequality in metropolitan America. *Organ Environ*. 21(3):270–294, 2008.

88. Payne-Sturges, D., and Gee, G.C.: National environmental health measures for minority and low-income populations: tracking social disparities in environmental health. *Environmental Research* 102(2):154–171, 2006.

89. Morello-Frosch, R., and Jesdale, B.M.: Separate and unequal: residential segregation and estimated cancer risks associated with ambient air toxics in US metropolitan areas. *Environmental Health Perspectives* 114(3):386, 2006.

90. Poehlmann, J.: Children's family environments and intellectual outcomes during maternal incarceration. *Journal of Marriage and Family* 5:1275–1285, 2005.

91. Gomez, S., Glaseer, S., McClure, L., Shema, S., Kealey, M., and Keegan, T., et al.: The California neighborhoods data system: a new resource for examining the impact of neighborhood characteristics on cancer incidence and outcomes in populations. *Cancer Causes Control* 22(4):631–647, 2011.

92. Tian, N., Goovaets, P., Zhan, F., Wilson, J.: Identification of racial disparities in breast cancer mortality: Does scale matter? *International Journal of Health Geographics* 9(35):1–14, 2010.

93. Brennan, P.K.: An intermediate sanction that fosters the mother-child bond: A process evaluation of summit house. *Women & Criminal Justice* 18(3):47–80, 2008.

94. Schulz, A.J., Williams, D.R., Israel, B.A., and Lempert, L.B.: Racial and spatial relations as fundamental determinants of health in Detroit. *Milbank Quarterly* 80(4):677-707, 2002.

95. Pastor Jr., M., Sadd. J,L,, and Morello-Frosch R.: Waiting to inhale: The demographics of toxic air release facilities in 21st-century California. *Social Science Quarterly* 85(2):420-440, 2004.

96. Maantay, J.: Zoning, equity, and public health. *Am.J.Public Health* 91(7):1033 – 1041, 2001.

97. Schmidt, C.W.: The market for pollution. *Environmental Health Perspectives* 109(8):A378, 2001.

98. Woodruff, T.J., Parker, J.D., Kyle, A.D., and Schoendorf, K.C.: Disparities in exposure to air pollution during pregnancy. *Environmental Health Perspectives* 111(7):942, 2003.

99.. American Lung Association.: Urban air pollution and health inequities: A workshop report. *Health Perspect* 109 (Suppl 3):357–374, 2001.

100. Evans, G., Colome, S., and Shearer, D.: Psychological reactions to air pollution. *Environ. Res.* 45(1):1–15, 1988.

101. Mohai, P., and Saha, R.: Reassessing racial and socioeconomic disparities in environmental justice research. *Demography* 43(2):383–399, 2006.

102. Faber, D.R., and Krieg, E.J.: Unequal exposure to ecological hazards: environmental injustices in the Commonwealth of Massachusetts. *Environmental Health Perspectives* 110(Suppl 2):277, 2002.

103. Gelobter, M.: Toward a model of environmental discrimination. In: Race and the Incidence of Environmental Hazards: A Time for Discourse, eds. P. Mohai and B. Bryant, pp. 64–81. Colorado, Westview 1992.

104. Gianessi, L.P., Peskin, H.M., and Wolff, E.: The distributional effects of uniform air pollution policy in the United States. *The Quarterly Journal of Economics* 281–301, 1979.

105. Faber, D.R., and Krieg, E.J.: Unequal exposure to ecological hazards: environmental injustices in the Commonwealth of Massachusetts. *Environmental Health Perspectives* 110(Suppl 2):277, 2002.

106. Faber, D.R., and Krieg, E.J.: Unequal exposure to ecological hazards: environmental injustices in the Commonwealth of Massachusetts. *Environmental Health Perspectives* 110(Suppl 2):277, 2002

107. Frickel, S.: Toxic bodies/toxic environments: An interdisciplinary forum. *Environmental History Issue* 13(4):643–650, 2008.

108. Logan, J.R., and Molotch, H.L.: *Urban Fortunes: The Political Economy of Place*. University of California Press, 1987.

109. Morello-Frosch, R.A.: Discrimination and the political economy of environmental inequality. *Environment and Planning C: Government & Policy* 20(4):477–496, 2002.

110. Sinton, J.: Rights discourse and mandatory HIV testing for pregnant women and newborns. *JL & Pol'y* 6:187, 1997.

111. Lopez, R.: Segregation and black/white differences in exposure to air toxics in 1990. *Environmental Health Perspectives* 110(Suppl 2):289, 2002.

