12-4-2006

Spatial Analysis of Substantiated Child Maltreatment in Metro Atlanta, Georgia

Yueqin Zhou

SPATIAL ANALYSIS OF SUBSTANTIATED CHILD MALTREATMENT IN

METRO ATLANTA, GEORGIA

by

YUEQIN ZHOU

Under the Direction of Jeremy Crampton

ABSTRACT

Identifying high-risk areas for child maltreatment to ultimately aid public health agencies for interventions is necessary for protecting children at high risk. Rates of substantiated neglect and physical/emotional abuse in 2000-2002 are computed for the census tracts in the urban area of five counties in Metro Atlanta, Georgia, and analyzed using spatial regression to determine their relationships with twelve risk variables computed from the Vital Records births and the 2000 Census data.

After accounting for multicollinearity among risk variables and spatial autocorrelation among observations for neighboring locations, it is found that high percentages of (1) births to non-married mothers, (2) births to mothers who smoked or drank alcohol during pregnancy, (3) unemployed males and females, and (4) single-parent families with children under age six best predict the rates of substantiated neglect, and that high percentage of births to mothers who smoked or drank alcohol during pregnancy best predicts the rates of substantiated physical/emotional abuse.

INDEX WORDS: Substantiated child maltreatment, Neglect, Physical abuse, Emotional abuse, Young children, Ecological theory, Spatial autocorrelation, Spatial regression

SPATIAL ANALYSIS OF SUBSTANTIATED CHILD MALTREATMENT IN METRO ATLANTA, GEORGIA

by

YUEQIN ZHOU

Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

In the College of Arts and Sciences

Georgia State University

2006

SPATIAL ANALYSIS OF SUBSTANTIATED CHILD MALTREATMENT IN

METRO ATLANTA, GEORGIA

by

YUEQIN ZHOU

Major Professor:	Jeremy Crampton
Committee:	Jeremy Diem
	Michael Eriksen
	Elaine Hallisey

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2006

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to the many people who have guided and helped me complete the Master's Degree program in the Department of Geosciences at Georgia State University.

First, I would like to thank my committee chair, Dr. Jeremy Crampton, for his guidance through my thesis research and for spending time on organizing my comprehensive exams. I would also like to thank the other members of my thesis committee, Dr. Jeremy Diem, Dr. Michael Eriksen, and Ms. Elaine Hallisey, for their valuable time on my comprehensive exams and thesis research.

I would also like to thank Dr. Susan Walcott, former graduate director in the Department of Anthropology and Geography, for her assistance in helping me obtain the financial support and for her encouragement.

Special thanks go to Ms. Hallisey, who is both a thesis committee member and my work supervisor, for her daily encouragement and great assistance in word-by-word reviewing of and commenting on my thesis writing. Ms. Hallisey also offered hands-on GIS help when I was a graduate research assistant during my first one and half years of graduate study. Foremost, she helped me find employment that best fits to my interests and specialty and helped me handle the difficulties in playing multiple roles: a full time employee, a thesis researcher, and a mother, whose husband works and lives in the west coast.

I pursued my thesis research while working as a full time employee at the Georgia Department of Human Resources. I would like to thank my employer, specifically Mr. Gordon Freymann, Director of Office of Health Information and Policy, for allowing me to use the resources I needed, including the data (DHR Institutional Review Board approval No. 060807), software and hardware, and time. This thesis could not have been completed without supports from my employer.

Lastly, thanks to my husband, Jun, and my son, Hanssen, for their love, encouragement, and patience.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AIC	Akaike Information Criteria
CAPEA	Child Abuse Prevention and Enforcement Act
CAPTA	Child Abuse Prevention and Treatment Act
CPS	Child Protective Services
DCDC	Detailed Case Data Component
DFCS	Division of Family and Children Services
DHHS	Department of Health and Human Services
DHR	Department of Human Resources
ESDA	Exploratory Spatial Data Analysis
GWR	Geographically Weighted Regression
IVE	Instrumental Variables Estimation
LISA	Local Indicator of Spatial Autocorrelation
LM	Lagrange Multiplier
MCN	Multicollinearity Condition Number
MLE	Maximum Likelihood Estimation
MAUP	Modifiable Areal Unit Problem
NACC	National Association of Counsel for Children
NCANDS	National Child Abuse and Neglect Data System
NCCAN	National Center on Child Abuse and Neglect
NCCANI	National Center on Child Abuse and Neglect Information
NIS	National Incidence Study
OLS	Ordinary Least Square

s.e.	Standard Error
SDC	Summary Data Component
WHO	World Health Organization
WIC	Women, Infants and Children
WLS	Weighted Least Squares

CHAPTER ONE

INTRODUCTION

1.1 CHILD MALTREATMENT AND CONSEQUENCES

Child maltreatment is the general term used to describe all types of child abuse and neglect done to a child by his or her primary caregiver. According to the US federal guidelines stated in the Child Abuse Prevention and Treatment Act (CAPTA), there are three types of child abuse: physical abuse, sexual abuse, and emotional abuse. Neglect is failure to provide for a child's basic physical, educational, medical, and emotional needs.

In the United States, the nationwide rate of maltreatment for each year in 1998-2003 was about 12 per 1,000 children aged 0-17 years. The rate of child maltreatment was inversely related to the age of the child: children from 0 to 3 years of age had the highest rate (DHHS, 2005a).

In Georgia, the statewide rate of child maltreatment was higher than the nationwide rate for each year in the years 1998-2003. Also, while the nationwide rate in this time period remained stable, Georgia experienced monotonic increase in maltreatment rate from 12.1 per 1,000 children in 1998 to 19.1 per 1,000 children in 2003 (DHHS, 2006).

The consequences of child maltreatment are striking. The most tragic consequence of maltreatment is child fatality. Infants and very young children have the highest percentage of deaths. An estimated 1,500 children died from abuse or neglect in

2003 – a rate of 2.00 deaths per 100,000 children. Of these 1500 children who died from abuse or neglect, 78.7% children were younger than four years of age (DHHS, 2005a).

The economic consequence of child maltreatment is immense. It is estimated that the nationwide costs resulting from abuse and neglect are as high as $94 billion per year, of which $24.3 billion are used for the immediate needs of abused or neglected children including hospitalization, treatment of chronic health problems, mental health care, child welfare, law enforcement, and the judicial system; and $69.7 billion are spent as the costs associated with the long-term and/or secondary effects of child abuse and neglect (Fromm, 2001). A study assessing the economic burden of hospitalization associated with child abuse and neglect found that children whose hospitalization was due to abuse or neglect were significantly more likely to have longer hospital stays and double the total charges than other hospitalized children and that nearly two-thirds of the primary payer were Medicaid (Rovi, Chen, & Johnson, 2004).

Child maltreatment has pronounced negative medical and social consequences. A large number of studies can be found in medical literature to confirm the association between childhood maltreatment and adverse adult health outcomes (Springer, Sheridan, Kuo, & Carnes, 2003). Examples include smoking (Anda et al., 1999), drug abuse (Dembo et al., 1988), depression (Kessler & Magee, 1994; Bifulco, Bernazzani, Moran, & Ball, 2000), stress disorder (Widom, 1999), and certain chronic diseases (Felitti et al., 1998). For example, a study on the long-term consequences of maltreatment in the early years using the longitudinal data from infancy through late adolescence confirmed adverse impact of early maltreatment on later antisocial behavior (Egeland, Yates, Appleyard, & van Dulmen, 2002).

1.2 PREVIOUS RESEARCH ON CHILD MALTREATMENT

There has been an increasing research interest in this issue since the publication of the seminal paper on child abuse by Kempe and colleagues (Kempe, Silverman, Steele, Droegemueller, & Silver, 1962). A variety of theories has been developed to account for the etiology of child maltreatment (Tzeng, Jackson, & Karlson, 1991). Among them, the ecological framework developed by Garbarino (1977) and by Belsky (1980) has been noted as the best theoretical model, for it considers child maltreatment the product of a variety of factors at multiple levels (Scannapieco & Connell-Carrick, 2005).

In Belsky’s ecological framework, child maltreatment is conceptualized as a social-psychological phenomenon that is multiply determined by forces at work in the individual abuser (ontogenic development) and the family (the microsystem), as well as the community (the exosystem) and the culture (the macrosystem) in which both the individual and the family are embedded (Belsky, 1980). It suggests that the characteristics of the individual (the child and caregiver), the family, the community, and the society all contribute to the increased risk of child maltreatment. The essence of the ecological approach is that it focuses not only on risk factors in individual systems, but also interactions among variables across systems.

The empirical studies employing the ecological framework of child maltreatment found in the literature can be categorized into two types: individual-level studies and area-based ecological studies. In individual-level studies, data are collected from individuals through interviews and analyzed using logistic regression methods. This type of study allows the incorporation of multiple-level factors and their interactions in the etiology of child maltreatment and aids in understanding causal relationships between multidimensional factors and child maltreatment. They are rarely seen in the literature, however, in part because gathering data in various facets at multiple levels has been a big challenge, as Belsky pointed out more than two decades ago (Belsky, 1980).

The area-based ecological approach views child maltreatment as a community problem and studies child maltreatment problems in relation to community level factors in the exosystem, such as economic resources and social supports derived from the decennial census data. County or the census tract is typically chosen as a surrogate for the community. The typical research question is: to which extent the community level factors characterized by socioeconomic and demographic variables are related to rates of child maltreatment. However, studies that address both the microsystem and exosystem factors have not been identified in existing ecological studies of child maltreatment.

1.3 MOTIVATION OF THE PRESENT STUDY

The present study is motivated by the desire to determine: (1) the scope of substantiated child maltreatment in Georgia, specifically, the Metro Atlanta urban area; (2) the variation of child maltreatment rates by community; (3) the variation of maltreatment rates by community in relation to the variation of both the microsystem and exosystem risk factors identified in previous studies; and (4) a set of risk factors that best predicts maltreatment rates.

Although the child maltreatment statistics for Georgia as a whole and for individual counties are reported annually, no in-depth research has been found addressing the problem at a more detailed level of geography, such as the census tract. Identification of community level factors associated with child maltreatment may be used to ultimately aid public health agencies in identifying geographic areas for intervention and prevention.

The theoretical framework in which the present study lies is the ecological perspective on child maltreatment developed by Garbarino and by Belsky. The 2000 census tract is chosen as a surrogate for the community. Data on substantiated maltreatment and risk factors are managed and analyzed at the census tract level. Spatial effects including non-constant variance of observations, non-constant relationships across space, and spatial autocorrelation, which together are well-known characteristics of spatial data, are taken into account in the data analysis. Special attention is focused on spatial autorrelation, meaning that data collected at a location in space tend to be similar to those at nearby locations. The presence of spatial autocorrelation violates the assumption of independent observations for traditional statistical methods. To obtain reliable results, spatial autocorrelation must be taken into consideration in the analysis.

CHAPTER TWO

Literature Review

2.1 CHILD MALTREATMENT: DEFINITIONS, PREVENTION ACTS, AND STATISTICS

2.1.1 Child Maltreatment Definitions

Child maltreatment, or child abuse and neglect, is a widespread social problem in all parts of the world, not only in poor countries, but also in rich nations, including the United States (UNICEF, 2003). It has been recognized that a complex combination of individual, relational, communal, and societal factors contributes to its occurrence.

Due to the differences in perception of what is considered maltreatment in different communities and societies, there has been no universal unifying definition across countries. The World Health Organization (WHO) defines child maltreatment as follows (WHO, 1999, p15):

> "Child abuse or maltreatment constitutes all forms of physical and/or emotional ill-treatment, sexual abuse, neglect or negligent treatment or commercial or other exploitation, resulting in actual or potential harm to the child's health, survival, development or dignity in the context of a relationship of responsibility, trust or power."

In this definition, child maltreatment is subcategorized into five types: physical abuse, emotional abuse, neglect and negligent treatment, sexual abuse, and exploitation.

In the United States, the national definition of child abuse and neglect was first introduced in the Child Abuse Prevention and Treatment Act (CAPTA, Public Law 93-247) enacted in 1974, and included in its amendments. By this definition, child abuse and neglect means (CAPTA, 2004, p44):

> "at a minimum, any recent act or failure to act on the part of a parent or caretaker, which results in death, serious physical or emotional harm, sexual abuse or exploitation, or an act or failure to act which presents an imminent risk of serious harm."

Further, neglect is defined as failure to provide for a child's basic needs. These include physical needs (necessary food or shelter, or appropriate supervision), medical needs (necessary medical or mental health treatment), educational needs (normal or special education), and emotional needs (psychological care, etc.). Abuse is subcategorized into physical abuse, sexual abuse, and psychological/emotional abuse.

Within the minimum standards set by CAPTA, each state is responsible for providing its own definitions of child abuse and neglect. The Child Abuse and Neglect Prevention Act of Georgia provides the following definitions of child abuse and neglect (Georgia General Assembly Ch. 19-14):

> "Child abuse means harm or threatened harm to a child´s health or welfare by a person responsible for the child´s health or welfare, which harm occurs or is threatened through nonaccidental physical or mental injury or the commission of a crime involving physical or sexual abuse of a child."

> "Neglect means harm to a child´s health or welfare by a person responsible for the child´s health or welfare which occurs through negligent treatment, including the failure to provide adequate food, clothing, shelter, or medical care."

Despite the differences among the definitions, the commonality is that child maltreatment consists of four general types: neglect, physical abuse, sexual abuse, and emotional (psychological) abuse. These categorizations are used to guide national and statewide child maltreatment data collection and management.

2.1.2 Child Maltreatment Prevention Acts

Modern child protection movement began the early 1800s (Scannapieco & Connell-Carrick, 2005). In the United States, child protection began with the House of Refuge movement driven by the doctrine of "Parens patriae," i.e., "parent of the country," which represented the first attempt to intervene on behalf of abused and neglected children (Scannapieco & Connell-Carrick, 2005). During the century following that event, a series of actions were taken, including the creation of the United States Children's Bureau as the result of President Roosevelt's 1909 White House Conference on Children in 1912, and the passage of the Shappard-Towner Act, which established Children's Bureaus at the state level in 1921 (NACC, 2005).

It was not until 1962 that child maltreatment was brought to the attention of medical professionals and the general public. Following a medical symposium in the previous year, Dr. Henry Kempe and colleagues published an article titled "The Battered Child Syndrome" in the *Journal of the American Medical Association* (Kempe et al., 1962). It was this landmark article that led to professional and public awareness of the existence and magnitude of child abuse and neglect in the United States and throughout the world (NACC, 2005).

The need for federal intervention led to the Child Abuse and Prevention and Treatment Act (Public Law 93-247) signed into law in 1974. It is one of the key pieces of legislation that guides child protection in the United States. The act sets forth minimum standards of what is considered child abuse and neglect; provides Federal funding to States in support of prevention, assessment, investigation, prosecution, and treatment activities; establishes the National Center on Child Abuse and Neglect (NCCAN), and mandates the National Clearinghouse on Child Abuse and Neglect Information (NCCANI) (NCCANI, 2004). The act was reauthorized in 1978, 1984, 1988, 1992, 1996, and 2003, with each reauthorization, amendments have been made to expand and refine the scope of the law.

In addition to CAPTA, other federal acts on child abuse prevention and welfare protection were also enacted, such as the Child Abuse Prevention and Enforcement Act (CAPEA, Public Law 106-177), which focuses on improving the criminal justice system's ability to provide timely, accurate criminal-record information to agencies engaged in child protection, and enhancing prevention and law enforcement activities (NACC, 2005).

Each state has its own legislation acts on child abuse prevention. In Georgia, one of the key legislation acts is the Child Abuse and Neglect Prevention Act of Georgia, which provides definitions of child abuse and neglect; establishes child abuse and neglect prevention programs, and the State Children’s Trust Fund (Georgia General Assembly Ch. 19-14). Another is the Children and Youth Act (Georgia General Assembly Ch. 49-

5), which authorizes and empowers the Department of Human Resources (DHR), through its Division of Family and Children Services (DFCS) and the county and district departments of family and children services, to establish programs to provide Child Protective Services (CPS) and other services (Section 49-5-8); and to establish and maintain the CPS Information System (Sections 49-5-180 and 49-5-181).

2.1.3 Child Maltreatment Statistical Data

2.1.3.1 Child Maltreatment Data Sources

Child maltreatment statistics come from many different sources, including nationwide systems, statewide systems, and various research agencies. Two key nationwide sources to provide national child maltreatment statistical data are: the National Child Abuse and Neglect Data System (NCANDS), and the National Incidence Study (NIS) of Child Abuse and Neglect. The State of Georgia CPS Information System is the key statewide source in Georgia

2.1.3.1.1 National data collection systems

The NCANDS was established by NCCAN, DHHS, as response to the Amendment of the Child Abuse Prevention and Treatment Act (Public Law 100-294) passed on April 25, 1988, which directed the secretary of the DHHS to establish a national data collection and analysis program on child abuse and neglect (DHHS, 2001). Starting in 1991, the NCANDS annually gathers and analyzes data reported by the states (including the District of Columbia, the territories, and the Armed Services). Key elements are the number of children abused and neglected, the types of abuse, the number of fatalities due to maltreatment, and the types of services provided to address maltreatment and prevent future abuse etc. In 1992, the DHHS produced its first NCANDS report based on data from 1990. Since then, 14 annual child maltreatment reports have been published. The most recent report is *Child Maltreatment 2003*, published in 2005 (DHHS, 2005a).

The NCANDS collects state child abuse and neglect data at different levels of detail through two data components. The Summary Data Component (SDC) collects state-level aggregate data through an annual survey, while the Detailed Case Data Component (DCDC) collects case-level data on children who are subjects of alleged maltreatment reports. An example of the instruments for these two data collection components can be found in (DHHS, 2005b). The SDC data were used as the primary sources for the child maltreatment reports 1990 through 1999. The DCDC data have been used as the primary sources since the publication of *Child Maltreatment 2000* (DHHS, 2005a). For the year 2003 data, all but six states reported DCDC data. Georgia was among the six states reporting SDC data.