112. Subramanian, D.: Riots and the immigrant community. *Economic and Political Weekly* 5156–5158, 2005.

113. Wilson, W.J.: *When Work Disappears*. Random House, 1997.

114. Wilson, S.: Environmental justice movement: a review of history, research, and public health issues. *J Public Manage Soc Policy* 16:19-50, 2010.

115. Payne-Sturges, D., and Gee, G.C.: National environmental health measures for minority and low-income populations: tracking social disparities in environmental health. *Environmental Research* 102(2):154–171, 2006.

116. Maantay, J.: Zoning, equity, and public health. *Am.J.Public Health* 91(7):1033–1041, 2001.

117. Jones, C.: The impact of racism on health. *Ethnicity & Disease* 12(1):S2, 2002.

118. Williams, D.R., and Collins, C. Racial residential segregation: a fundamental cause of racial disparities in health. *Public Health Reports* 116(5):404, 2001.

119. Williams, D.R., and Collins C.: Reparations. *American Behavioral Scientist* 47(7):977, 2004.

120. Stuart, A.L., Mudhasakul, S., and Sriwatanapongse. W.: The social distribution of neighborhood-scale air pollution and monitoring protection. *Journal of the Air & Waste Management Association* 59(5):591–602, 2009.

121. Hockman, E.M., and Morris, C.M.: Progress towards environmental justice: A five-year perspective of toxicity, race and poverty in Michigan, 1990-1995. *Journal of Environmental Planning and Management* 41(2):157–176, 1998.

122. Mohai, P., and Saha, R.: Racial inequality in the distribution of hazardous waste: a national-level reassessment. *Social Problems* 54(3):343–370, 2007.

123. Reynolds, P., Von Behren, J., Gunier, R., Goldberg, D., Hertz, A., and Smith, D.: Traffic patterns and childhood cancer incidence rates in California, United States. *Cancer Causes Contro*l 13:665–673, 2002.

124. Shenassa, E., Stubbendick, A., and Brown, M.: Social disparities in housing and related pediatric injury: A multilevel study. *Am.J.Public Health* 94:633–639, 2004.

125. Smith, D., and Jarjoura G.: Social structure and criminal victimization. *Journal of Research in Crime and Delinquency* 25(1):27–52, 1988.

126. Sampson, R.: Local friendship ties and community attachment in mass society: A multi-level systemic model. *American Sociological Review* 53:766-779,1988.

127. Bursik, J., and Grasmick H.: *Neighborhoods and Crime: The Dimensions of Effective Community Control*. Lexington Books, 1993.

128. Berkman, L.F., and Kawachi, I.: *Social Epidemiology*. Oxford University Press, 2000.

129. Kawachi, I., Kennedy, B.P., Lochner, K., and Prothrow-Stith, D.: Social capital, income inequality, and mortality. *Am.J.Public Health* 87(9):1491–1498, 1997.

130. Sampson, R., Raudenbush, S.,and Earls, F.: Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science* 277(15):918–924, 1997.

131. Sampson, R., and Morenoff, J.: Spatial (dis)advantage and homicide in Chicago neighborhoods. In: *Spatially Integrated Social Science*, eds. J. MGaD New York, Oxford University Press, 2004.

132. Sampson, R., Morenoff, J., and Earls, F.: Beyond social capital: Spatial dynamics of collective efficacy for children. *Am Sociol Rev* 64(5):633–660,1999.

133. Perlin, S.A., Wong, D., and Sexton K.: Residential proximity to industrial sources of air pollution: interrelationships among race, poverty, and age. *Journal of the Air & Waste Management Association* 51(3):406-421, 2001.

134. Wilson, W.J.: *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. University Of Chicago Press, 1990.

135. Gee, G.C., and Payne-Sturges, D.C.: Environmental health disparities: a framework integrating psychosocial and environmental concepts. *Environmental Health Perspectives* 112(17):1645, 2004.

136. Berkman, L.F., Glass, T., Brissette, I., and Seeman, T.E:. From social integration to health: Durkheim in the new millennium. *Social Science & Medicine* 51(6):843–857, 2000.

137. Marmot, M.: *Multi-Level Approaches to Understanding Social Determinants, in Social Epidemiology*. Oxford University Press, 1999.

138. Susser, M.: Does risk factor epidemiology but epidemiology at risk? peering into the future. *J Epidemiol Communtiy Health* 52:608–611, 1998.

139. Galea, S., Nandi, A., Vlahov, D.: The social epidemiology of substance use. *Epidemiologic Reviews* 26:36-52, 2004.

1401. Sampson, R.: The neighborhood context of wellbeing. *Perspectives in Biology and Medicine* 46(Suppl. 3):S53–S64, 2003.

141. Bullard, R.D., and Johnson, G.S.: Environmentalism and public policy: Environmental justice: Grassroots activism and its impact on public policy decision making. *Journal of Social Issues* 56(3):555–578, 2000.

142. Boeglin, M., Wessels, D., and Henshel, D.: An investigation of the relationship between air emissions of volatile organic compounds and the incidence of cancer in Indiana counties. *Environ Res* 100(2):242–254, 2006.

143. Clapp, R., Jacobs, M., and Loechler E.: Environmental and occupational causes of cancer new evidence, 2005–2007. *Rev Environ Health* 23(1):1–37, 2008.

144. Boffetta, P., and Nyberg, F.: Contribution of environmental factors to cancer risk. *Br Med Bull* 68(1):71–94, 2003.

145. Gammon, M., Wolff, M., Neugut, A., Eng, S., Teitelbaum, S., and Britton, J., et al.: Environmental toxins and breast cancer on Long Island organochlorine compoundlevels in blood. *Cancer Epidemiology Biomarkers & Prevention* 11:686–697, 2002.

146. Mohai, P., Pellow, D., and Roberts, J.: Environmental justice. *Annu. Rev. Environ. Resour*.34:405–430, 2009.

147. Morrison, R.S., Wallenstein, S., Natale, D.K., Senzel, R.S., and Huang, L.L.: " We don't carry that"--failure of pharmacies in predominantly nonwhite neighborhoods to stock opioid analgesics. The New England Journal of Medicine 342(14):1023, 2000.

148. Massey, D.S., and Shibuya, K.: Unraveling the tangle of pathology: The effect of spatially concentrated joblessness on the well-being of african americans. *Social Science Research* 4:352–366, 1995.

149. Bach, P., Pham, H., Schrag, D., Tate, R., and Hargraves, J.: Primary care physicians who treat blacks and whites. *N Engl J Med* 351:575–584, 2004.

150. Hiatt, R., and Breen, N. The social determinants of cancer: A challenge for transdiciplinary science. American Journal of Preventative Medicine 35(2 Suppl):S141–S150, 2008.

151. Lopez, R., and Hynes P.: Obesity, physical activity, and the urban environment: Public health research needs. *Environmental Health: A Global Access Science Source* 5(25):1–10, 2006.

152. Heinrich, K., Lee, R., Regan, G., Reese-Smith, J., Howard, H., and Haddock, C., et al.: How does the built environment relate to BMI and obesity prevalence among public housing residents? *Am J Health Promot* 22(3):187–194, 2008.

153. Cohen, L., Chehimi, S., and Chavez, V.: *Prevention is Primary: Strategies for Community Wellbeing.* Prevention Institute, 2010.

154. Kawachi, I., and Berkman, L.: *Neighbourhoods and Health*. Oxford University Press Inc, 2003.

155. Maranville, A., Ting, T., and Zhang, Y.: An environmental justice analysis: superfund sites and surrounding communities in Illinois. *Environmental Justice* 2(2):49–56, 2009.

156. Bevc, C., Marshall, B., and Picou J.: Environmental justice and toxic exposure: Toward a spatial model of physical health and psychological well-being. *Social Science Research* 36:48–67, 2007.

157. Maantay, J.: Asthma and air pollution in the Bronx: Methodological and data considerations in using GIS for environmental justice and health research. *Health & Place* 13:32–56, 2007.

158. Downey, L.: Environmental Injustice: Is Race Or Income a Better Predictor?: National Emergency Training Center; 1998.

159. Bowen, W.M., Salling, M.J., Haynes, K.E., and Cyran, E.J.: Toward environmental justice: Spatial equity in Ohio and Cleveland. *Annals of the Association of American Geographers* 85(4):641–663, 1995.

160, Stretesky, P., and Hogan M.J.: Environmental justice: An analysis of superfund sites in Florida. *Soc. Probs*. 45:268, 1998.

161. Jones, C.: The impact of racism on health. *Ethnicity & Disease* 12(1):S23, 2002.

162. Henschke, C., Yankelevitz, D., Libby, D., Pasmantier, M., Smith, J., and Miettinen, O.: Survival of patients with stage I lung cancer detected on CT screening. *N Engl J Med* 355(17):1763–1771, 2006.

163. Gadgeel, S., Severson, R., Kau, Y., Graff, J., Weiss, L., and Kalemkerian, G.: Impact of race in lung cancer: Analysis of temporal trends from a surveillance, epidemiology, and end results database. *Chest* 120(1):55–63, 2001.