Another national key data source is the National Incidence Study of Child Abuse and Neglect. The NIS is a congressionally mandated, periodic effort of the National Center on Child Abuse and Neglect (Sedlak & Broadhurst, 1996). The first NIS (NIS-1), mandated under the CAPTA (Public Law 93-247), was conducted in 1979 and 1980 and published in 1981. The second and third NISs, NIS-2 and NIS-3, were conducted in 1986 and 1987, and in 1993 and 1994, respectively. The work of the fourth NIS (NIS-4) began in April 2004 (Westat, 2004).

The principal purpose of the previous national incidence studies was to go beyond cases of child maltreatment that come to the attention of the official CPS system and attempt to assess the overall national incidence of the problem (Sedlak, 2001). The NISs gather data in a nationally representative sample of counties selected to ensure the necessary mix of geographic regions and of urban and rural areas. The CPS agencies serving the selected counties are asked to provide data about all children in cases they accepted for investigation during a specific period of time. In addition, professionals working in a wide range of agencies in the same counties are asked to serve as the "sentinels" to remain on the lookout for children they believe are maltreated during the study period. Some of the agencies include elementary and secondary public schools; public health departments; public housing authorities; short-stay general and children's hospitals; state, county and municipal police/sheriff departments; licensed day care centers; juvenile probation departments; voluntary social services and mental health agencies; and shelters for runaway and homeless youth shelters or victims of domestic violence. Children identified by sentinels and those whose alleged maltreatment is investigated by CPS during the same period are evaluated against standardized definitions of abuse and neglect.

2.1.3.1.2 State of Georgia data collection system

The State of Georgia CPS Information System, the Protective Services Data System (PSDS), was established in 1990 in response to the requirement of the Children and Youth Act. The PSDS is administered by DFCS and operated by each of the 159 county DFCS offices. When a suspected maltreatment case is reported to a county DFCS office, the county DFCS office determines whether it meets the criteria for a CPS investigation, i.e., the child is under 18 years of age and alleged to be mistreated by the parent or caretaker (DFCS, 2004). A report that meets the criteria is investigated by the CPS agencies. An investigated report is substantiated when the preponderance of evidence supports the allegation. It is unsubstantiated when a preponderance of evidence does not exist or there is no evidence to support the allegation. The county DFCS office enters the alleged abuse/neglect report to the PSDS.

The PSDS collects all relevant information of all cases investigated by CPS for alleged maltreatment of children, including the demographics of the child, address of residence, maltreater's demographics, types of alleged and substantiated maltreatment, and the consequences of maltreatment (physical injury and/or death). Types of maltreatment include neglect, physical abuse, sexual abuse, emotional abuse, and other abuse. The data collected by the PSDS are used to generate the Georgia PSDS annual reports and to report to the NCANDS.

2.1.3.2 Child Maltreatment Statistics

Despite various prevention legislation acts and programs, child maltreatment is still prevalent everywhere in the country, in all segments of population, regardless of individual differences in cultural background, geographic locations, or socioeconomic status, although the extent of prevalence may differ in different groups. The most recent child maltreatment report based on the 2003 NCANDS data provides the following statistics (DHHS, 2005a):

- Approximately 906,000 children nationwide were determined to be maltreated, of which, more than 60% were neglected; approximately 20% were physically abused; 10% were sexually abused; 17% suffered from other types of maltreatment; and 5% were emotionally maltreated. A child could be a victim of more than one type of maltreatment.
- The nationwide rate of child maltreatment for all children was 12.4 per 1,000 children. Overall, the maltreatment rate was inversely related to the age of the

child. The youngest children in the age group 0-3 years had the highest rate, and the oldest children in the age group 16-17 years had the lowest rate (Table 2.1).

- An estimated 1,500 children were confirmed to have died from maltreatment. Younger children had higher percentage of deaths. Of the deaths, 79% children were younger that 4 years old; 10% were 4-7 years old; 5% were 8-11 years old; and 6% were 12-17 years old.
- The most common single type of maltreatment leading to deaths was neglect followed by physical abuse. Of the deaths in 2003, 35.6% resulted from neglect only, 28.9% from multiple maltreatment types, 28.4% from physical abuse only, 6.7% from emotional (psychological) abuse only, other type, or unknown type only, and 0.4% from sexual abuse only.
- More than 80% of victims were maltreated by at least one parent including biological parent, step-parent or adoptive parent. Children maltreated by nonparental caregivers accounted for 13.4% of the total.

Child maltreatment rates are not homogeneous temporally and geographically. The NCANDS data from 1990 to 2003 (DHHS, 2005a) show temporal and geographical variations. Temporally, the nationwide maltreatment rate experienced a monotonic increase during 1990-1993 to the highest rate of 15.3 per 1,000 children in 1993; a monotonic decrease during 1994-1999 to the lowest rate of 11.8 in 1999; a slight increase in 2000, and has remained stable since 2000 (Figure 2.1). Georgia and Massachusetts experienced an apparent monotonic increase since 1998.

Geographically, maltreatment rates vary from state to state (Figure 2.2). Table 2.2 lists the rates of some states in 1998-2003. Also listed in the table are the US overall rates. New Hampshire, New Jersey, and Pennsylvania were the only states that had the lowest rate (less than 5 per 1,000 children) in all six years. During the same period, Florida and Massachusetts were the only states that had the highest rate (greater than 20 per 1,000 children) in five out of six years. Georgia had the rates in all six years lower than 20 per 1,000 children, but increasing monotonically.

2.2 ECOLOGICAL THEORIES OF CHILD MALTREATMENT

Recognition of the seriousness of child maltreatment in the early 1960s not only propelled federal and state legislation on child maltreatment prevention and welfare protection, but also stimulated broad theoretical and empirical studies of the etiology of child maltreatment. In the past four decades, research on child maltreatment passed through four progressive stages: the “speculations” of the 1960s, the “introspective explorations” of the 1970s, the various “diversities” of the 1980s, and the “multidisciplinary integration” of the 1990s (Tzeng et al., 1991; Scannapieco & Connell-Carrick, 2005). According to (Tzeng et al., 1991), more than 40 theories, models, and/or perspectives had been proposed to address one or more types of child maltreatment. The authors grouped these theories into nine paradigms (or schools of thought), including individual determinants, offender typology, family systems, sociocultural, individual-environmental interaction, parent-child interaction, sociobiological, Learning/situational, and ecological (Appendix A). The ecological theory, which is used to guide the present study, belongs to the ecological paradigm originated from the ecological approach of human development.

Building on the previous fragmented ecologically oriented research work, Bronfenbrenner (1974; 1977) proposed the ecological approach to research in human development, which studies the progressive, mutual accommodation, throughout the life span, between a growing human organism and the changing immediate settings in which the organism lives, as well as the larger social contexts, both formal and informal, in which the immediate settings are embedded. The ecological environment of human development consists of a topologically nested arrangement of the microsystem, mesosystem, exosystem, and macrosystem, each contained within the next (Figure 2.3). The essence of the ecological approach is that it focuses on not only the forces of the individual systems, but also interactions among systems.

About the same time Bronfenbrenner proposed the ecological approach in human development research, Garbarino practiced ecologically oriented research on child maltreatment. Following an empirical research (Garbarino, 1976), Garbarino (1977) conceptualized the ecological model of child maltreatment. In Garbarino's ecological framework, child maltreatment was placed in the perspective of family development, and considered the product of a multiplicity of factors, which were categorized into sufficient conditions, and necessary conditions. The sufficient conditions refer to family asynchrony, i.e., mismatch of parent to child and of family to neighborhood and community. These conditions would lead to child maltreatment if the necessary conditions were met, which include the role of cultural support for the use of physical force against children, and lack of or failure to use family support systems.

Before long, Belsky applied Bronfenbrenner's ecological approach to integrate divergent viewpoints of child maltreatment, particularly the psychiatric model emphasizing the role of the individual abuser; sociological model highlighting the role of social factors in abuse; the child abuse-eliciting characteristics model pointing toward the role the child plays in stimulating his or her own maltreatment; and the ecological model proposed by Garbarino. Belsky (1980) conceptualized child maltreatment as a social-psychological phenomenon that is multiply determined by forces at work in the individual abuser (ontogenic development) and the family (the microsystem), as well as the community (the exosystem) and the culture (the macrosystem) in which both the individual and the family are embedded.

Belsky's ecological framework consists of four levels of analysis: ontogenetic development, the microsystem, the exosystem, and the macrosystem. Table 2.3 summarizes the social units of analysis and causal factors at each level. This ecological framework recognizes the roles played not only by individual factors in the etiology of child maltreatment, but also by their interactions. The co-existence of causative factors at different levels increases the likelihood of child maltreatment. For example, the likelihood that a child would be abused by his/her mother would be increased if the child was born prematurely to the mother who herself was victim of child abuse; the likelihood would be further increased if the family was struggling with economic resources or marital conflict; and it would be greatly increased if the disorganized family was embedded in a community where social support systems were lacking, and violence toward children was perceived as the normal disciplinary means.

However, it has been recognized that the ecological theories alone do not fully explain why maltreatment rates vary across areas with similar risks, and there exist some factors that play roles in decreasing the probability of maltreatment. Developed upon the ecological theories are the ecological/transactional theory (Cicchetti & Lynch, 1993) and the ecological/developmental theory (Belsky, 1993). Both recognize not only the risk (potentiating) factors or stressors that increase the probability of child maltreatment, but also the protective factors or supports that decrease the risk for maltreatment. According to these theories, child maltreatment occurs only when potentiating factors outweigh protective factors (Cicchetti & Lynch, 1993), or stressors outweigh supports (Belsky, 1993). This implies that in order to prevent child maltreatment more effectively, prevention programs should not only focus on reducing risks but also on strengthening protective factors (Tomison & Wise, 1999).

2.3 APPLICATIONS OF THE ECOLOGICAL THEORIES

Empirical studies employing the ecological theories of child maltreatment fall into two general categories: individual-level studies and area-based ecological studies, the latter of which are the focus of this review.

In the individual-level studies, data are collected from individuals through interviews, and analyzed using the logistic regression methods. Examples of this type of study are found in (Kotch et al., 1995; Kotch et al., 1997; Kotch, Browne, Dufort, Winsor, & Catellier, 1999). In their studies, Kotch et al. (1995) recruited 1,111 mothers of newborn infants from community and regional hospitals and local health departments in 42 counties of North Carolina and South Carolina selected for geographic distribution, 80% of whom had biomedical and sociodemographic risk factors. They interviewed 842 mothers shortly after discharges from hospitals. Questions were asked regarding factors in ontogeneic development, micro-, exo-, and macro-systems. These include the mother’s history of child maltreatment (mother’s separation as a young adolescent from her own mother), characteristics of the mother (depression, health opinion, self-esteem, education, and maternal health), the infant (infant health, and characteristics), the family (marital status, number of children, stress, income, employment, receipt of cash and in-kind public support such as Aid to Families with Dependent Children, Medicaid, Food Stamps, and Women, Infants and Children (WIC)), social networks/social support, and cultural beliefs. The data about these mother-infant pairs were linked to the child abuse and neglect registry data; and processed to create Boolean fields, 1 indicating occurrence of maltreatment, if report(s) of child maltreatment occurred to an infant before the first birthday, and 0, otherwise. Among the interviewed, 749 mother-infant pairs met the predefined criteria and were included in the data analysis. Logistic regression was used to identify risk factors for reported infant abuse or neglect. The study found that receipt of Medicaid, low maternal education, the presence of any other dependent children in the home, maternal depression, and mother's separation as a young adolescent from her own mother were predictive risk factors of reports of child maltreatment during the first year of life.

The researchers extended their follow-up period up to 4 years after the discharges, and identified neonatal risk factors that were predictive of reports of child maltreatment in the second and third years of life (Kotch et al., 1997), and those in the first four years of life (Kotch et al., 1999).

Individual-level studies such as the above examples allow incorporating multi-level factors and their interactions in the etiology of child maltreatment and help understand causal relationships between multidimensional factors and child maltreatment. However, they are rarely seen in the literature, in part because gathering data in various facets at multiple levels has been a big challenge, as Belsky (1980) pointed out in more than two decades ago.

The second category of child maltreatment studies under the ecological framework is the area-based ecological approach, in which child maltreatment is viewed as a community problem and studied in relation to community (ecological) level factors, such as economic resources and social supports. The goal is to determine the extent to which the community level factors characterized by socioeconomic and demographic variables are related to rates of child maltreatment. An administrative unit such as county or the census tract is used as a surrogate for the community. In contrast to individual-level studies where data are collected from interviews, ecological studies obtain child maltreatment data from the state or local official child maltreatment data collection systems, and obtain community level factors from the U.S. Census database.

In the first ecological study of child maltreatment, Garbarino (1976) examined 12 socioeconomic and demographic indices (Table 2.4) reflecting five dimensions of community economic resources and social supports to determine the extent to which they were associated with child maltreatment reports in New York counties. It was found that five variables (displayed in Italic in Table 2.4) accounted for 36% of the variance in rates of child maltreatment (rates were calculated as the number of child abuse and neglect combined reports per 1,000 population). Among these five variables, two characterize the extent to which mothers' experience of the stress induced by economic disadvantage and double responsibilities (working outside the home and taking care of children), and three characterize the belief in the value of education as well as the existence of institutionalized opportunities for education.

In a later study to test the hypothesis that child maltreatment is an indicator of the overall quality of life for children and families, Garbarino and Crouter (1978) applied the ecological approach to identify socioeconomic, demographic, attitudinal, and housing correlates of rates of reported abuse and of neglect for 93 census tracts in Douglas County, Nebraska. Among 12 variables examined, five were obtained from survey data of 1,992 respondents in 93 census tracts; seven from the 1970 U.S. Census data. The census variables were 1) percent of families with income less than $8,000 a year; 2) percent of families with income more than $15,000 a year; 3) percent of families headed by females; 4) percent of married women (with children under six years old) in the work force outside the home; 5) percent of families living in current residence less than one year; 6) percent of single-family housing; and 7) percent of vacant housing. The housing variables reflect the physical and social quality of the surroundings. Multivariate regression analysis was used to examine the relationships between rates of reported child maltreatment and individual or a combination of variables. It was found that the rates of reported child maltreatment had negative relationship with the variable "percent of families with income more than $15,000 a year", and all eight different combinations of variables accounted for substantial proportion of the variances in the rates. But the study did not report the statistical significance for individual variables included in each model.

In the last two decades, many researchers addressing child maltreatment problems reexamined the socioeconomic, demographic, and housing variables initiated in the above studies, with some modifications, to determine the extent to which these variables were associated with increased risk of child maltreatment in their selected population. Examples are (Young & Gately, 1988; Zuravin, 1989; Coulton, Korbin, Su, & Chow, 1995; Krishnan & Morrison, 1995; Drake & Pandey, 1996; Ernst, 2000; Weissman, Jogerst, & Dawson, 2003; Freisthler, 2004) (Appendix B), among others. These studies combine all age children together to a single age group, but differ in the unit of analysis, definition of maltreatment rates, and to some extent the variables examined.

2.4 DEALING WITH SPATIAL EFFECTS

Not only do the ecological theories put great demands on data gathering, but they also put great demands on the data analysis techniques to handle multiple variables. Without exception, the above-referenced ecological studies used multivariate regression techniques to determine relationships between rates and a set of variables. Some also used bivariate correlation techniques to determine the correlation between a single variable and the child maltreatment rate (e.g., (Weissman et al., 2003)). One problem with the multivariate regression techniques is multicollinearity among explanatory variables. This problem has been widely noted elsewhere and may be ameliorated by including only uncorrelated variables, or by transforming variables to their orthogonal components using principal component analysis, e.g., in (Coulton et al., 1995). Another problem is the existence of spatial effects embedded in spatially aggregated data. This problem has not attracted much attention.

As noticed in the previous section, data used in the ecological studies of child maltreatment were observations aggregated by spatial units (county, census tract, or census block group). Spatially aggregated data are characterized by spatial dependence (i.e., spatial autocorrelation among observations) and spatial heterogeneity (non-constant variance of observations and non-constant relationships across space), which together are referred to as spatial effects (Anselin, 1988).

Spatial effects were ignored in all the ecological studies cited above except the study by Freisthler (2004). Data were treated as independent observations with constant variances, and analyzed using standard methods of regression (i.e., ordinary least squares (OLS) regression) or correlation (such as Pearson's product-moment correlation). However, when spatial effects exist in the underlying data generating process but ignored in analysis, such as the OLS regression or Pearson's correlation, the results are biased (Anselin & Griffith, 1988). In other words, a significant relationship between the response variable and an explanatory variable suggested by the OLS regression analysis may actually be not significant, and the goodness-of-fit measure (R^2) is upward biased (Benirschka & Binkley, 1994).

A common method to handle spatial autocorrelation is to minimize spatial autocorrelation effects by resampling data to create a subset of data by either manually selecting data locations or using a random process (Mitchell, 2005). However, both methods have some drawbacks. Manual selection may be subjective, random selection may not be free of spatial dependency, and both methods may result in loss of information, that is, the selected subset may not represent all of the characteristics of the dataset.

A less commonly used but more objective method is to separate the spatial component from the non-spatial component of each explanatory variable using the so-called spatial filtering process (Getis, 1990). The spatial and nonspatial components are both considered independent variables in the regression analysis.

A third method is so-called spatial regression which considers spatial autocorrelation an additional variable in the regression equation and solves its effect simultaneously with the effects of other explanatory variables (Anselin, 1988). This method uses all available information in the dataset, and is implemented in the free software GeoDA developed by Luc Anselin and colleagues (Anselin, 2003, 2004, 2005).

Table 2.1 Child maltreatment rate per 1,000 children in the United States, 2003

Age in Years	0-3	4-7	8-11	12-15	16-17
Rate	16.4	13.8	11.7	10.7	5.9

Table 2.2 Child maltreatment rate per 1,000 children in 1998-2003

States	1998	1999	2000	2001	2002	2003
Pennsylvania	1.9	1.8	1.7	1.6	1.8	1.6
New Hampshire	3.9	3.0	2.7	3.5	3.1	3.4
New Jersey	4.9	4.6	4.2	4.1	3.8	3.8
Georgia	12.1	13.1	14.2	16.6	18.2	19.1
Massachusetts	18.9	20.2	21.6	22.1	22.8	24.6
Florida	23.2	18.9	26.3	33.3	31.5	35.3
US Overall	12.9	11.8	12.2	12.4	12.3	12.4

Table 2.3 Belsky's ecological framework on child maltreatment

Level of Analysis	Domain	Potential Causal Factors
Ontogenic development: Abuser	• Developmental history	❖ Exposure to, or experience with, violence as a child ❖ Lack of practice in parenting role
Microsystem: Family	• Child • Parent • Parent-child • Family	❖ Low birth weight, premature birth, temperament, colicky ❖ Young/unmarried, marital conflict and discord, unprepared transition from husband-wife dyad to mother-father-infant triad ❖ Negative parent-child interaction ❖ Large family size, economic stress
Exosystem: Community	• World of work • Neighborhood	❖ Unemployment, occupational stress ❖ Isolation from formal and informal social support systems (either lack of support systems or failure to use support systems)
Macrosystem: Society	• Societal attitudes • Cultural beliefs	❖ Societal willingness to tolerate high levels of violence ❖ General acceptance of physical punishment as a means of controlling children's behavior ❖ Cultural beliefs that children are property to be handled as parents choose

Table 2.4 Ecological correlates examined in Garbarino's study of child maltreatment

Dimension	**Indication**	**Variables**
Transience	• The degree to which families are "rootless"	1) Percentage of the population born in a different state 2) Percentage of families in the same house as 5 years ago
Economic development	• The degree to which families are deprived of necessary material resources and thus experience economic stress	3) Percentage unemployed 4) Percentage of families with income less than 125% of the poverty 5) Median income of all families
Educational development	• To some degree the belief in education as well as the existence of institutionalized opportunities for education	*6) Percentage of adults who are high school graduate* [a] *7) Percentage of 18-19-year olds who are enrolled in educational institutions* [a] *8) Percentage of 3-4-year olds who are enrolled in educational programs* [a]
Rural-urban	• The residential organization, the concentration of resources, and the isolation of families	9) Percentage urban 10) Population density
Socioeconomic situation of mothers	• The extent to which mothers experience the stress induced by economic disadvantage and double responsibilities (working outside the home and taking care of children)	*11) Percentage of women in the labor force who have children under 18 years of age* [a] 12) *Median income of households headed by females* [a]

[a] Variables identified as the predictors of child maltreatment

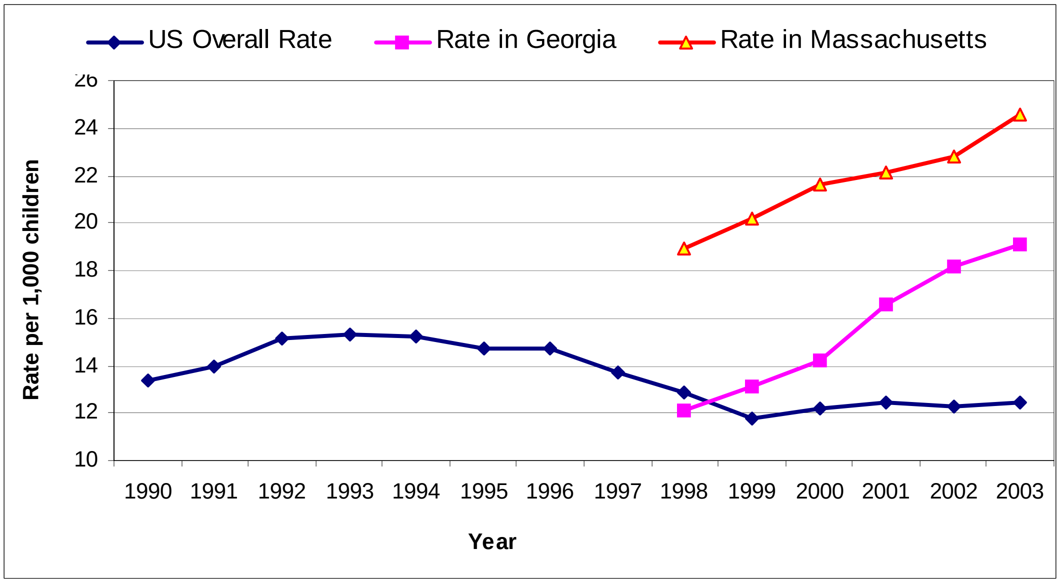


Figure 2.1 Child maltreatment rates in the United States, and in selected states.

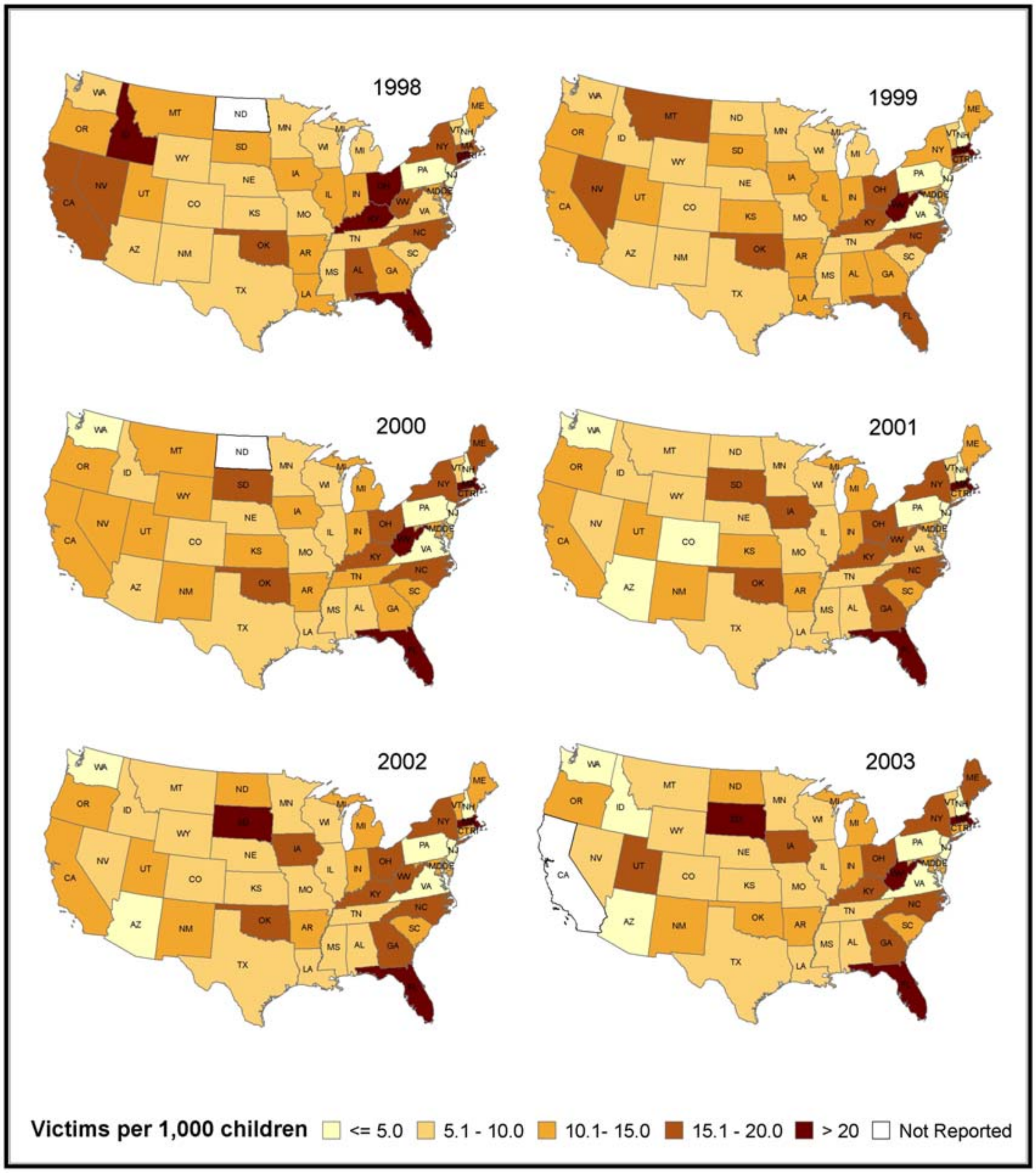
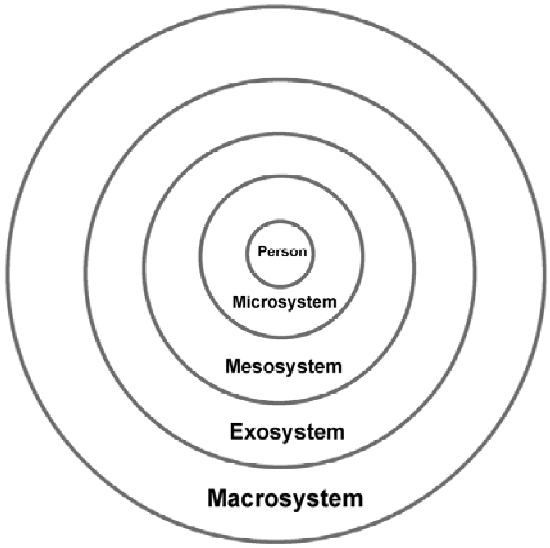


Figure 2.2 Child maltreatment rates in the continental United States, 1998-2003. The break values for classification are based on *Child Maltreatment* 2003.



Microsystem is the complex of relations between the developing person and environment in an immediate setting containing that person.

Mesosystem comprises the interrelations among major settings containing the developing person at a particular point in his or her life.

Exosystem is an extension of the mesosystem embracing other specific social structures, both formal and informal, which include the major institutions of the society.

Macrosystem refers to the overarching institutional patterns of the culture and the subculture.

Figure 2.3 Structure of Bronfenbrenner's human development ecology

CHAPTER THREE

Purpose of the Study and Study Design

3.1 PURPOSE OF THE STUDY

The theoretical framework guiding the present study is the ecological theory of child maltreatment developed by Garbarino and by Belsky. The study has two purposes. The primary purpose is to examine the selected factors in the microsystem as well as in the exosystem to determine if the individual factors are positively related to increased risk of child maltreatment among children under the age of four, and to identify a combination of independent factors that best predicts maltreatment rates. The secondary purpose is to examine how spatial autocorrelation affects the parameter estimates of regression models. This study is an extension of another study whose purpose is to determine if a set of selected perinatal risk factors, both individually and in various combinations, is associated with increased risk of infant maltreatment (Zhou, Hallisey, & Freymann, 2006).

Infancy and early childhood are the years in which the human brain develops most rapidly; maltreatment during this period can seriously disrupt the course of healthy development, leading to physical, mental, emotional, social, and cognitive problems (Scannapieco & Connell-Carrick, 2005). Also, children under the age of four are most vulnerable to serious injuries and deaths from maltreatment (DHHS, 2001, 2003, 2005a).

Identifying high-risk areas to ultimately aid public health agencies for interventions is necessary for protecting children at high risk.

Spatial autocorrelation must be taken into account in the analysis if the presence of spatial autocorrelation does upward bias the absolute values of the test statistic for testing significance of parameter estimates of regression models. Ignoring its effects using traditional statistical methods with nonspatial data may lead to false significant relationships.

3.2 STUDY DESIGN

3.2.1 Ecological Approach

In the present study, an area-based ecological approach is used to examine child maltreatment in relation to characteristics at the level of communities in which the maltreatment victims lived. The 2000 census tract is chosen as a surrogate for the community. The reasons for choosing census tracts are as follows. First, census tracts are small, relatively permanent statistical subdivisions of a county, and designed to be relatively homogeneous units with respect to population characteristics, economic status, and living conditions at the time they are established (U.S. Census Bureau, 2001). They generally contain between 1,000 and 8,000 people, with an optimum size of 4,000 people. Second, the census tract yields maximal geocoding. The geocoding procedures used to geocode data for this study ensure maximal geocoding with reliable results at the census tract level (see Chapter Four for more details). Third, the census tract can be considered the most disaggregated areal unit to allow reliable calculations of rates. Although the census block group and census block are more disaggregated units, they are usually too small to contain sufficient numbers of cases and populations to allow computing reliable health measures including rates. Last, the census tract is readily interpretable to and can feasibly be used by public health staff for intervention purposes. The census tract has been found most apt for monitoring socioeconomic inequalities in health (Krieger, Waterman, Chen, Rehkopf, & Subramanian, 2004).

3.2.2 Variables to Be Examined

The response variable for this study is the rate of substantiated child abuse and neglect by their biological parents among children under four years old for the years 2000-2002. Rates of abuse and rates of neglect are calculated and analyzed separately. Abuse is referred to as a combination of physical and emotional/psychological abuse because they may be associated with similar factors (Tzeng et al., 1991). Sexual abuse is excluded from the study due to its low incidence rate, particularly among children under four years old. Also, community level factors examined in previous studies are more likely related to physical abuse, emotional/psychological abuse, and neglect rather than to sexual abuse (Belsky, 1993; Tomison & Wise, 1999). Explanatory variables to be examined are based on previous research. Scannapieco and Connell-Carrick (2005) provide a list of risk factors associated with child maltreatment among children 0-36 months of age. Some of the factors in this list are chosen to examine based on data availability (Table 3.1). Actual variables related to each factor are listed in Column V of Table 3.1.

The rationale for choosing these variables is as follows. A recent study identified a set of perinatal risk factors for infant maltreatment, which include: mother smoked during pregnancy, more than two siblings, Medicaid beneficiary, unmarried marital status, birth weight less than 2,500 grams, and maternal age less than 20, among others (Wu et al., 2004). Other studies linked premature birth and low Apgar score to infant maltreatment (Frodi et al., 1978; Bugental & Happaney, 2004). In the present study, a composite risk, denoted as CHILDRISK, is used to represent the presence of one or more neonatal difficulties: low birth weight, premature birth, or a low Apgar score. This variable was found to be significantly associated with high rates of infant maltreatment at the census tract level (Zhou et al., 2006). Kotch et al. (1999) found mothers who consumed alcohol during pregnancy to be predictive of child maltreatment in the first four years of life. Based on findings in (Kotch et al., 1999; Wu et al., 2004), a single variable, denoted as SUBSTANCE, that represents the percentage mothers who smoked or consumed alcohol during pregnancy is used as a surrogate for substance abuse. Large family size with four or more children for whom to care induces stress in family environment (Belsky, 1993). The variable Medicaid beneficiary as a surrogate for poverty status is used because Medicaid is a program that pays for medical assistance for certain individuals and families with low income and resources. Five variables related to lack of social support and economic resources stem from two previous studies (Garbarino & Crouter, 1978; Zuravin, 1989).

In Table 3.1, the factor "Unemployment" is not included in Scannapieco and Connell-Carrick's list, but is included in the present study. Unemployment was considered an important factor in the exosystem in Belsky's ecological framework. Belsky argued (Belsky, 1980, p327), "… maltreatment may simply be a consequence of the increased parent-child contact (and thus conflict) that results from the unemployed parent's spending more time at home." Young & Gately (1988) found unemployment was correlated with child abuse by male maltreaters.

3.2.3 Analysis Methods and Research Questions

The hypothesis for this study is that rates of substantiated neglect and of substantiated physical/emotional abuse are positively related to the risk variables defined in Table 3.1. The hypothesis is first examined through visual analyses including reviewing maps and investigating scatter plots, and then tested using quantitative methods of regression analysis to answer the following questions:

1. To what extent are the selected variables individually related to increased risk of child neglect, and physical/emotional abuse?
2. Do the relationships differ by type of maltreatment?
3. What is the set of risk variables that best predicts the rates of child neglect and what is the set of risk variables that best predicts the rates of physical/emotional abuse?
4. Do the risk variables that best predict the rates of child maltreatment differ by type of maltreatment?

Bivariate linear regression techniques are used to determine if the relationships between rates of substantiated neglect and of physical/emotional abuse and individual variables are statistically significant. Multivariate linear regression techniques are used to identify the combinations of variables that best predicts the rates of substantiated neglect and of substantiated physical/emotional abuse.

Spatial effects including spatial heterogeneity (i.e., non-constant variance of observations and non-constant relationships across space) and spatial autocorrelation are controlled. To control for the effect of non-constant relationships, the study area is confined to relatively small area and urban settings because the same risk factors may have different influences on child maltreatment in urban areas vs. rural areas (Belsky, 1993; Weissman et al., 2003). To control for the effect of non-constant variance of observations, the response variables are transformed using a variance-stabilizing function (Waller & Gotway, 2004). Spatial autocorrelation is controlled by the use of the spatial regression method (Anselin, 1988; Anselin & Bera, 1998). Results from OLS regression and those from spatial regression are compared to examine how the presence of spatial autocorrelation affects the parameter estimates of regression models. The reasons for using the spatial regression method rather than resampling data to minimize spatial autocorrelation or using the spatial filtering method are that the spatial regression method uses all available information in the dataset and that software having this function is ready for use.

Table 3.1 Ecological variables to be examined in the study

Column I **Ecological Level**	II **Unit of Analysis**	III **Factors**	IV **Examined?**	V **Variables**
Ontogenic development		• Parent experienced child maltreatment as a child	• No	
Microsystem	Child	• Age • Born prematurely • Physical or mental disability • Infant tests positive for AOD • Race	• Yes • Yes • Yes • No • No	• Controlled (under 4 years old) • **CHILDRISK**: % births experiencing neonatal difficulties (low birth weight, premature birth, or a low Apgar score)
	Parent	• Not satisfied with the child • Biological or genetic factors • Not enjoying parenting • Young parent • Not understanding role as caregiver • Lacking knowledge of child development • Substance abuse	• No • No • No • Yes • No • No • Yes	• **MAGELT20**: % births to mothers less than 20 years old • **SUBSTANCE**: % births to mothers who smoked or drank alcohol during pregnancy
	Family	• Poverty • Stress in family environment • Interpersonal conflict between parents • Single parenting	• Yes • Yes • No • Yes	• **MEDICAID**: % births to Medicaid beneficiaries • **SIBLINGS3**: % births having three or more siblings • **NMARRIED**: % births to non-married mothers
Exosystem	Community	• Lack of social support • Unemployment indicating socioeconomic resource drain	• Yes • Yes	• **SINGPARCH6**: % single parent families with children under 6 years old • **FEMLBCH6**: % females 16 and older (with children under 6 years old) in the labor force outside the home • **RESIDLT1Y**: % families in the current residence < 1 year • **SINGFAMHSE**: % single-family housing units • **VACANTHSE**: % vacant housing units • **UNEMPMF**: % of males and females 16 years and older in the labor force who are unemployed
Macrosystem	Society	• Cultural values that support violence • Attitudes toward how a mother should behave as a parent	• No • No	

CHAPTER FOUR

Study Area and Data Description

4.1 STUDY AREA

The geographic context of the present study is the urban area covering five core counties, including Fulton, DeKalb, Cobb, Gwinnett, and Clayton, which make up much of metropolitan Atlanta, Georgia (Figure 4.1). The City of Atlanta is located in the middle of the study area. The area is divided into 478 census tracts for the 2000 census. The census data show that all census tracts were populated, but population densities varied greatly, from 3 to 36,503 residents per square mile, with an average of 3,138 residents per square mile. The median population density was 2,694 residents per square mile. Figure 4.2 displays the frequency distribution (4.2 (a)) and geographic distribution (4.2 (b)) of population density. The histogram of population density is positively skewed with one extremely densely populated census tract of 36,503 residents per square mile. Most densely populated census tracts were clustered inside the Perimeter (I-285).