164. Orsi, J.M., Margellos-Anast, H., and Whitman, S.: Black-White health disparities in the United States and Chicago: a 15-year progress analysis. *Am J Public Health*. 100(2):349–56, 2010.

165. Orsi, J.M., Margellos-Anast, H., and Whitman, S.: Black-White health disparities in the United States and Chicago: a 15-year progress analysis. *Am J Public Health*. 100(2):349–56, 2010.

166. Clarke, C., West, D., Edwards, B., Figgs, L., Kerner, J., and Schwartz, A.: Existing data on breast cancer in African-American women: What we know and what we need to know. *Cancer* 97(1 Suppl):211–221, 2003.

167. Wan, N., Zhan, F., and Cai, Z.: Socioeconomic Disparities in prostate cancer mortality and the impact of geographic scale. *South Med Journal* 104(8):553–559. 2011.

168. Bilheimer, L., and Klein R.: Data and measurement issues in the analysis of health disparities. Health Services Research 45:1489–1507, 2010.

169. Griffith, D., Moy, E., Reischl, T., and Dayton, E.: National data for monitoring and evaluating racial and ethnic health inequities: Where do we go from here? *Health Education and Behavior* 33:470-487, 2006.

170. Ghetian, C., Rarrott, R., Volkman, J., and Lengerich, E.: Cancer registry policies in the United States and geographic information systems applications in comprehensive cancer control. *Health Policy* 87:185-193, 2008.

171. USEPA: 2005 National-Scale Air Toxics Assessment: Data Tables State-Specific Emission by County. North Carolina, 2011.

172. USEPA: Background on Risk Characterization, 2002 National-Scale Air Toxics Assessment. North Carolina, 2010.

173. USEPA: An Overview of Methods for EPA's National-Scale Air Toxics Assessment. North Carolina, 2011.

174. USEPA: An Overview of Methods for EPA's National-Scale Air Toxics Assessment. North Carolina, 2011.

175. USEPA: An Overview of Methods for EPA's National-Scale Air Toxics Assessment. North Carolina, 2011.

176. Tian, N., Goovaets, P., Zhan, F., Wilson, J.: Identification of racial disparities in breast cancer mortality: Does scale matter? *International Journal of Health Geographics* 9(35):1–14, 2010.

177. USEPA: An Overview of Methods for EPA's National-Scale Air Toxics Assessment. North Carolina, 2011.

178. USEPA: An Overview of Methods for EPA's National-Scale Air Toxics Assessment. North Carolina, 2011.

179. USEPA: An Overview of Methods for EPA's National-Scale Air Toxics Assessment. North Carolina, 2011.

180. USEPA: Framework for Cumulative Risk Assessment. EPA/600/P-02/001F, Environmental Protection Agency, Risk Assessment Forum, Office of Research and Development., Washington, DC, 2003.

181. USEPA: An Overview of Methods for EPA's National-Scale Air Toxics Assessment. North Carolina, 2011.

182. USEPA: An Overview of Methods for EPA's National-Scale Air Toxics Assessment. North Carolina, 2011.

183. USEPA. The Toxics Release Inventory (TRI) and factors to consider when using TRI data, 2009. http://www.epa.gov/tri/triprogram/FactorsToConPDF.pdf . Accessed on June 5, 2013.

184. Osiecki, K.M., Kim, S., Chukwudozie, I.B., Calhoun, E.A.: Utilizing exploratory spatial data analysis to examine health and environmental disparities in disadvantaged neighborhoods. *Environmental Justice* 6(3):81–87, 2013.

185. USEPA Toxics Release Inventory (TRI) Program Factsheet http://www.epa.gov/tri/triprogram/RY_2011_TRI_Factsheet.pdf. Accessed on June 6, 2013.

186. Illinois Department of Public Health. Cancer in Illinois. In: Illinois State Cancer Registry (ISCR), http://www.idph.state.il.us/cancer/statistics.htm. Accessed on June 10, 2013.

187. Department of Housing and Urban Development. HUD USPS zip code crosswalk files. In: Development DoHaU, 2012. http://www.huduser.org/portal/datasets/usps_crosswalk.html. Accessed on August 30 2012.

188. Rytkönen, M.P.: Not all maps are equal: GIS and spatial analysis in epidemiology. *International Journal of Circumpolar Health* 63(1), 2004.

189. Young, L., and Gotway, C.: Using geostatistical methods in the analysis of public health data: The final frontier. *geoENV VII - Geostatistics for Environmental Applications* 16: 89-98, 2010.

190. Guo, D.,Gahegan, M., MacEachren, A.M., and Zhou, B..: Multivariate analysis and geovisualization with an integrated geographic knowledge discovery approach. *Cartography and Geographic Information Science*. 32(2):113–132, 2005.