In the study area, there were 214,915 children under the age of five in 2000. All census tracts except one (corresponding to the least populated census tract) were occupied by children in this age group. Figure 4.2 displays the frequency distribution (4.2 (c)) and geographic distribution (4.2 (d)) of the percentage of young children.

In contrast to population density, the histogram of the percentage of young children is approximately normally distributed; census tracts with the percentage of young children in the highest two categories were mainly located outside the Perimeter, particularly Gwinnett and Clayton counties.

4.2 DATA DESCRIPTION

4.2.1 The Response Variables

The response variables are rates of substantiated neglect and of substantiated physical/emotional abuse for children under the age of four for the years 2000-2002. Rates are presented as the number of maltreated children per 1,000 children per year. They are calculated as the ratios of the counts of children who were maltreated during the three-year period and the counts of children during the same period amplified by 1,000. The former are derived from the data on substantiated child abuse and neglect, and the latter from the vital records birth data.

4.2.1.1 Data on Substantiated Child Abuse and Neglect

Data on substantiated child abuse and neglect are obtained from DFCS, Georgia DHR. The data were collected, via the State of Georgia CPS Information System, i.e., the PSDS system, from 2000 through 2002 between January 1st and December 31st of each year. In the years 2000, 2001, and 2002, respectively, 1,343, 1,711, and 1,908 children under four years old in the study area were determined to be victims of one or more types of maltreatment including neglect, physical abuse, emotional abuse, sexual abuse and other types of abuse. Of the maltreaters, 91.33% are biological parents; other maltreaters account for 8.67% (Table 4.1).

Associated with each child record is the address information of place of residence including street address, city, county, state, and zip code. Addresses are geocoded using Centrus software (Group 1 Software, 2003). About 88.7% of total records are address-matched to a street address location, and assigned latitudes and longitudes. About 3% of total records are matched to the accuracy of the census tract level. Their latitudes and longitudes are assigned based on five or nine digit zip code centroids. The remaining 8.3% of the records have inappropriate address information, such as P.O. Boxes, or incorrect and/or incomplete addresses. For each of these records, latitude and longitude are randomly assigned, using the Spatial Imputation Method, within the census tract where the child had the highest probability to live based on the child's age-sex-race-specific information (Millard & Freymann, 2001).

Children who were neglected or physically/emotionally abused by their biological parents are included in the analysis. Of total 4962 maltreated children, 3,793 children meet this criterion, of which, 3,526 children were neglected and 313 were physically and/or emotionally abused. 46 children were both neglected and physically/emotionally abused, and are counted in both calculating the number of neglect victims and the number of physical/emotional abuse victims. The individual records are then aggregated by census tract to determine the number of neglect and the number of physical/emotional abuse in each tract.

4.2.1.2 Vital Records Births

Vital records birth data are obtained from the Division of Public Health (DPH), Georgia DHR. The birth records for 1996 through 2002 are extracted from the database. The births in 1996 through 2000, 1997 through 2001, and 1998 through 2002 are used to derive the counts of children under four years old for the year 2000, 2001, and 2002, respectively. The total counts of children during the three-year period are the sum of counts in individual years. It should be noted that the counts of child population derived from birth records are an approximation of actual counts because the calculation does not account for population migration effects.

To adjust for the effects of varying lengths of time that individual children are considered children under four years old in a given target calendar year, the person-year concept is applied (Simpson, Imrey, Geling, & Butkus, 2000; Timmreck, 2002). In doing so, a weighting factor is calculated for each child for each target year, each denoting the proportion of time over a one-year period that a child was under four years old. For example, if a child was under four years old in entire year in 2000, then the weighting factor for this child for the year 2000 is one. Otherwise, the weighting factor is calculated as the number of days during which a child was under four years old divided by total number of days in that year, i.e., 366 days for a leap year, and 365 days otherwise.

As an example, suppose a child was born on 7/11/1997. The weights for 2000, 2001, 2002, are calculated as follows: *Weight2000* = 366 days /366 days = 1; *Weight2001* = 191/365 = 0.5233; and *Weight2002* = 0. For a child born on 7/11/2001, the weights for 2000, 2001, 2002, are calculated as follows: *Weight2000* = 0; *Weight2001* = 173/365 = 0.4740; and *Weight2002* = 1. The number of weighted counts of children in census tract *i* for the year 2000 is calculated as $WN_i\ (2000) = \sum_{j=1}^{N_i} Weight2000(j)$, here N_i is the total number of births residing in tract *i*. $WN_i(2001)$ and $WN_i(2002)$ are calculated in a similar manner. The total number of weighted counts during a three-year period in tract *i* is $WN_i = \sum_k WN_i(k), k$ = 2000, 2001, and 2002.

4.2.1.3 Calculating the Response Variables

Let RATENEG and RATEPE denote the rate of substantiated neglect and rate of substantiated physical/emotional abuse, respectively. The calculation of rates is as follows.

Let Y_i be the number of victims of substantiated neglect or physical/emotional abuse in tract i, and WN_i the weighted counts of children under four years old in the same tract. The rate is calculated as:

$$R_i = 1000 * Y_i / WN_i \tag{4.1}$$

Figure 4.3 displays the histogram of the substantiated neglect rates (4.3 (a)) and that of physical/emotional abuse rates (4.3 (b)). Both histograms are highly positively skewed.

To ensure the normal distribution, which is required for linear regression, the rates are transformed to their natural logarithmic form. Let TRATENEG and TRATEPE denote the transformed rate of substantiated neglect and that of substantiated physical/emotional abuse, respectively. Waller and Gotway (2004) suggest the following transformation formula:

$$z_i = \ln(\frac{1000 * (Y_i + 1)}{WN_i}) \tag{4.2}$$

where ln() is the natural logarithmic transformation function; and z_i is the transformed rate in census tract *i*. This formula is useful because it not only gives valid values for those tracts with $Y_i = 0$, but also helps discriminate the tracts with $Y_i \le 1$ but with different WN_i, and reduces the dependence of variance on the mean, i.e., heteroskedasticity

(Waller & Gotway, 2004). The histograms of the transformed rates are displayed in Figure 4.3 ((c) and (d)). Obviously, the transformation reduces skewness (compare to the histograms in 4.3 (a) and (b)).

4.2.2 Explanatory Variables

The explanatory variables consist of child-, parent-, and family-risk variables in microsystem, as well as socioeconomic, demographic, and housing variables indicating inadequate social supports and socioeconomic resource drain in exosystem (see Table 3.1 for the variable names and their definitions). The former are obtained from the vital records birth data, and the latter from the 2000 U.S. Census database.

4.2.2.1 Microsystem Variables

The birth records for 1996 through 2002 are used to derive child-, parent-, and family-risk variables. The birth records include data for calculating variables CHILDRISK, MAGELT20, SUBSTANCE, SIBLINGS3, and NMARRIED. Medicaid data, obtained from the Georgia Department of Community Health and linked to the birth records, are used to calculate variable MEDICAID. All birth records for the years 1996 through 2002 are processed to create Boolean fields, 1 meaning present and 0 meaning not present, for each of the individual risk variables including maternal age less than 20, having three or more siblings, non-married mother, and Medicaid beneficiary. The risk composite for a child is coded 1 if any of the three neonatal difficulties are present: birth weight less than 2,500 grams, gestation less than 37 weeks, or 5-minute Apgar score less than 7. Similarly, the risk composite indicating substance abuse of the mother is coded 1 if the mother smoked or drank alcohol during pregnancy. For any record, if any of the risks are unknown, the record is omitted for the calculations of risk variables. The value of each risk variable in a tract is obtained by calculating the percentage of births coded 1. Figure 4.4 displays the histograms of six microsystem risk variables.

4.2.2.2 Exosystem Variables

The socioeconomic, demographic, and housing variables at the census tract level are obtained from the 2000 Census data. Table 4.2 lists their source variables and files in the US Census database. Figure 4.5 displays the histograms of six exosystem risk variables. Four variables, SINGPARCH6, FEMLBCH6, VACANTHSE, and UNEMPMF, have highly positively skewed frequency distribution.

Table 4.1 Descriptive statistics of child relationships with the maltreaters

Label	Relationship	Count	Percent
1	Biological parent	4,532	91.33
2	Adoptive parent	13	0.26
3	Step-parent	12	0.24
4	Foster Parent	33	0.67
5	Grandparent	107	2.16
6	Uncle/Aunt	59	1.19
7	Biological Sibling	33	0.67
8	Step Sibling	3	0.06
9	Other Relative	16	0.32
10	Babysitter/Childcare	54	1.09
11	Other Non-Related Person	46	0.93
12	Relationship Unknown	13	0.26
13	Live In boyfriend or Girlfriend's house	32	0.64
14	School Personnel	7	0.14
15	Residential/ Facility Staff	2	0.04
Total		4,962	100

Table 4.2 Risk variables related to inadequate social supports and unemployment from the 2000 census data

Variable Name	**Conceptual Definition**	**Variable Description**	**Census Variable**	**Census Dataset**
SINGPARCH6	Single parents	% of families with own children under six years old where "male householder only, no wife present" or "female householder only, no husband present"	P015	SF3
FEMLBCH6	Females in labor force	% of females 16 years and older in the labor force who have own children under six years old	P045	SF3
RESIDLT1Y	New residents	% of persons who moved to the housing units (owner occupied or renter occupied) in 1999 to March 2000	HCT009	SF3
SINGFAMHSE	Single family dwellings	% of housing units with single dwelling structure ("1, detached" or "1, attached")	H030	SF3
VACANTHSE	Vacant housing	% of housing units with occupancy status "vacant"	H003	SF1
UNEMPMF	Unemployment	% of males and females 16 years old and older in the labor force who are unemployed	P043	SF1

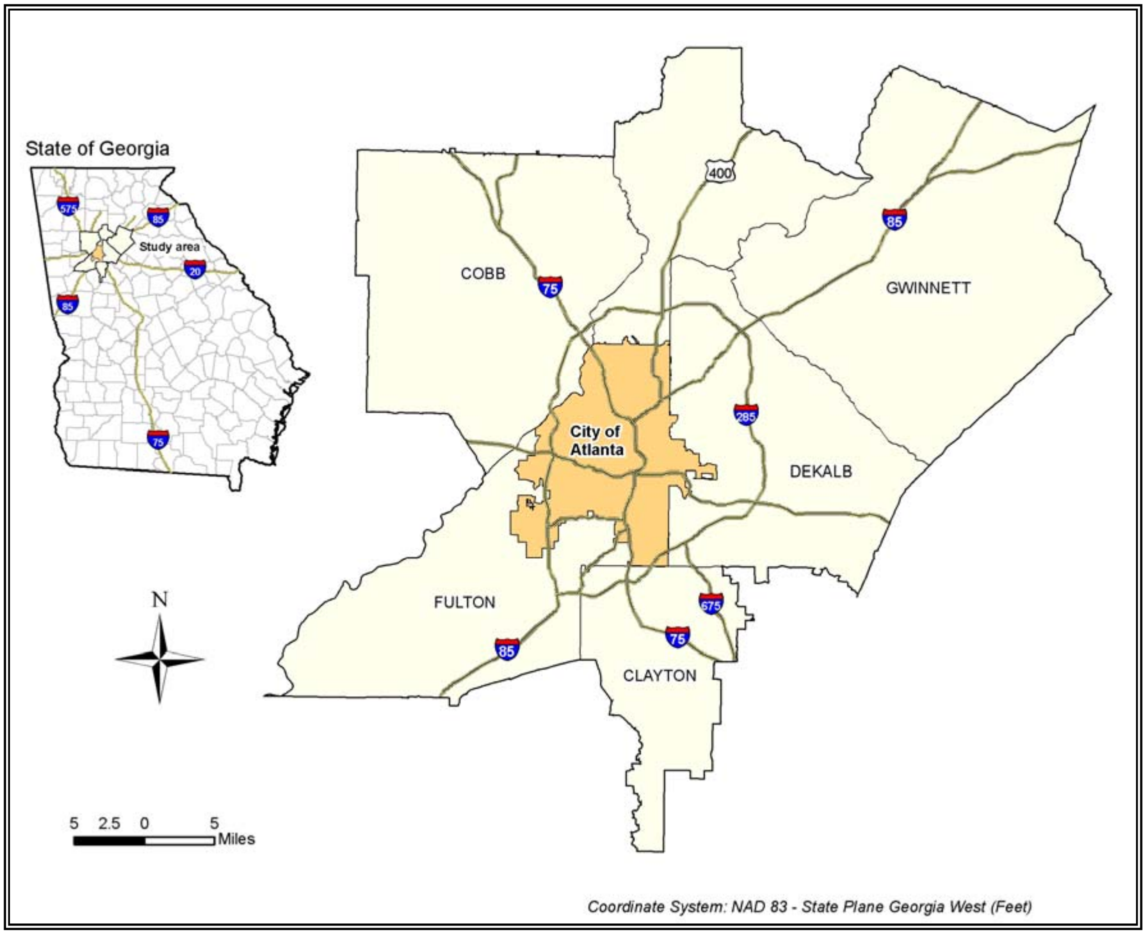


Figure 4.1 Study area covering five core counties in Metro Atlanta, Georgia

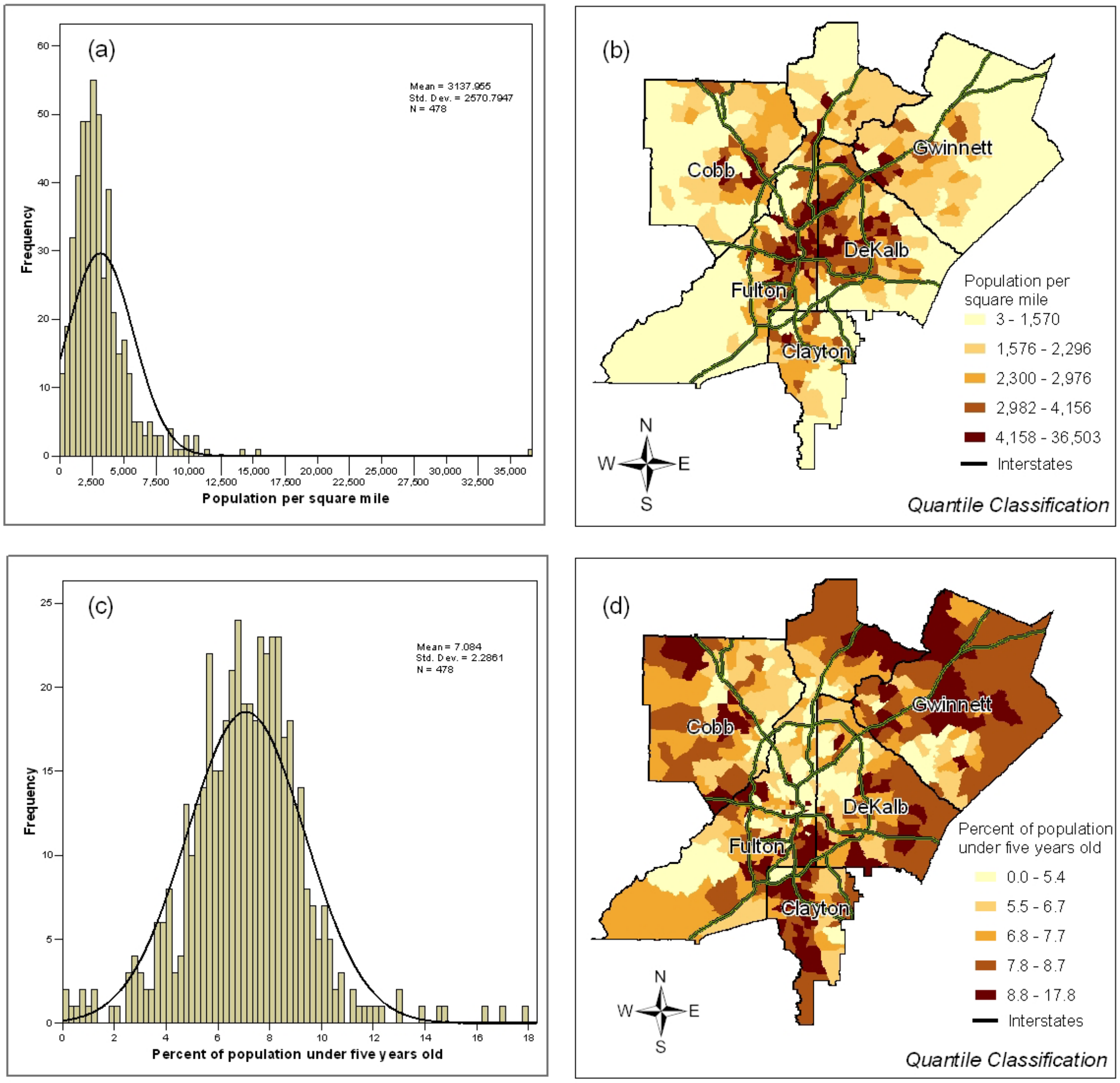


Figure 4.2 Distribution of population by census tract in the study area: (a) histogram of population density; (b) geographic distribution of population density; (c) histogram of percent of population under five years old; (d) geographic distribution of percent of population under five years old.