191. Arizona State University GeoDa Center for Geospatial Analysis and Computation, Glossary of Key Terms. https://geodacenter.asu.edu/node/390#e. Accessed on June 11, 2013.

192. Osiecki, K.M., Kim, S., Chukwudozie, I.B., Calhoun, E.A.: Utilizing exploratory spatial data analysis to examine health and environmental disparities in disadvantaged neighborhoods. *Environmental Justice* 6(3):81–87, 2013.

193. Arizona State University GeoDa Center for Geospatial Analysis and Computation, Glossary of Key Terms. https://geodacenter.asu.edu/node/390#e. Accessed on June 11, 2013.

194. Hockman, H., and Morris, C.: Progress towards environmental justice: A five-year perspective of toxicity, race and poverty in Michigan, 1990–1995. *Journal of Environmental Planning and Management* 41:157–176, 1998.

195. Bullard, R.D., and Johnson, G.S.: Environmentalism and public policy: Environmental justice: Grassroots activism and its impact on public policy decision making. *Journal of Social Issues* 56(3):555–578, 2000.

196. Boeglin, M., Wessels, D., and Henshel, D.: An investigation of the relationship between air emissions of volatile organic compounds and the incidence of cancer in Indiana counties. *Environ Res* 100(2):242–254, 2006.

197. Clapp, R., Jacobs, M., and Loechler E.: Environmental and occupational causes of cancer new evidence, 2005–2007. *Rev Environ Health* 23(1):1–37, 2008.

198. Boffetta, P., and Nyberg, F.: Contribution of environmental factors to cancer risk. *Br Med Bull* 68(1):71–94, 2003.

199. Cohen, L., Chehimi, S., and Chavez, V.: *Prevention is Primary: Strategies for Community Wellbeing.* Prevention Institute, 2010.

200. Kawachi, I., and Berkman, L.: *Neighbourhoods and Health*. Oxford University Press Inc, 2003.

201. Gilbert, A., and Chakraborty J.: Using geographically weighted regression for environmental justice analysis: Cumulative cancer risks from air toxics in Florida. *Social Science Research* 40(1):273-286, 2010.

202. Guo, D.,Gahegan, M., MacEachren, A.M., and Zhou, B.: Multivariate analysis and geovisualization with an integrated geographic knowledge discovery approach. *Cartography and Geographic Information Science*. 32(2):113–132, 2005.

203. Wilkinson, P., Grundy,C., Landon, M., Stevenson, S.: GIS in Public Health. In: *GIS and Health: GISDATA 6,* eds. A. Gattrell and L. Taylor, pp. 179–189. Pennsylvania, Taylor & Francis, 2004.

204. Young, L., and Gotway, C.: Using geostatistical methods in the analysis of public health data: The final frontier. *geoENV VII - Geostatistics for Environmental Applications* 16: 89-98, 2010.

205. Chen, J., MacEachren, A., and Guo, D.: Supporting the process of exploring and interpreting space-time multivariate patterns: The visual toolkit. *Cartography and Geographic Information Science* 35:33, 2009.

206. Chen, J., MacEachren, A., and Guo, D.: Supporting the process of exploring and interpreting space-time multivariate patterns: The visual toolkit. *Cartography and Geographic Information Science* 35:33, 2009.

207. Chen, J., MacEachren, A., and Guo, D.: Supporting the process of exploring and interpreting space-time multivariate patterns: The visual toolkit. *Cartography and Geographic Information Science* 35:33, 2009.

208. Anselin, L.: The moran scatterplot as an ESDA tool to assess local instability in spatial association. In: *Analytical Perspective on GIS*: *GISDATA*, eds. M.M. Fischer, pp. 121–138. Pennsylvania, Taylor & Francis Ltd., 1996.

209. Anselin, L.: The moran scatterplot as an ESDA tool to assess local instability in spatial association. In: *Analytical Perspective on GIS*: *GISDATA*, eds. M.M. Fischer, pp. 121–138. Pennsylvania, Taylor & Francis Ltd., 1996.

210. Wilkinson, P., Grundy,C., Landon, M., Stevenson, S.: GIS in Public Health. In: *GIS and Health: GISDATA 6,* eds. A. Gattrell and L. Taylor, pp. 179–189. Pennsylvania, Taylor & Francis, 2004.

211. Wilkinson, P., Grundy,C., Landon, M., Stevenson, S.: GIS in Public Health. In: *GIS and Health: GISDATA 6,* eds. A. Gattrell and L. Taylor, pp. 179–189. Pennsylvania, Taylor & Francis, 2004.

212. Guo, D.,Gahegan, M., MacEachren, A.M., and Zhou, B.: Multivariate analysis and geovisualization with an integrated geographic knowledge discovery approach. *Cartography and Geographic Information Science*. 32(2):113–132, 2005.

213. Guo, D.,Gahegan, M., MacEachren, A.M., and Zhou, B.: Multivariate analysis and geovisualization with an integrated geographic knowledge discovery approach. *Cartography and Geographic Information Science*. 32(2):113–132, 2005.

214. Goovaerts, P., and Jacquez G.: Detection of temporal changes in the spatial distribution of cancer rates using local moran's I and geostatistically simulated spatial neutral models. *Journal of Geographical Systems* 7:137–159, 2008.

215. Goovaerts, P., and Jacquez G.: Detection of temporal changes in the spatial distribution of cancer rates using local moran's I and geostatistically simulated spatial neutral models. *Journal of Geographical Systems* 7:137–159, 2008.

216. USEPA: An Overview of Methods for EPA's National-Scale Air Toxics Assessment. North Carolina, 2011.

217. Wilkinson, P., Grundy, C., Landon, M., Stevenson, S.: GIS in Public Health. In: GIS and Health: GISDATA 6, ed. A. Gattrell and L. Taylor, pp. 179–189. Pennsylvania, Taylor & Francis, 2004.

218. Krieger, N.: Place, space, and health: GIS and epidemiology," *Epidemiology* 14:384-385, 2003.

219. Gilboa,S., Mendola, P., Olshan, A., Harness, C., Loomis, D.,Langlois, P.,Savitz, D., Herring, A.: Comparison of residential geocoding methods in population-based study of air quality and birth defects. *Environmental Res*earch 101:256-262, 2006.

220. Wilkinson, P., Grundy,C., Landon, M., Stevenson, S.: GIS in Public Health. In: *GIS and Health: GISDATA 6,* eds. A. Gattrell and L. Taylor, pp. 179–189. Pennsylvania, Taylor & Francis, 2004.

221. Wilkinson, P., Grundy,C., Landon, M., Stevenson, S.: GIS in Public Health. In: *GIS and Health: GISDATA 6,* eds. A. Gattrell and L. Taylor, pp. 179–189. Pennsylvania, Taylor & Francis, 2004.

222. Rytkönen, M.P.: Not all maps are equal: GIS and spatial analysis in epidemiology. *International Journal of Circumpolar Health* 63(1), 2004.

223. Krieger, N.: Place, space, and health: GIS and epidemiology," *Epidemiology* 14:384-385, 2003.

224. Krieger, N.: Place, space, and health: GIS and epidemiology," *Epidemiology* 14:384-385, 2003.

225. Wilkinson, P., Grundy, C., Landon, M., Stevenson, S.: GIS in Public Health. In: GIS and Health: GISDATA 6, ed. A. Gattrell and L. Taylor, pp. 179–189. Pennsylvania, Taylor & Francis, 2004.

226. Wilkinson, P., Grundy, C., Landon, M., Stevenson, S.: GIS in Public Health. In: GIS and Health: GISDATA 6, ed. A. Gattrell and L. Taylor, pp. 179–189. Pennsylvania, Taylor & Francis, 2004.

227. Frumkin, H., and Jackson, R.J.: *Health and the Environment in the Southeastern United States.* National Academies Press, 2002.

228. Frumkin, H.: Healthy places: Exploring the evidence. *Am.J.Public Health* 93(9):1451, 2003.

229. Northridge, M.E., Stover, G.N., and Rosenthal, J.E., Sherard D.: Environmental equity and health: Understanding complexity and moving forward. *Am.J.Public Health* 93(2):209, 2003.

230. Davison, K.K., and Lawson C.T.: Do attributes in the physical environment influence children's physical activity? A review of the literature. *International Journal of Behavioral Nutrition and Physical Activity* 3(1):19, 2006.

231. Giles-Corti, B., and Donovan, R.J.: Relative influences of individual, social environmental, and physical environmental correlates of walking. *Am.J.Public Health* 93(9):1583, 2003.

232. Gee, G.C., and Payne-Sturges, D.C.: Environmental health disparities: a framework integrating psychosocial and environmental concepts. *Environmental Health Perspectives* 112(17):1645, 2004.

233. Berkman, L.F., Glass, T., Brissette, I., and Seeman, T.E.: From social integration to health: Durkheim in the new millennium. *Social Science & Medicine* 51(6):843–857, 2000.

234. Marmot, M.: *Multi-Level Approaches to Understanding Social Determinants, in Social Epidemiology*. Oxford University Press, 1999.