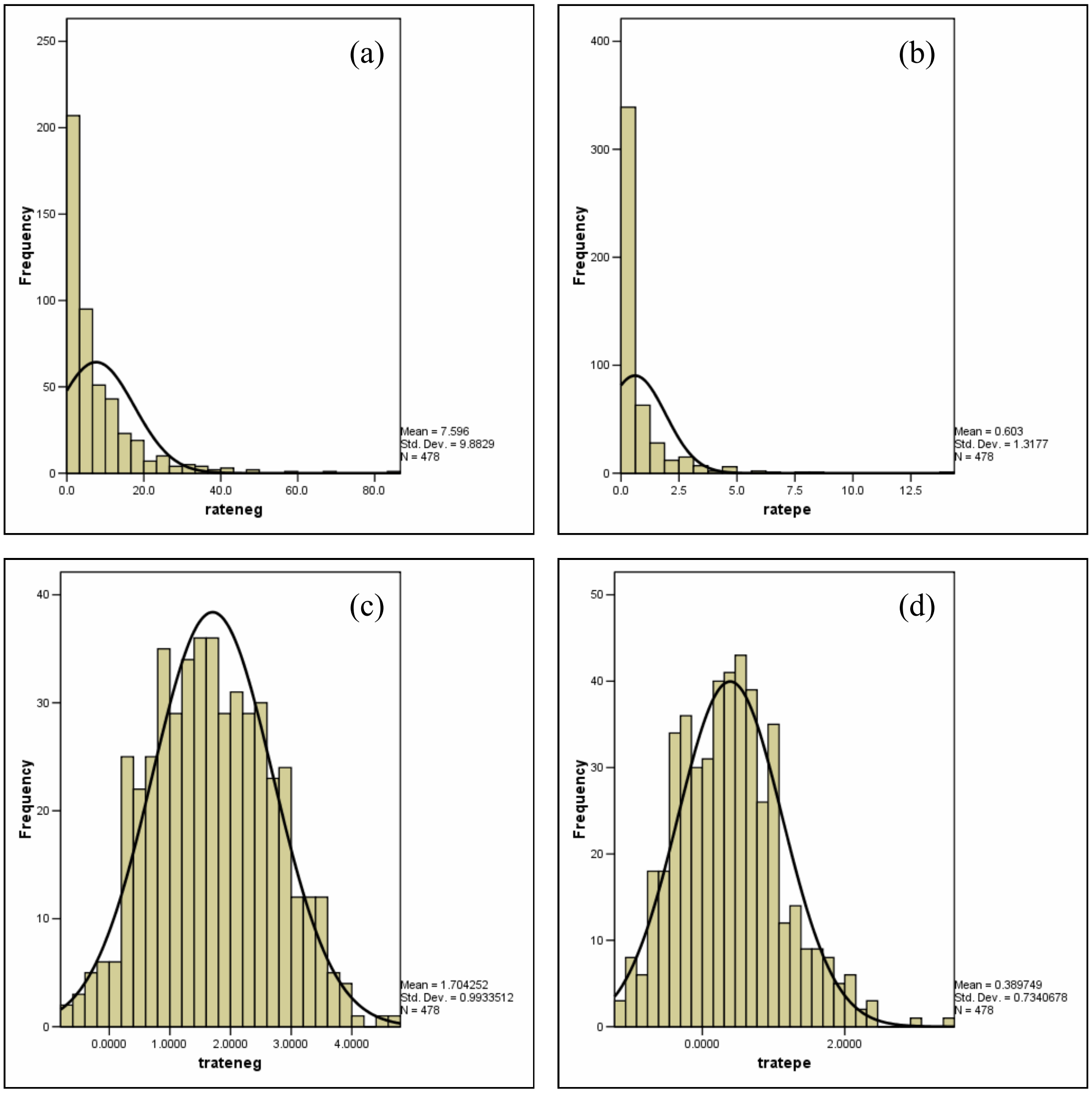


Figure 4.3 Histograms of rates and transformed rates of substantiated neglect, and substantiated physical/emotional abuse: (a) rate of substantiated neglect; (b) rate of substantiated physical/emotional abuse; (c) transformed rate of substantiated neglect; and (d) transformed rate of substantiated physical/emotional abuse.

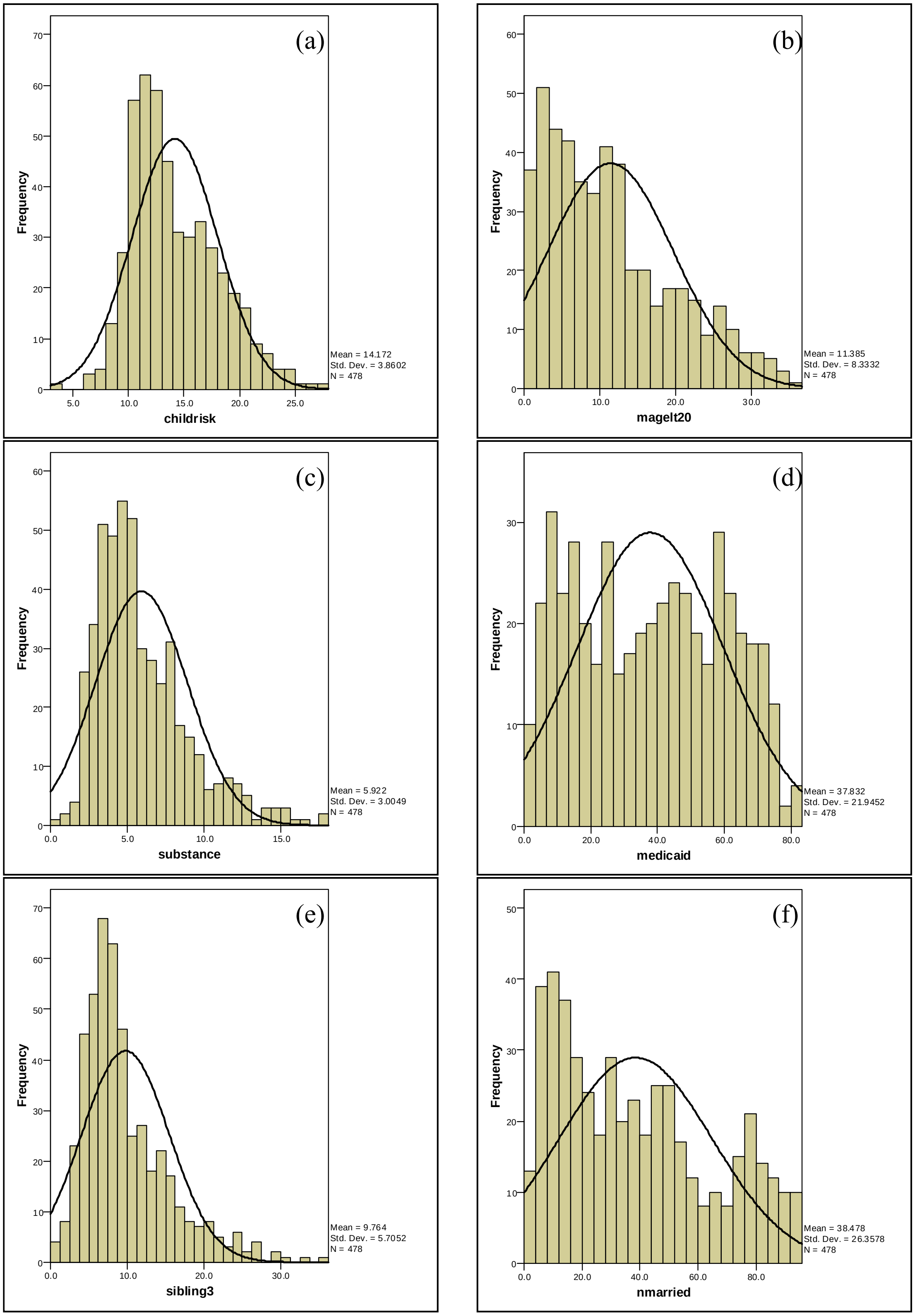


Figure 4.4 Histograms of the microsystem variables

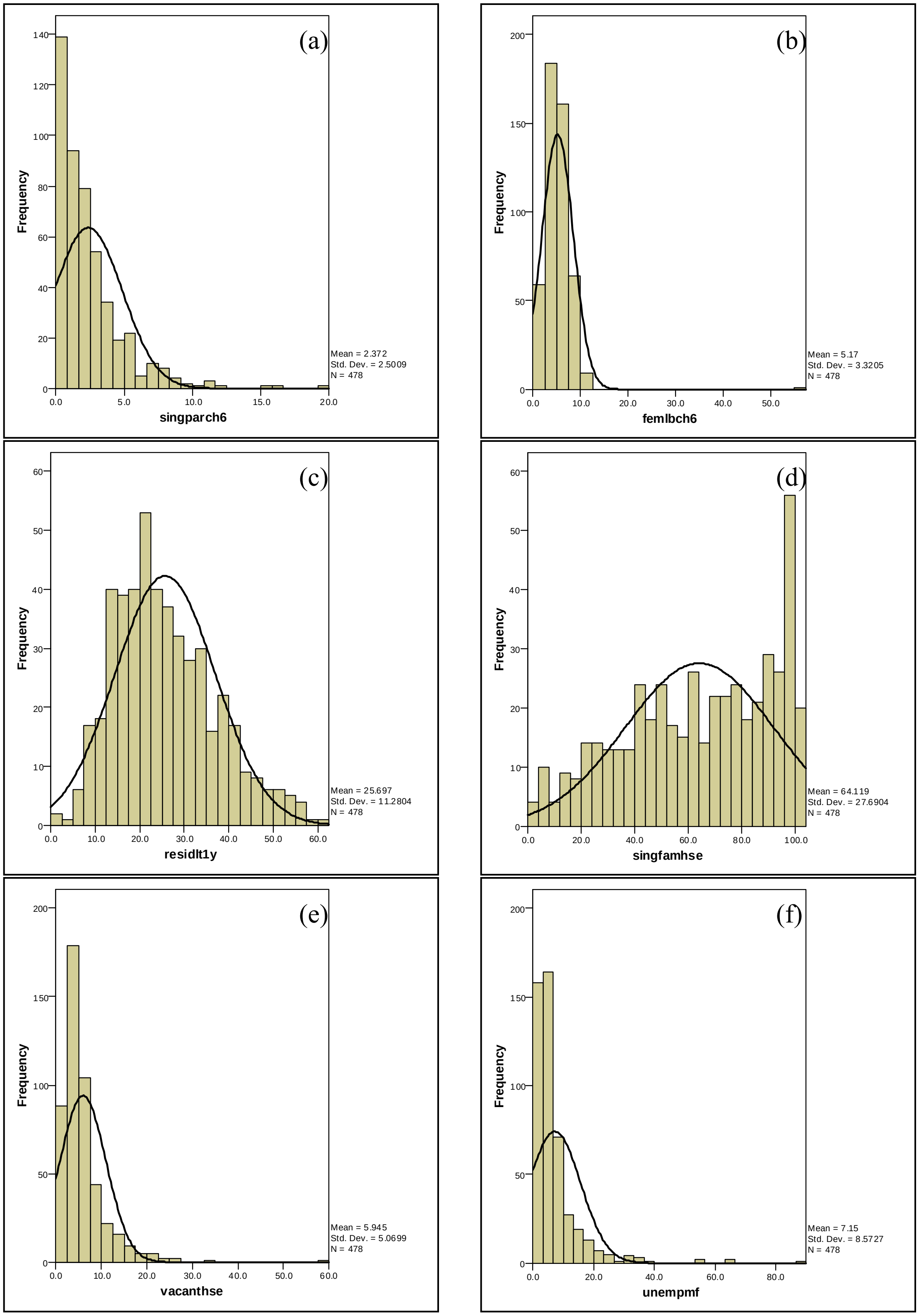


Figure 4.5 Histograms of the exosystem variables

CHAPTER FIVE

Method of Analysis

5.1 VISUAL ANALYSIS

Before quantitative analyses, all the explanatory variables and both response variables are mapped. The mapped data are reviewed to determine if the distribution suggests any patterns or relationships among mapped features (Hallisey, 2005).

Also, each pair of the response-explanatory variables is displayed in a scatter plot. The scatter plots are investigated to determine if there is a relationship, and if the relationship is linear or nonlinear, positive or negative. If a scatter plot reveals apparent linear relationship, the bivariate linear regression method is used to determine if the relationship is statistically significant.

5.2 TRADITIONAL LINEAR REGRESSION

5.2.1 Bivariate Linear Regression

Bivariate linear regression is used to quantitatively determine if there is a relationship between a response variable and an individual risk variable.

Let z_i denote the value of the response variable, and x_i denote the value of an explanatory variable in census tract *i*. A bivariate regression equation is expressed as:

$$z_i = b_0 + b_1 x_i + e_i ,\ i = 1,2,\cdots,n \qquad (5.1)$$

where, b_0 and b_1 are regression coefficients, in which, b_0 is the intercept, and b_1 is the slope reflecting the relationship between x and z (without loss of generality, the subscript index *i* is removed); e_i is the error term; and n is the number of census tracts.

The system of *n* equations associated with *n* census tracts is expressed in matrix notation as:

$$z = X\beta + e \tag{5.2}$$

where $z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$; $X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$; $\beta = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$; and $e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$.

The regression coefficients, b_0 and b_1, are unknown parameters. In traditional statistics with nonspatial data, the OLS method is used to estimate these parameters. The OLS estimator of β is obtained by minimizing the sum of squared differences between the observed and predicted values of the response variable, i.e., residuals (Rogerson, 2006), that is,

$$\underset{b_0,b_1}{Min}\sum_{i=1}^{n}(z_i - \hat{z}_i)^2 = \underset{b_0,b_1}{Min}\sum_{i=1}^{n}(z_i - b_0 - b_1 x_i)^2 \tag{5.3}$$

leading to the following system of equations in matrix notation:

$$(X^T X)\cdot\beta = X^T z \tag{5.4}$$

where X^T is the transpose of X. Estimates of β are obtained by solving Equation (5.4), leading to:

$$\hat{\beta} = \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \end{bmatrix} = (X^T X)^{-1} \cdot X^T z \tag{5.5}$$

in which $(X^T X)^{-1}$ is the inverse of $X^T X$ provided $X^T X$ is invertible (Chatterjee, Hadi, & Price, 2000).

In order for the statistical inference about parameter estimates to be valid, it is assumed that z_i ($i = 1,2,\cdots,n$) are independent and normally distributed observations; e_i ($i = 1,2,\cdots,n$) are independent and normally distributed with a constant mean of zero and constant variance of σ^2 (homogeneity); and $\beta = [b_0\ b_1]^T$ are constant across the whole dataset.

When these assumptions are satisfied, the estimates of b_0 and b_1, denoted as $\hat{b}_0$ and $\hat{b}_1$, are the best linear unbiased estimates. The sign of $\hat{b}_1$ gives the direction of the relationship between x and z. The standard hypothesis testing procedures are then used to test if the value of $\hat{b}_1$ is statistically significantly different from the null hypothesis H_0: b_1=0. The test can be done via t-statistic calculated as:

$$t = \frac{\hat{b}_1}{s.e.(\hat{b}_1)} \tag{5.6}$$

where $s.e.(\hat{b}_1)$ is the estimated standard error of the slope. When the null hypothesis is true, *t* has a *t*-distribution with *n*-2 degrees of freedom (Rogerson, 2006). Here the number “2” reflects two unknown parameters: slope and intercept.

One of the measures to assess how good the observations of the response variable are fitted by the regression model is R^2 (Chatterjee et al., 2000). The measure R^2, which also is called the coefficient of determination, measures the proportion of the total variability of the observed values of the responsible variable explained by the regression model, i.e.,

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{z}_i - \bar{z})^2}{\sum_{i=1}^{n} (z_i - \bar{z})^2} \tag{5.7}$$

where $\hat{z}_i$ and $\bar{z}$ are the predicted value and the expected mean of z_i, respectively.

Another interpretation of R^2 is it measures the strength of correlation between the observed (z) and predicted ($\hat{z}$) values of the response variable, that is, $\sqrt{R^2} = Cor(z, \hat{z})$. Here, $Cor(z, \hat{z})$ is the Pearson's correlation coefficient between *z* and $\hat{z}$. A third interpretation is it measures the strength of the linear association between the response variable (z) and the explanatory variable (x), i.e., $\sqrt{R^2} = Cor(z, x)$, where $Cor(z, x)$ is the Pearson's correlation coefficient between *z* and *x*.

5.2.2 Multivariate Linear Regression

Multivariate linear regression is used to identify a set of explanatory variables that best predicts the response variable. Let $x_1, x_2, \cdots x_k$ denote k risk variables chosen to be included in the regression equation, and z the response variable. A multivariate regression equation is expressed in matrix notation as:

$$z = X\beta + e \tag{5.8}$$

where $X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix}$; $\beta = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix}$; and z and e are defined the same as for Equation (5.2). b_j is called the partial regression coefficient. It represents the contribution of variable x_j to the response variable after it has been adjusted for the other explanatory variables (Chatterjee et al., 2000). The OLS estimates of β are $\hat{\beta} = (X^T X)^{-1} \cdot X^T z$ provided that $X^T X$ is invertible.

In addition to the assumptions given for the bivariate regression, there is one additional assumption for the multivariate regression. That is, there is no multicollinearity among the explanatory variables, which means the correlation among the explanatory variables should not be high (Rogerson, 2006). If any two explanatory variables are perfectly correlated, it is impossible to estimate the regression coefficients because in this case $X^T X$ is not invertible. If correlation is high but not perfect, which is commonly encountered in real applications, $X^T X$ is ill-conditioned. Inversion of such an ill-conditioned matrix is unstable, and thus the parameter estimates will have large errors, which in turn affect both prediction and inference of the regression model (Chatterjee et al., 2000; Rogerson, 2006).

A diagnostic to suggest the overall multicollinearity of the explanatory variables is the multicollinearity condition number (MCN) (Belsley, Kuh, & Welsch, 1980; Anselin, 2005), which is used in the present study. It is found that the effect of multicollinearity on parameter estimation becomes observable when MCN takes a value about 10; and a value between 30 and 100 is associated with moderate to strong multicollinearity (Belsley et al., 1980). In the present study, 10 is used as the threshold.