235. Susser, M.: Does risk factor epidemiology but epidemiology at risk? Peering into the future. *J Epidemiol Communtiy Health* 52:608–611, 1998.

236. Galea, S., Nandi, A., and Vlahov, D.: The social epidemiology of substance use. *Epidemiologic Reviews* 26:36–52, 2004.

237. Sampson, R.: The neighborhood context of wellbeing. *Perspectives in Biology and Medicine* 46(Suppl. 3):S53–S64, 2003.

238. Lee, C.: Environmental justice: building a unified vision of health and the environment. *Environmental Health Perspectives* 110(Suppl 2):141, 2002

239. Sexton, K.: Socioeconomic and racial disparities in environmental health: Is risk assessment part of the problem or part of the solution? *Human and Ecological Risk Assessment: An International Journal* 6(4):561-574, 2000.

240. Yen, I.H., Syme, S.L.: The social environment and health: a discussion of the epidemiologic literature. *Annual Review of Public Health* 20(1):287-308, 1999.

241. Corburn, J.: Combining community-based research and local knowledge to confront asthma and subsistence-fishing hazards in Greenpoint/Williamsburg, Brooklyn, New York. *Environmental Health Perspectives* 110(Suppl 2):241, 2002.

242. Morello-Frosch, R.A.: Discrimination and the political economy of environmental inequality. *Environment and Planning C: Government & Policy* 20(4):477-496, 2002.

243. Morello-Frosch, R., Pastor Jr., M., Porras, C., and Sadd, J.: Environmental justice and regional inequality in southern California: implications for future research. *Environmental Health Perspectives* 110(Suppl 2):149, 2002.

244. Burger, J., Gaines, K.F., and Gochfeld, M.: Ethnic differences in risk from mercury among Savannah River fishermen. *Risk Analysis* 21(3):533-544, 2001.

245. Calderon, R.L., Johnson Jr., C.C., Craun, G.F., Dufour, A.P., Karlin, R.J., and Sinks, T., et al.: Health risks from contaminated water: Do class and race matter? Toxicology and Industrial Health 9(5):879, 1993.

246. Morello-Frosch, R., Pastor, M., and Sadd, J.: Environmental justice and southern California's "riskscape". *Urban Affairs Review* 36(4):551, 2001.

247. Moses, M., Johnson, E.S., Anger, W.K., Burse, V.W., Horstman, S.W., and Jackson R.J., et al.: Environmental equity and pesticide exposure. *Toxicology and Industrial Health* 9(5):913, 1993.

248. Perera, F.P., Rauh, V., Tsai, W.Y., Kinney, P., Camann, D., and Barr, D., et al.: Effects of transplacental exposure to environmental pollutants on birth outcomes in a multiethnic population. *Environmental Health Perspectives* 111(2):201, 2003.

249. Pastor, M., Sadd, J., and Hipp, J.: Which came first? toxic facilities, minority move in, and environmental justice. *Journal of Urban Affairs* 23(1):1-21, 2001.

250. Margi, F.: Environmental Health Hazards and Social Justice: Geographical Perspectives on Race and Class Disparities. Earthscan LLC, 2010.

251. Osiecki, K.M., Kim, S., Chukwudozie, I.B., Calhoun, E.A.: Utilizing exploratory spatial data analysis to examine health and environmental disparities in disadvantaged neighborhoods. *Environmental Justice* 6(3):81–87, 2013.

252. Anselin, L.: Local indicators of spatial association – LISA. *Geographic Analysis* 27(2):93-115, 1995.

253. Anselin, L., Syabri, I., and Youngihn, K.: GeoDa: An introduction to spatial data analysis. *Geographical Analysis* 38(1): 5-22, 2006.

254. Osiecki, K.M., Kim, S., Chukwudozie, I.B., Calhoun, E.A.: Utilizing exploratory spatial data analysis to examine health and environmental disparities in disadvantaged neighborhoods. *Environmental Justice* 6(3):81–87, 2013.

255. Anselin, L.: Local indicators of spatial association – LISA. *Geographic Analysis* 27(2):93-115, 1995.

256. Margi, F.: Environmental Health Hazards and Social Justice: Geographical Perspectives on Race and Class Disparities. Earthscan LLC, 2010.

257. Margi, F.: Environmental Health Hazards and Social Justice: Geographical Perspectives on Race and Class Disparities. Earthscan LLC, 2010.

258. Anselin, L.: Local indicators of spatial association – LISA. *Geographic Analysis* 27(2):93-115, 1995.

259. Arizona State University GeoDa Center for Geospatial Analysis and Computation, Glossary of Key Terms. https://geodacenter.asu.edu/node/390#e. Accessed on June 11, 2013.