The formula to calculate the *t*-statistic for testing individual parameters is similar to Equation (5.6) and takes the form:

$$t = \frac{b_j}{s.e.(b_j)}, \quad j = 1, 2, \cdots, k$$

but *t* has a *t*-distribution with *n*-*k*-1 degrees of freedom, where *k* is the number of explanatory variables. $s.e.(b_j)$ is the estimated standard error of the *j*th parameter. The goodness-of-fit measure R^2 for the multivariate linear regression has the same formula and definition as for bivariate linear regression. Besides, $R = \sqrt{R^2}$ is called the multiple correlation coefficient and measures the association between the responsible variable *z* and *k* variables $x_1, x_2, \cdots, x_k$ (Chatterjee et al., 2000).

5.3 DEALING WITH SPATIAL HETEROGENEITY

As noted in Chapter Two, spatially aggregated data are characterized by spatial autocorrelation and spatial heterogeneity (Anselin, 1988). Spatial heterogeneity has two aspects. The first is related to the lack of stationarity of geographic phenomena over space, which means that relationships between the response variable and the explanatory variables change geographically (Fotheringham, Brunsdon, & Charlton, 2002). The other is related to the varying size and shape of spatial aggregation units, which may result in heteroskedasticity, that is, the variance depends on the mean (Anselin, 1988)

To deal with the first aspect of spatial heterogeneity, the study area is confined to relatively small and homogeneous urban settings. Heteroskedasticity is reduced through the transformation of the response variables using a variance-stabilizing transformation expressed by Equation (3.2).

In the present study, special attention is paid to spatial autocorrelation in the error term. The presence of spatial autocorrelation violates the assumption of independence. Spatial regression is used to account for the effect of spatial autocorrelation.

5.4 SPATIAL REGRESSION

5.4.1 Dealing with Spatial Autocorrelation by Spatial Regression

Assuming there are no serious problems with heteroskedasticity and multicollinearity, the OLS method provides the best linear unbiased estimates only if the regression model is correctly specified so that the residuals are independent and normally distributed with zero mean and constant variance. A regression model expressed by Equation (5.8) is considered misspecified in several situations: 1) the response variable is inherently spatially dependent; 2) the unit of analysis does not match the unit of actual phenomena; 3) important explanatory variables are missing (not included in the model); and 4) the observations of the response and/or explanatory variables are not free of errors (Anselin, 1988; Anselin & Bera, 1998; Waller & Gotway, 2004).

If a regression model is misspecified, the residuals after the OLS fitting are not independent; instead, the residual at one location may be correlated with the residuals at nearby locations, resulting in the clustering of similar residuals among nearby locations (Anselin & Bera, 1998). When the residuals are spatially autocorrelated, the OLS estimates are no longer best linear unbiased and the estimated standard errors are likely to be downward biased (Benirschka & Binkley, 1994). The direct consequence of the downward biasedness of standard errors is that the absolute values of the test statistic are upward biased. This implies a significant relationship between the response variable and an explanatory variable suggested by the regression analysis may actually be insignificant. To obtain reliable parameter estimates, the spatial autocorrelation effect must be accounted for. This can be achieved by the use of the spatial regression method (Anselin, 1988; Anselin & Bera, 1998).

If one or more of the above situations exist, the error term e in equation (5.8) is spatially autocorrelated. In a spatial regression model, spatial autocorrelation in e is considered an additional variable in the model specification; its effect is solved simultaneously with the effects of other explanatory variables (Anselin, 1988).

There are two methods to incorporate spatial autocorrelation in a regression model. One is to model spatial autocorrelation in the error term as the spatially lagged response variable, which is defined as the average of the values for neighboring locations. That is,

$$e = \rho Wz + \varepsilon \tag{5.9}$$

where W is the spatial weights matrix characterizing the spatial relationship (interaction) between every pair of spatial units; Wz is called the spatial lag of the response variable z ; ρ is the spatial autoregressive parameter characterizing the spatial autocorrelation effect; and ε is the independent and normally distributed error term with a constant mean of zero and constant variance. This is referred to as the spatial lag model.

The other method is to model spatial autocorrelation in the error term as the spatially lagged error term, that is,

$$e = \lambda We + \varepsilon \quad \text{or} \quad e = (I - \lambda W)^{-1}\varepsilon \tag{5.10}$$

in which W and ε are defined the same as for Equation (5.9); We is called the spatial lag of the error term e; I is the identity matrix; and λ is similar to ρ. This is referred to as the spatial error model.

Substituting the error term in (5.8) with Equation (5.9) gives the expression of a spatial lag model:

$$z = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} \varepsilon \tag{5.11}$$

Similarly, a spatial error model is obtained by substituting the error term in (5.8) with Equation (5.10), and expressed as:

$$z = X\beta + (I - \lambda W)^{-1} \varepsilon \tag{5.12}$$

The OLS method is no longer appropriate for estimating the parameters in Equations (5.11) and (5.12); instead, the maximum likelihood estimation (MLE) method or the instrumental variables estimation (IVE) method should be used (Anselin, 1988). The MLE method estimates model parameters by maximizing the *Likelihood Function* of the observations (Anselin & Bera, 1998).

In both spatial lag and spatial error models, the statistic for testing significance on explanatory variable parameters as well as on the spatial autocorrelation parameter is approximately the *z-score* calculated as (Anselin, 1988):

$$z - score = \frac{\hat{r}}{s.e.(\hat{r})} \sim N(0,1)$$

in which, $\hat{r}$ denotes the MLE estimate of any of the explanatory variable parameters ($b_0, b_1, \cdots, b_k$) or the spatial autocorrelation parameters (ρ or λ), and $s.e.(\hat{r})$ is the estimated standard error.

When the MLE method is used to estimate the parameters in Equation (5.11) or (5.12), the traditional goodness-of-fit measure R^2 is no longer valid for assessing model fits (Anselin, 1988). One of the appropriate measures is the Akaike Information Criteria (AIC). A model is considered the best among a set of alternatives if the model gives the smallest AIC value. An approximate measure that mimics the traditional measure R^2 is so-called pseudo-R^2, which provides a measure of linear association between the observed and predicted values of the response variable; but it is no longer related to the variance component explained by the model (Anselin, 1988). A small pseudo-R^2 may suggest a low predictive ability of the model; however, a model with the highest pseudo-R^2 value cannot be considered the best among a set of alternatives.

The design criteria discussed by Gilbert and cited in (Haining, 2003) are used to guide the identification of a set of explanatory variables that best predicts the response variable. Gilbert and Haining contend a model should: 1) be fit-for-purpose, meaning the model must enable the analyst to answer the research question; 2) be robust, meaning there is no serious multicollinearity among the explanatory variables; 3) give uncorrelated residuals.

The first criterion implies that a variable needs to be excluded from the regression model if the sign of its estimated coefficient is at the opposite direction to its relationship with the response variable shown in the scatter plot. The second criterion is regarding satisfaction of no multicollinearity. The third criterion is fulfilled using the spatial regression method.

To meet the criterion of no multicollinearity, the backward selection method is used. That is, all relevant variables are first included in the regression model, and solved using the OLS method. Then the value of MCN is checked to see if multicollinearity is a problem, i.e. if the value of MCN is equal to or greater than 10. If yes, one variable is taken out and the remaining variables are solved using the OLS method. Repeat the process until multicollinearity is not a problem, i.e., MCN<10. The variable that needs to be taken out each time is either having an opposite sign or not statistically significant. The minimum number of variables retained in the final model is one.

5.4.2 Spatial Weights Matrix

Spatial weights are essential in spatial regression models. They represent the spatial relationship between observation units, that is, whether two units are in each other's neighborhood. The rationale is that interaction between any two observations occurs if the two units are in each other's neighborhood.

The spatial relationship between any two units can be determined based on either the distance or contiguity between them (Mitchell, 2005). The former works best for point observation units; the latter is often used when the observation units are areas. In the present study, spatial weights are determined based on contiguity among census tracts.

In matrix notation, spatial weights among n census tracts are represented using a spatial weights matrix W, which is a $n \times n$ binary (0-1 values) and symmetric matrix (Anselin & Bera, 1998), expressed as:

$$W = \begin{bmatrix} 0 & \cdots & w_{1j} & \cdots & w_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ w_{i1} & \cdots & w_{ij} & \cdots & w_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ w_{n1} & \cdots & w_{nj} & \cdots & 0 \end{bmatrix}$$

By convention, the diagonal elements are set to zero. For any non-diagonal element w_{ij}, $w_{ij} = 1$ when *i* and *j* are neighbors, and $w_{ij} = 0$ otherwise.

There are three options to determine contiguity. They are referred to as the Rook contiguity (only common boundaries), the Bishop contiguity (only common vertices), and the Queen contiguity (both common boundaries and vertices) (Anselin, 2002). Furthermore, contiguity needs not to be limited to first order, i.e., direct adjacency; higher order contiguity can also be determined (Anselin, 2003). Which type and order of spatial weights should be used, however, is still subjective although the use of different types of spatial weights leads to different results (Anselin, 1988).

In the present study, Queen contiguity is used. The rationale is that the interaction among areas should not be limited to areas that share non-zero length boundaries; it occurs among areas that share vertices as well. The order is determined empirically. That is, first run regression with the *1*st, *2*nd, ···, *k*th weights, and then choose the order that provides the best model (i.e., having the smallest AIC value).

5.5 SPATIAL REGRESSION SOFTWARE

The software used for the spatial regression analysis in the present study is GeoDA (Version 0.9.5i_6) developed by the Spatial Analysis Laboratory in the Department of Geography at the University of Illinois, Urbana-Champaign. The software provides tools to calculate and manipulate spatial weights, and provides functions for descriptive spatial data analysis, such as spatial autocorrelation statistics, as well as spatial regression functionality (Anselin, 2003, 2004).

GeoDA can create contiguity-based spatial weights for polygon spatial files and distance-based spatial weights for any input files with x- and y- coordinates available. For contiguity-based spatial weights, there are two options: Rook contiguity and Queen contiguity. The default order of contiguity is one, but higher order weights can be created as well.

GeoDA provides functions to generate graphs for exploratory spatial data analysis (ESDA) including histograms, scatter plots, box plots, and other types of plots for advanced ESDA purposes such as parallel coordinate plots, 3D plots, and conditional plots.

When spatial weights have been created and opened, GeoDA can perform global and local spatial autocorrelation analysis for single variable (Univariate) or a pair of variables (Multivariate). Global spatial autocorrelation analysis is handled by means of the Moran's *I* statistic and can be visualized in the form of a Moran scatter plot. A Univariate Moran's *I* statistic represents the correlation between a variable and its spatial lag; while a Multivariate Moran's *I* statistic represents the correlation between one variable and the spatial lag of another variable. Similarly, a Univariate Moran scatter plot shows the standardized values of a variable on the horizontal axis and the standardized values of the spatial lag of the same variable on the vertical axis; a Multivariate Moran scatter plot shows the standardized values of one variable on the horizontal axis and the standardized values of the spatial lag of another variable on the vertical axis. The slope of the regression line in a Moran scatter plot is Moran's *I*. Inference for Moran's *I* (both

Univariate and Multivariate) is based on a permutation approach, which uses a randomization algorithm to generate a number of random replications of the data set under the null hypothesis, and the test statistic is then calculated for each random replication, from which the critical value for inference is derived (Anselin, 2003).

Local spatial autocorrelation analysis is based on the local indicator of spatial autocorrelation (LISA) statistics and can be visualized in the form of the significance map, the cluster map, the box plot, or the Moran scatter plot.

GeoDA can run regression analysis based on three types of regression models: Classic (OLS), Spatial Lag, and Spatial Error models. The output of the OLS regression includes diagnostics for multicollinearity (the value of MCN), nonnormality and heteroskedasticity, as well as five Lagrange Multiplier (LM) test statistics against spatial autocorrelation. Among the five LM test statistics, LM-Lag and Robust LM-Lag pertain to the spatial lag model, while LM-Error and Robust LM-Error refer to the spatial error model. The last test, LM-SARMA, is related to the higher order model that includes both the spatial lag and spatial error terms. This last test is not useful in the current version of software because the software does not allow the user to select both models. The software allows the user to first run the OLS model; then examine the test statistics to see if spatial autocorrelation is significant to consider, and if so, decide which spatial model should be used. Figure 5.1 illustrates the decision process of spatial regression model selection.

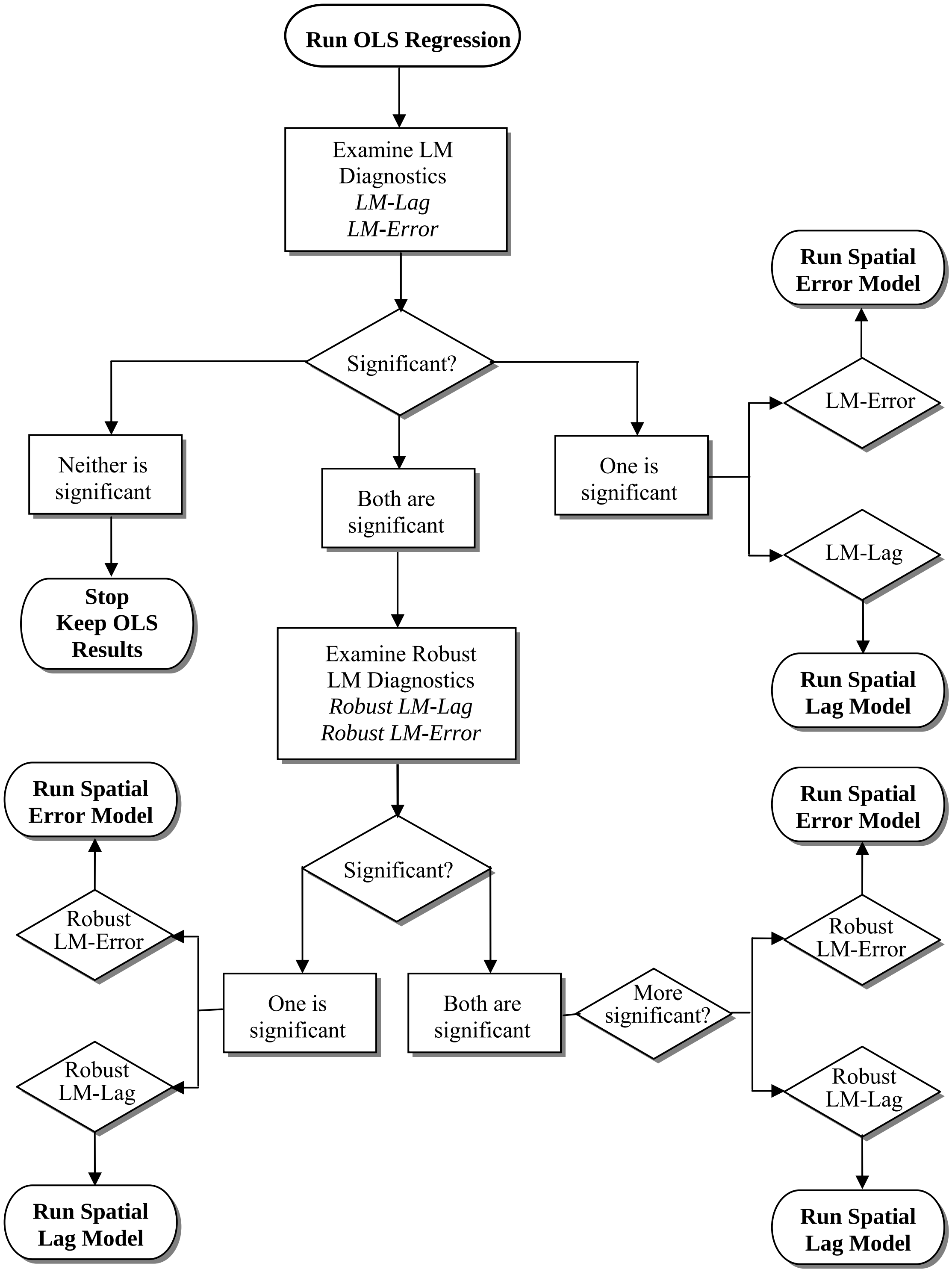


Figure 5.1 Decision process of the spatial regression model selection

CHAPTER SIX

Results

6.1 VISUAL ANALYSIS

6.1.1 Spatial Distribution of Rates and Transformed Rates

The distribution of substantiated child maltreatment in 2000 through 2002 in the study area varies by type of maltreatment and by community. Table 6.1 displays the descriptive statistics. Substantiated neglect occurred in more census tracts than substantiated physical/emotional abuse did. For substantiated neglect, 3,526 victims lived in 405 out of 478 census tracts; 73 census tracts had no victims and one tract had a maximum of 67 victims. The rate of substantiated neglect varies from 0.0 to 86.1 per 1,000 weighted counts of children under four years old. For substantiated physical/emotional abuse, 313 victims lived in 167 census tracts; 311 census tracts had no victims and one tract had a maximum of 9 victims. The rate of substantiated physical/emotional abuse varies from 0.0 to 14.3 per 1,000 weighted counts of children.

Figure 6.1 displays the maps of the rates of substantiated neglect (denoted as RATENEG and shown in 6.1 (a)) and rates of substantiated physical/emotional abuse (denoted as RATEPE and shown in 6.1 (b)) by census tract. Spatial autocorrelation is clearly seen in Figure 6.1 (a). The tracts with high rates of substantiated neglect are mainly concentrated in the south, specifically in southern Fulton County and in Clayton County. They are also found in central DeKalb County and the most urbanized area of Cobb County. The tracts having no substantiated neglect victim or low rates are in the north, i.e., in Gwinnett County, the central and north parts of Fulton County, and to a lesser extent eastern Cobb County.

Spatial autocorrelation is less pronounced with regards to the rates of substantiated physical/emotional abuse in contrast to substantiated neglect. The tracts with high rates of substantiated physical/emotional abuse are found in all five counties, although the tracts in the highest classification are more often found in Fulton County along Interstate 20. The tracts with no substantiated physical/emotional abuse victims occupy the majority of the study area.