260. Arizona State University GeoDa Center for Geospatial Analysis and Computation, Glossary of Key Terms. https://geodacenter.asu.edu/node/390#e. Accessed on June 11, 2013.

261. Arizona State University GeoDa Center for Geospatial Analysis and Computation, Glossary of Key Terms. https://geodacenter.asu.edu/node/390#e. Accessed on June 11, 2013.

262. Arizona State University GeoDa Center for Geospatial Analysis and Computation, Glossary of Key Terms. https://geodacenter.asu.edu/node/390#e. Accessed on June 11, 2013.

263. Arizona State University GeoDa Center for Geospatial Analysis and Computation, Glossary of Key Terms. https://geodacenter.asu.edu/node/390#e. Accessed on June 11, 2013.

264. Arizona State University GeoDa Center for Geospatial Analysis and Computation, Glossary of Key Terms. https://geodacenter.asu.edu/node/390#e. Accessed on June 11, 2013.

265. Arizona State University GeoDa Center for Geospatial Analysis and Computation, Tutorials and Reports. https://geodacenter.asu.edu/learning/tutorials. Accessed on June 4, 2013.

266. Arizona State University GeoDa Center for Geospatial Analysis and Computation, Tutorials and Reports. https://geodacenter.asu.edu/learning/tutorials. Accessed on June 4, 2013.

267. Arizona State University GeoDa Center for Geospatial Analysis and Computation, Tutorials and Reports. https://geodacenter.asu.edu/learning/tutorials. Accessed on June 4, 2013.

268. Anselin, L.: Local indicators of spatial association – LISA. *Geographic Analysis* 27(2):93-115, 1995.

269. Arizona State University GeoDa Center for Geospatial Analysis and Computation, Tutorials and Reports. https://geodacenter.asu.edu/learning/tutorials. Accessed on June 4, 2013.

270. Arizona State University GeoDa Center for Geospatial Analysis and Computation, Tutorials and Reports. https://geodacenter.asu.edu/learning/tutorials. Accessed on June 4, 2013.

271. Arizona State University GeoDa Center for Geospatial Analysis and Computation, Tutorials and Reports. https://geodacenter.asu.edu/learning/tutorials. Accessed on June 4, 2013.

272. Arizona State University GeoDa Center for Geospatial Analysis and Computation, Tutorials and Reports. https://geodacenter.asu.edu/learning/tutorials. Accessed on June 4, 2013.

273. Osiecki, K.M., Kim, S., Chukwudozie, I.B., Calhoun, E.A.: Utilizing exploratory spatial data analysis to examine health and environmental disparities in disadvantaged neighborhoods. *Environmental Justice* 6(3):81–87, 2013.

274. Osiecki, K.M., Kim, S., Chukwudozie, I.B., Calhoun, E.A.: Utilizing exploratory spatial data analysis to examine health and environmental disparities in disadvantaged neighborhoods. *Environmental Justice* 6(3):81–87, 2013.

275. Osiecki, K.M., Kim, S., Chukwudozie, I.B., Calhoun, E.A.: Utilizing exploratory spatial data analysis to examine health and environmental disparities in disadvantaged neighborhoods. *Environmental Justice* 6(3):81–87, 2013.

276. USEPA Toxics Release Inventory (TRI) Program Factsheet http://www.epa.gov/tri/triprogram/RY_2011_TRI_Factsheet.pdf. Accessed on June 6, 2013.

277. Margi, F.: Environmental Health Hazards and Social Justice: Geographical Perspectives on Race and Class Disparities. Earthscan LLC, 2010.

278. USEPA Toxics Release Inventory (TRI) Program Factsheet http://www.epa.gov/tri/triprogram/RY_2011_TRI_Factsheet.pdf. Accessed on June 6, 2013.

VITA

NAME: Kristin Marie Osiecki

EDUCATION: B.S., Health Education, University of Illinois at Urbana, Urbana, Illinois, 1989

M.S., Public Health, University of Illinois at Chicago, Chicago, Illinois 2011

HONORS: Magna Cum Laude, School of Public Health, University of Illinois at Chicago, 2011

Dean’s List, University of Illinois at Urbana, College of Applied Life Studies, 1987 - 1989

PUBLICATIONS: Osiecki, K.M., Kim, S., Chukwudozie, I.B., Barrett, R.E., Calhoun, E.A. (2013). Utilizing exploratory spatial data analysis to examine health and environmental disparities in disadvantaged neighborhoods. *Accepted (May 2013)*; *Environmental Justice*