Figure 6.2 presents the maps of the transformed rates of substantiated neglect (denoted as TRATENEG and shown in 6.2 (a)) and the transformed rates of substantiated physical/emotional abuse (denoted as TRATEPE and shown in 6.2 (b)) by census tract. As expected, Figure 6.2 (a) is very similar to Figure 6.1 (a) because the transformation function expressed by Equation (4.2) is monotonic. However, Figure 6.2 (b) looks different from Figure 6.1 (b). The reason is as follows. The transformation function has the ability to discriminate the tracts having no victim but different weighted counts of children (see Subsection 4.2.1.3). Thus the 311 census tracts, where the values of the raw rate of substantiated physical/emotional abuse equal zero and categorized into one class in Figure 6.1 (b), are categorized into different classes in Figure 6.2 (b). Although the transformation has a similar impact on the73 tracts, where the raw rates of substantiated neglect equal zero and categorized into one class in Figure 6.1 (a), so that they may be classified into different classes in Figure 6.2 (a), the changes in the classification between Figure 6.2 (a) and Figure 6.1 (a) are not visually noticeable as opposed to the changes between Figures 6.2 (b) and 6.1 (b).

6.1.2 Spatial Distribution of Risk Variables

The risk variables are defined in Table 3.1. Figure 6.3 displays the spatial distribution of the microsystem risk variables. All variables except SUBSTANCE (percent of births to mothers who smoked and/or drank alcohol during pregnancy, shown in 6.3 (c)) have a similar pattern of distribution, i.e., tracts in the highest two categories are in the south, while those in the lowest two categories are in the north. It is interestingly noted that DeKalb County has low percentage of births to mothers who smoked and/or drank alcohol during pregnancy. Other tracts in the lowest two categories of variable SUBSTANCE are clustered in northern Fulton County, and found in some parts of Cobb and Gwinnett counties.

Figure 6.4 shows the spatial distribution of the exosystem risk variables. In contrast to the microsystem variables, in which five out of six variables have a similar distribution pattern, the patterns of the exosystem risk variables are quite different. For variable SINGPARCH6 (6.4 (a)), tracts in the highest two categories are mainly found in southern Fulton and Clayton counties, central to south DeKalb County, and some parts of Cobb County. Variables FEMLBCH6 (6.4 (b)) and SINGFAMHSE (6.4 (d)) have inner-outer differentiation. That is, tracts inside the Perimeter (I-285) generally have lower percentage of single-family housing units and females in the labor force with children under six years of age, and tracts in the highest two categories of these two variables are mainly found outside the Perimeter. Tracts with high percentage of residential instability (variable RESIDLT1Y, 6.4 (c)) are mainly along the expressways where transportation is more convenient. Tracts inside the Perimeter have high percentage of vacant houses (variable VACANTHSE, 6.4 (e)). Tracts in the highest classification of variable

UNEMPMF (6.4 (f)) are mainly clustered in the south part of the City of Atlanta (refer to Figure 4.1 to locate the boundary of the City of Atlanta).

6.1.3 Map Comparison Between the Transformed Rates and Risk Variables

In general, similarity of distribution patterns is more visible between the response variable TRATENEG and the microsystem risk variables than any of the other comparisons (TRATENEG with the exosystem risk variables; TRATEPE with the microsystem risk variables, and TRATEPE with the exosystem risk variables).

As noted earlier, five out of six microsystem risk variables have a similar pattern of distribution (Figure 6.3: (a), (b), (d)-(f)). This pattern also is seen in the distribution of TRATENEG (Figure 6.2 (a)). However, similarity is less pronounced between the map of TRATENEG and that of the risk variable SUBSTANCE (Figure 6.3 (c)). Among six maps of the exosystem variables, the map of UNEMPMF (Figure 6.4 (f)) has most visible similarity with the map of TRATENEG. Next is the map of variable SINGPARCH6 (Figure 6.4 (a)). Similarity is hardly visible between the map of TRATENEG and maps of any of the other exosystem variables: FEMLBCH6 (6.4 (b)), RESIDLT1Y (6.4 (c)), SINGFAMHSE (6.4 (d)), and VACANTHSE (6.4 (e)).

In contrast to TRATENEG, similarity of distribution patterns is very much less pronounced between the response variable TRATEPE (Figure 6.2 (b)) and any of the microsystem and exosystem variables. Only is the map of variable SUBSTANCE (Figure 6.3 (c)) that has some visually noticeable similarity with Figure 6.2 (b).

To further visually examine the relationships between the response and the explanatory variables, the scatter plots are investigated. Investigating scatter plots can help determine if a relationship is linear or nonlinear, positive or negative.

6.1.4 Investigating Scatter Plots

Figure 6.5 displays the scatter plots of the response variable TRATENEG with the microsystem variables. It is seen that positive, linear relationships exist between TRATENEG and four microsystem variables: CHILDRISK (6.5 (a)), MAGELT20 (6.5 (b)), MEDICAID (6.5 (d)), and NMARRIED (6.5 (f)). Positive, linear, but weaker relationship is found with variable SUBSTANCE (6.5 (c)). The relationship with variable SIBLING3 looks more nonlinear than linear (6.5 (e)).

Figure 6.6 presents the scatter plots of the response variable TRATENEG with the exosystem variables. Consistent with map comparison results, relationships are much less pronounced in Figure 6.6 compared with Figure 6.5. In-depth investigation reveals that no relationship is found with variable FEMLBCH6 (6.6 (b)) or RESIDLT1Y (6.6 (c)). Linear relationship is found with SINGFAMHSE (6.6 (d)), but the relationship is negative. The relationships between TRATENEG and variables SINGPARCH6 (6.6 (a)), VACANTHSE (6.6 (e)), and UNEMPMF (6.6 (f)) are more nonlinear than linear.

Figure 6.7 and 6.8 display the scatter plots of the response variable TRATEPE with the microsystem variables (6.7: (a) – (f)) and exosystem variables (6.8: (a) – (f)). Positive, linear relationship can be seen in all six plots in Figure 6.7, but is weak in strength. In Figure 6.8, no relationships are found with variables FEMLBCH6 (6.8 (b)) and with RESIDLT1Y (6.8 (c)); linear, but negative relationship is found with SINGFAMHSE (6.8 (d)). For the remaining three variables, SINGPARCH6 (6.8 (a)), VACANTHSE (6.8 (e)), and UNEMPMF (6.8 (f)), some sort of nonlinear relationship can be seen, but not definitively.

In summary, the visual analysis results do not support the idea that there are positive relationships between the response variables and risk variables FEMLBCH6, RESIDLT1Y, and SINGFAMHSE. These variables are then eliminated from the further statistical analysis. Variables SIBLING3, SINGPARCH6, VACANTHSE, and UNEMPMF have nonlinear relationships with both response variables. They cannot be directly included in linear regression models.

To include variables SIBLING3, SINGPARCH6, VACANTHSE in the further statistical analysis, they must be transformed to achieve linearity. Two transformations—square root and natural logarithm—are applied to all four variables. The rationale is that these transformations can achieve not only normality, but also linearity (Chatterjee et al., 2000). Variable SINGPARCH6 can only be transformed by the square root function because it has zero values in some census tracts. To determine which function is appropriate for the other three variables, both the square root and natural logarithmic transformations are applied. Figure 6.9 presents the histograms of the three variables by the two transformations, in which, (a), (c), and (e) correspond to the transformed variables by the natural logarithmic transformation, whereas (b), (d), and (f) to the transformed variables by the square root transformation. Comparison of all histograms in Figure 6.9 with the histograms in Figures 3.4 (e) (variable SIBLING3), 3.5 (e) (Variable VACANTHSE) and 3.5 (f) (variable UNEMPMF) suggests that both transformations achieved normality. However, in terms of providing better shape of histograms, the natural logarithmic transformation works better for variables VACANTHSE and UNEMPMF, whereas the square root transformation works slightly better for variable SIBLING3.

The four transformed variables are denoted as SSIBLING3 and SSINGPARCH6 (by the square root transformation), and LUNEMPMF and LVACANTHSE (by the natural logarithmic transformation). Figure 6.10 shows their scatter plots with the response variables TRATENEG ((a), (c), (e), and (g)) and TRATEPE ((b), (d), (f), and (h)), respectively. In general, the transformations archived linearity. However, it is seen that the scatter plot of SIBLING3 with TRATEPE in Figure 6.7 (e) demonstrates a better shape of linearity than Figure 6.10 (b). This suggests that the original variable SIBLING3 is more suited than its transformed variable SSIBLING3 in the regression of TRATEPE.

Finally, the explanatory variables to be examined quantitatively by the regression analysis are listed in Table 6.2. Table 6.3 presents their descriptive statistics.

6.2 REGRESSION ANALYSIS

6.2.1 Bivariate Regression

6.2.1.1 OLS Regression

Table 6.4 displays the results of bivariate OLS regression. The values displayed in the column "$\hat{b}$" are unstandardized regression parameter estimates. The models are ranked according to the AIC values. Ranking is made separately for the two response variables. The smaller the AIC value, the better the model. If the difference between the AIC values of two models is less than 3.0, these two models are considered tied, meaning not different from each other (Fotheringham et al., 2002).

The values of the test statistic of all the models are positive. The smallest value is 1.958 in the regression of TRATEPE on SSINGPARCH6, corresponding to a probability (*p-value* or *p)* of 0.0508. The second smallest value of the test statistic is 7.07, which is the smallest value of the test statistic of all the regression models of TRATENEG. Its corresponding *p-value* is smaller than 0.0000. Therefore, the response variable

TRATENEG is statistically significantly, positively related to all nine explanatory variables at $p<0.0000$. The response variable TRATEPE is statistically significantly, positively related to eight explanatory variables at $p<0.0000$; its relationship with SSINGPARCH6 is positive but not significant at $p=0.05$.

The next step is to examine the residuals. The OLS regression results cannot be accepted as final if spatial autocorrelation in the residuals is significant to consider. Figures 6.11 - 13 display the standard deviation maps of the residuals. Blue colors illustrate negative residuals (over-predication), and brown colors illustrate positive residuals (under-prediction). The darkest colors display the areas where the absolute residuals are greater than two standard deviations. Visual comparison of these maps with the maps in Figure 6.2 suggest that all the bivariate OLS regression models over-predict the low values and under-predict the high values of both TRATENEG and TRATEPE. Furthermore, visual inspection of spatial patterns suggests the presence of spatial autocorrelation in all the residual maps. Therefore, spatial regression must be conducted to account for the effects of spatial autocorrelation.

6.2.1.2 Spatial Regression

Before performing spatial regression, two issues must be resolved: what is the appropriate order of spatial weights and which spatial regression model (the spatial lag or spatial error model) should be chosen. To determine which model should be chosen, the spatial regression decision process illustrated in Figure 5.1 is followed. To determine which order of spatial weights should be used, an empirical method is used. That is, first run the OLS and spatial regression with the 1st, 2nd,···, 5th order of weights; then choose the order that provides the best models (i.e., have the smallest AIC values). If two AIC values are not significantly different, i.e., their difference is smaller than 3.0 (Fotheringham et al., 2002), the lower order is chosen. The results of determining the order of weights are shown in Table 6.5. For all models, the AIC values first decrease and then increase, as the order of weights gets higher. The bolded cell in each row is the smallest AIC value among five values except three cells displayed in Bolded Italic, where the values are the second smallest because their differences from the smallest values are less than 3.0.

Once the order of spatial weights is specified based on Table 6.5, the Moran's *I* value of the OLS regression residuals and an array of test statistics are reported in the OLS regression outputs. The results are listed in Table 6.6. The last four columns display the values of the standard LM test and robust LM test statistics. All these test statistics are distributed as χ^2 with one degree of freedom (Anselin, 2003). The values in the column "z-score" indicate that spatial autocorrelation is highly significantly present in the residuals of all the OLS models. Moreover, both standard LM-Lag and LM-Error test statistics are significant for all the models. So, the robust LM test statistics are used to make decisions of the spatial regression model selection. The results are highlighted in bold. For example, the spatial lag model should be chosen in the regression of TRATENEG on variable CHILDRISK because both standard LM-Lag and LM-Error test statistics are significant; and both robust LM-Lag and LM-Error test statistics are significant as well, but the robust LM-Lag statistic is more significant than the robust LM-Error statistic (67.2 vs. 13.3).

Once the order of weights and the type of the spatial regression models are determined, spatial regression is performed. To check if the models meet the requirement of giving uncorrelated residuals (the third criterion of model design described in Subsection 5.4.1), the Moran scatter plots of the residuals are drawn and displayed in Figures 6.14 - 16. The scatter plots indicate no or little spatial autocorrelation in the residuals of any of the spatial regression models.

Table 6.7 presents the results of spatial regression. The values displayed in the columns "$\hat{b}$" and "$\hat{\rho}$ or $\hat{\lambda}$" are unstandardized regression parameter estimates. The models are ranked according to the AIC values. Ranking is made separately for the two response variables.

To examine how spatial autocorrelation affects the parameter estimation of regression models, comparison is made between results from the OLS regression and from spatial regression. First, compare the parameter estimation in the category "Risk variable" in Table 6.7 with that in Table 6.4. The absolute values in all the cells of the column "z-score" in Table 6.7 are about half of those in the column "*t*-statistics" in Table 6.4. This comparison supports the idea that when spatial autocorrelation is present in OLS regression residuals, the absolute values of the test statistic are upward biased (see Subsection 5.3.1). The impact is clear on the statistical inference for the regression models with the response variable of TRATEPE. For example, TRATEPE is significantly related to MAGELT20 at $p<0.0000$ based on the OLS estimation, but the relationship is significant only at $p=0.05$ based on the spatial regression estimation. Furthermore, TRATEPE is significantly related to MEDICAID at $p<0.0000$ in the OLS regression, but the relationship is not significant at $p=0.10$ in the spatial regression.

Second, compare the "AIC" columns in Table 6.7 and Table 6.4. It is seen that inclusion of spatial autocorrelation improves the predictability, for all the AIC values in

Table 6.7 are smaller than their counterparts in Table 6.4. Besides, the ranking results based on the AIC values are changed. In the OLS regression of TRATENEG, NMARRIED ranks the best and MEDICAID the second best; but the rankings are reverse in the spatial regression. NMARRIED ranks the best in the OLS regression of TRATEPE, but is the second best tied with SIBLING3 and LVACANTHSE in the spatial regression. In the OLS regression of TRATEPE, NMARRIED ranks the best and SUBSTANCE the second best. However, SUBSTANCE ranks the best predictor in the spatial regression.

Finally, use Table 6.7 to answer the first two research questions (see Subsection 3.2.3). After accounting for the effect of spatial autocorrelation, TRATENEG is significantly, positively related to all nine explanatory variables at p<0.0000. Compared with these relationships, the relationships between TRATEPE and the explanatory variables are relatively weaker. This finding is consistent with the results from reviewing maps and investigating scatter plots. SUBSTANCE is the only variable to which TRATEPE is significantly related at p<0.0000. TRATEPE is significantly related to variables SIBLING3 and LVACANTHSE at p=0.005; to variable NMARRIED at p=0.01; to variables CHILDRISK, MAGELT20, and LUNEMPMF at p=0.05; but not significantly related to variables MEDICAID (p=0.0666) and SSINGPARCH6 (p=0.2053).

MEDICAID, NMARRIED, LUNEMPMF, and MAGELT20 rank as the top four significant explanatory variables for TRATENEG in terms of giving smaller AIC values, while SUBSTANCE, SIBLING3, LVACANTHSE, and NMARRIED rank as the top four significant explanatory variables for TRATEPE. NMARRIED is only variable among the top four predicators for TRATENEG and those for TRATEPE. MEDICAID is the most significant predicator for TRATENEG, but is not a significant predicator for TRATEPE at p=0.05.

6.2.2 Multivariate Regression

6.2.2.1 OLS Regression

Table 6.8 presents the results of multivariate variable selection. The backward selection process starts from including all nine variables, and ends when MCN<10 and all remaining variables are statistically significant. For the regression of TRATENEG, when all nine variables are included, the problem with multicollinearity is serious because the value of MCN is 35.99, greater than 30. Variable MAGELET20 is then removed because its estimated coefficient is at the opposite direction of the relationship shown by the scatter plot (Figure 6.5 (b)). The model with eight remaining variables ("Model 2" in Table 6.8) gives a value of 33.21 for MCN, indicating multicollinearity is still a serious problem. Variable SSIBLINGS3 is removed from the next regression model because it is the most insignificant variable. Variables CHILDRISK, MEDICAID, and LVACANTHSE are removed in the following steps, leaving four variables including SUBSTANCE, NMARRIED, LUNEMPMF, and SSINGPARCH6 retained in the final model ("Model 6" in Table 6.8), which gives a value of 9.83 for MCN, smaller than the threshold of 10.

For the regression of TRATEPE, the inclusion of all nine variables gives a value of 35.14 for MCN. Then variables SSINGPARCH6, MEDICAID, MAGELT20, CHILDRISK, SIBLINGS3, and LUNEMPMF are removed at six steps, leaving three variables including SUBSTANCE, NMARRIED, and LVACANTHSE retained in the final model ("Model 7" in Table 6.8), in which, MCN=7.08 is smaller than the threshold of 10. Although the value of MCN (9.73) in “Model 6” is smaller than the threshold of 10, variable LUNEMPMF is statistically insignificant, and hence removed.

The OLS regression results with remaining variables are shown in Table 6.9. Variables are ordered according to the values of “*t*-statistic”. In the regression of TRATENEG, variables NMARRIED, SUBSTANCE, and LUNEMPMF are statistically significant at $p<0.0000$, while variable SSINGPARCH6 is significant to a lower degree. In the regression of TRATEPE, variable SUBSTANCE is statistically significant at $p<0.0000$, while other two variables are statistically significant to a lower degree. Reviewing the residual maps (not shown) suggests spatial autocorrelation is present. Therefore, spatial regression must be conducted to account for the effect of spatial autocorrelation.

6.2.2.2 Spatial Regression

Table 6.10 presents the results of determining the order of spatial weights for multivariate spatial regression. The results suggest the third order of weights is right both for the regression of TRATENEG and for the regression of TRATEPE because it provides the smallest AIC values in all the regression models.

The Moran’s *I* values of the OLS regression residuals and test statistics are reported in Table 6.11. The results indicate spatial autocorrelation is highly significant in both regression models. Furthermore, the values of the Robust LM test statistics suggest that the spatial error model is appropriate for the regression of TRAGENEG, while the spatial lag model is right for the regression of TRATEPE. Therefore, spatial regression is performed with the spatial error model selected for the regression of TRATENEG and the spatial lag model chosen for the regression of TRATEPE.

To check if the models meet the requirement of giving uncorrelated results, the Moran scatter plots of the residuals are drawn and displayed in Figure 6.17. The scatter plots indicate no spatial autocorrelation in the residuals of both spatial regression models.

The spatial regression results are presented in Table 6.12. All four variables are statistically significant in the regression of TRATENEG. However, for the regression of TRATEPE on all three variables found significant in the OLS regression (Model A), SUBSTANCE is the only significantly contributing variable. The other two variables, NMARRIED and LVACANTHSE, are not significant at p=0.05. After the two insignificant variables are removed from Model A, the results are displayed in the last row (Model B) of Table 6.12. They are the same as the results from the bivariate regression model of TRATEPE on variable SUBSTANCE, which ranks the best among the bivariate spatial regression models of the response variable TRATEPE (Table 6.7). Dropping the two insignificant variables does not change the model's predictive ability (Models A and B have the same pseudo-R^2 values).

Consistent with the results of the bivariate regression analysis, the multivariate regression results support the idea that when spatial autocorrelation is present in the OLS regression residuals, the absolute values of the test statistic are upward biased. This can be seen by comparing the numbers in the column "z-score" in Table 6.12 with the numbers in the column "*t*-statistics" in Table 6.9. It is seen that the z-scores in Table 6.12 for all variables except for variable SSINGPARCH6 are smaller than the *t*-statistics in Table 6.9.

Finally, use Table 6.12 to answer the last two research questions of the present study (see Subsection 3.2.3). The set of variables that best predicts TRATENEG includes

NMARRIED, LUNEMPMF, SUBSTANCE, and SSINGPARCH6. All four variables are statistically significantly contributive. The value of pseudo- R^2 (0.646) suggests the model has moderate predictive ability.

The set of variables that best predicts TRATEPE includes SUBSTANCE, LVACANTHSE, and NMARRIED; but the latter two variables are not statistically significantly contributive at p=0.05. The value of pseudo- R^2 is 0.300, suggesting the model has relatively low predictive ability.

Variable SUBSTANCE is the only variable that is significantly contributive both in the regression of TRATENEG and in the regression of TRATEPE. Variable NMARRIED is retained both in the final model of TRATENEG and in that of TRATEPE; however, it is significantly contributive in the former at p<0.0000 but not significantly contributive in the latter at p=0.05. Therefore, the results suggest that the combination of risk variables that best predicts the rates of substantiated maltreatment may differ by type of maltreatment.

Table 6.1 Descriptive statistics of substantiated child maltreatment

Measures	**Min (# of tracts)**	**Max (# of tracts)**	**Sum**	**Mean**	**Standard Deviation**	**# of tracts having at least one victim**
Number of substantiated neglect	0 (73)	67 (1)	3,526	7.4	8.7	405
Number of substantiated physical/emotional abuse	0 (311)	9 (1)	313	0.7	1.2	167
Number of weighted counts of children under the age of four	29.1 (1)	6427.6 (1)	563,661	1179.2	766.9	N/A[a]
Rates of substantiated neglect	0.0	86.1	N/A	7.6	9.9	N/A
Rates of substantiated physical/emotional abuse	0.0	14.3	N/A	0.6	1.3	N/A
Transformed rates of substantiated neglect	-0.6804	4.6352	N/A	1.7043	0.9923	N/A
Transformed rates of substantiated physical /emotional abuse	-1.1757	3.5366	N/A	0.3897	0.7333	N/A

[a] N/A: Not Applicable

Table 6.2 Explanatory variables examined in regression analyses

SUBSTANTIATED NEGLECT		SUBSTANTIATED PHYSICAL/EMOTIONAL ABUSE	
Response Variable	**Explanatory Variable**	**Response Variable**	**Explanatory Variable**
TRATENEG	CHILDRISK MAGELT20 SUBSTANCE MEDICAID SSIBLING3 NMARRIED LUNEMPMF SSINGPARCH6 LVACANTHSE	TRATEPE	CHILDRISK MAGELT20 SUBSTANCE MEDICAID SIBLING3 NMARRIED LUNEMPMF SSINGPARCH6 LVACANTHSE

Table 6.3 Descriptive statistics of the explanatory variables

Risk Variables	Mean	Standard Deviation	Min	Max	Range
CHILDRISK [a]	14.17	3.86	3.61	27.59	23.98
MAGELT20 [a]	11.38	8.33	0.00	35.09	35.09
SUBSTANCE [a]	5.92	3.00	0.57	17.75	17.19
MEDICAID [a]	37.83	21.95	0.95	82.46	81.50
SIBLING3 [a]	9.76	5.71	0.73	35.29	34.56
SSIBLING3	3.00	0.86	0.86	5.94	5.08
NMARRIED [a]	38.48	26.36	1.15	94.95	93.80
LUNEMPMF	1.59	0.82	-1.61	4.50	6.11
SSINGPARCH6	1.33	0.78	0.00	4.46	4.46
LVACANTHSE	1.54	0.68	-0.14	4.05	4.20

[a] Values presented as percentage

Table 6.4 Bivariate OLS regression results

Response Variable	Explanatory Variable	OLS Regression				AIC	
		$\hat{b}$	Std. Error	*t*-statistic	R^2	Value	Ranking
TRATENEG	CHILDRISK	0.16887	0.00890	18.97	0.431	1083.9	5
	MAGELT20	0.08372	0.00389	21.53	0.493	1028.2	3
	SUBSTANCE	0.14945	0.01351	11.06	0.204	1243.8	9
	MEDICAID	0.03344	0.00140	23.91	0.546	976.0	2
	SSIBLINGS3	0.75197	0.04018	18.71	0.424	1089.6	6
	NMARRIED	0.02822	0.00114	24.65	0.561	959.9	1
	LUNEMPMF	0.80521	0.04148	19.41	0.442	1074.4	4
	SSINGPARCH6	0.67480	0.04964	13.59	0.280	1196.3	7
	LVACANTHSE	0.70762	0.05902	11.99	0.232	1227.0	8
TRATEPE	CHILDRISK	0.06339	0.00822	7.71	0.111	1007.6	5
	MAGELT20	0.02865	0.00381	7.51	0.106	1010.5	5-tied
	SUBSTANCE	0.08584	0.01048	8.19	0.123	1001.0	2-tied
	MEDICAID	0.01031	0.00146	7.07[a]	0.095	1016.2	8
	SIBLINGS3	0.04615	0.00550	8.38	0.129	998.1	2
	NMARRIED	0.01037	0.00118	8.76	0.139	992.6	1
	LUNEMPMF	0.28681	0.03887	7.38	0.103	1012.2	5-tied
	SSINGPARCH6	0.08430	0.04305	1.958[b]	0.008	1060.1	9
	LVACANTHSE	0.38262	0.04657	8.21	0.124	1000.6	2-tied

[a] *p*-value (t=7.07) < 0.0000, two-sided; [b] *p*-value (t=1.958) = 0.0508, two-sided

Table 6.5 Determining the order of spatial weights for bivariate regression

Response Variable	**Explanatory Variable**	**Order of Spatial Weights**				
		AIC				
		1st	2nd	3rd	4th	5th
TRATENEG	CHILDRISK	986.4	**969.8**	982.7	987.0	994.1
	MAGELT20	956.1	**942.1**	950.4	954.5	960.2
	SUBSTANCE	1018.5	***992.7***	*990.2*	1002.8	1015.9
	MEDICAID	914.3	904.9	***891.9***	*891.4*	894.3
	SSIBLING3	970.6	**948.3**	961.0	961.7	965.7
	NMARRIED	915.9	912.1	**902.4**	905.6	909.7
	LUNEMPMF	952.5	**937.1**	948.6	948.8	952.4
	SSINGPARCH6	994.7	**963.7**	968.1	977.2	989.7
	LVACANTHSE	1026.0	***997.8***	*997.7*	1007.1	1025.0
TRATEPE	CHILDRISK	951.2	929.0	**921.7**	932.9	943.4
	MAGELT20	951.6	930.1	**923.2**	934.3	944.9
	SUBSTANCE	940.2	916.4	**908.0**	916.8	924.1
	MEDICAID	954.0	930.9	**923.6**	934.4	946.0
	SIBLING3	944.5	922.0	**916.8**	927.3	936.6
	NMARRIED	943.1	924.7	**919.3**	932.5	942.9
	LUNEMPMF	952.9	930.4	**922.7**	933.1	943.0
	SSINGPARCH6	969.1	936.3	**925.0**	936.1	947.6
	LVACANTHSE	944.8	924.4	**919.3**	929.6	939.4

Table 6.6 Test statistics of bivariate OLS regression residuals

Response Variable	Explanatory Variables	Order of *W*	Moran's *I* of Residuals				
			z-score	LM Test Statistics [a]			
				LM-Lag	LM-Error	Robust LM-Lag	Robust LM-Error
TRATENEG	CHILDRISK	2nd	12.48	193.7	139.7	**67.2**	13.3
	MAGELT20	2nd	11.11	142.7	110.1	**48.0**	15.4
	SUBSTANCE	2nd	28.72	742.7	765.3	66.9	**89.5**
	MEDICAID	3rd	16.54	149.3	226.8	41.5	**119.1**
	SSIBLINGS3	2nd	16.21	271.9	238.7	**73.0**	39.9
	NMARRIED	3rd	14.03	87.3	160.4	17.0	**90.1**
	LUNEMPMF	2nd	14.01	249.7	177.8	**92.9**	21.0
	SSINGPARCH6	2nd	23.84	619.7	528.3	**131.1**	39.6
	LVACANTHSE	2nd	26.30	610.4	637.4	62.8	**700.3**
TRATEPE	CHILDRISK	3rd	16.38	235.2	221.9	**28.3**	15.0
	MAGELT20	3rd	16.10	239.9	213.8	**34.9**	8.7
	SUBSTANCE	3rd	17.46	294.3	261.0	**46.3**	13.0
	MEDICAID	3rd	17.04	264.6	240.9	**34.7**	11.0
	SIBLINGS3	3rd	15.23	216.0	193.3	**35.4**	12.6
	NMARRIED	3rd	14.93	183.4	182.2	**19.3**	18.1
	LUNEMPMF	3rd	15.17	245.8	192.4	**55.8**	2.5
	SSINGPARCH6	3rd	25.18	591.7	551.7	**50.2**	10.2
	LVACANTHSE	3rd	13.17	207.9	144.5	**64.1**	0.7

[a] All LM test statistics are distributed as χ^2 with one degree of freedom; LM stands for Lagrange

Table 6.7 Bivariate spatial regression analysis results

Response Variable	**Explanatory Variable**	**SRM**	**Risk Variable**				**Spatial Autocorrelation**			**Pseudo $-R^2$**	**AIC**	
			$\hat{b}$	**Std. Error**	***z*-score**	***p*-value**	$\hat{\rho}$ **or** $\hat{\lambda}$	**Std. Error**	***z*-score**		**Value**	**Ranking**
TRATENEG	CHILDRISK	L	0.0761	0.0106	7.18	<0.0000	0.6557	0.0565	11.60	0.567	969.8	7
	MAGELT20	L	0.0452	0.0053	8.58	<0.0000	0.5734	0.0597	9.61	0.588	942.1	4
	SUBSTANCE	E	0.0679	0.0129	5.27	<0.0000	0.8635	0.0412	20.95	0.561	992.7	8
	MEDICAID	E	0.0244	0.0019	12.61	<0.0000	0.8306	0.0641	12.96	0.630	891.9	1
	SSIBLINGS3	L	0.3736	0.0441	8.48	<0.0000	0.6615	0.0528	12.53	0.587	948.3	5
	NMARRIED	E	0.0233	0.0018	12.67	<0.0000	0.7667	0.0795	9.64	0.620	902.4	2
	LUNEMPMF	L	0.4193	0.0449	9.33	<0.0000	0.6447	0.0516	12.49	0.595	937.1	3
	SSINGPARCH6	L	0.3232	0.0408	7.92	<0.0000	0.7689	0.0456	16.87	0.580	963.7	6
	LVACANTHSE	E	0.2990	0.0630	4.75	<0.0000	0.8616	0.0416	20.72	0.556	997.8	9
TRATEPE	CHILDRISK	L	0.0182	0.0079	2.31	0.0209	0.8131	0.0653	12.45	0.281	921.7	5
	MAGELT20	L	0.0071	0.0036	**1.97**	0.0493	0.8226	0.0639	12.88	0.280	923.2	6-tied
	SUBSTANCE	L	0.0432	0.0098	4.41	<0.0000	0.7866	0.0674	11.67	0.300	908.0	1
	MEDICAID	L	0.0025	0.0014	**1.83**	0.0666	0.8297	0.0623	13.32	0.280	923.6	N/S[a]
	SIBLINGS3	L	0.0173	0.0054	3.24	0.0012	0.7874	0.0684	11.51	0.287	916.8	2
	NMARRIED	L	0.0034	0.0012	2.76	0.0059	0.7860	0.0700	11.23	0.283	919.3	2-tied
	LUNEMPMF	L	0.0762	0.0358	2.13	0.0333	0.8216	0.0638	12.88	0.280	922.7	6
	SSINGPARCH6	L	-0.0464	0.0366	**-1.26**	0.2053	0.8888	0.0476	18.68	0.283	925.0	N/S[a]
	LVACANTHSE	L	0.1278	0.0440	2.90	0.0037	0.7964	0.0678	11.75	0.284	919.3	2-tied

[a] Not significant at the 0.05 level, two-sided test

Table 6.8 Multivariate OLS regression variable selection

Response Variable	Explanatory Variable	*t*-statistics						
		Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**
TRATENEG	CHILDRISK	0.59	0.50	**0.53**				
	MAGELT20	**-1.29**						
	SUBSTANCE	4.39	4.20	4.39	4.45	4.57	<u>5.06</u>	
	MEDICAID	1.71	1.45	1.51	**1.44**			
	SSIBLINGS3	0.42	**0.23**					
	NMARRIED	2.65	2.33	2.45	3.41	7.91	<u>8.74</u>	
	LUNEMPMF	3.61	3.67	3.72	3.76	4.06	<u>4.15</u>	
	SSINGPARCH6	2.30	2.45	2.44	2.41	2.83	<u>2.72</u>	
	LVACANTHSE	1.47	1.57	1.56	1.54	**1.54**		
	MCN[a]	35.99	33.21	23.27	21.39	11.31	9.83	
TRATEPE	CHILDRISK	-0.80	-0.57	0.19	**0.07**			
	MAGELT20	-2.72	-2.33	**-3.18**				
	SUBSTANCE	4.44	4.57	4.57	3.76	3.77	4.20	<u>4.23</u>
	MEDICAID	-2.58	**-3.45**					
	SIBLINGS3	0.64	1.04	1.13	0.69	**0.70**		
	NMARRIED	4.95	4.41	2.87	1.03	1.34	2.25	<u>3.88</u>
	LUNEMPMF	1.59	1.41	0.73	0.74	0.74	**0.82**	
	SSINGPARCH6	**-3.40**						
	LVACANTHSE	2.59	2.87	2.83	3.09	3.09	3.05	<u>3.10</u>
	MCN	35.14	33.28	27.00	25.08	12.18	9.73	7.08

[a] MCN stands for Multicollinearity Condition Number

Table 6.9 Multivariate OLS regression results

Response Variable	**Explanatory Variables**	**OLS Regression**				R^2	**AIC**
		$\hat{b}$	**Std. Error**	***t*-statistic**	***p*-value**		
TRATENEG	NMARRIED	0.0175	0.0020	8.74	<0.0000	0.603	917.2
	SUBSTANCE	0.0543	0.0107	5.05	<0.0000		
	LUNEMPMF	0.2315	0.0558	4.15	<0.0000		
	SSINGPARCH6	0.1325	0.0486	2.72	0.0067		
TRATEPE	SUBSTANCE	0.0485	0.0115	4.23	<0.0000	0.200	961.6
	NMARRIED	0.1715	0.0553	3.88	0.0001		
	LVACANTHSE	0.0055	0.0014	3.10	0.0021		

Table 6.10 Determining the order of spatial weights for multivariate regression

Response Variable	Explanatory Variable	Order of Spatial Weights AIC				
		1st	2nd	3rd	4th	5th
TRATENEG	SUBSTANCE NMARRIED LUNEMPMF SSINGPARCH6	878.8	876.3	**873.1**	874.3	873.9
TRATEPE	SUBSTANCE NMARRIED LVACANTHSE	927.2	913.0	**908.9**	918.0	925.3

Table 6.11 Test statistics of multivariate OLS regression residuals

Response Variable	**Explanatory Variables**	**Order of *W***	**Moran's *I* of Residuals**				
			***z*-score**	**LM Test Statistics [a]**			
				LM-Lag	**LM-Error**	**Robust LM-Lag**	**Robust LM-Error**
TRATENEG	SUBSTANCE NMARRIED LUNEMPMF SSINGPARCH6	3rd	11.63	75.3	103.8	23.2	**51.7**
TRATEPE	SUBSTANCE NMARRIED LVACANTHSE	3rd	9.52	103.0	66.2	**37.7**	0.8

[a] All LM test statistics are distributed as χ^2 with one degree of freedom; LM stands for Lagrange Multiplier

Table 6.12 Multivariate spatial regression analyses

Response Variable	Explanatory Variables	SRM	Risk Variable				Spatial Autocorrelation			Pseudo $-R^2$	AIC
			$\hat{b}$	Std. Error	z-score	p-value	$\hat{\rho}$ or $\hat{\lambda}$	Std. Error	z-score		
TRATENEG	NMARRIED	SEM	0.0138	0.0024	5.76	<0.0000	0.7334	0.0869	8.44	0.646	873.1
	LUNEMPMF		0.2095	0.0536	3.91	0.0001					
	SUBSTANCE		0.0436	0.0115	3.78	0.0002					
	SSINGPARCH6		0.1303	0.0467	2.79	0.0052					
TRATEPE (Model A)	SUBSTANCE	SLM	0.0365	0.0108	3.38	0.0007	0.7168	0.0791	9.07	0.300	908.9
	LVACANTHSE		0.0542[a]	0.0519	1.04	0.2962					
	NMARRIED		0.0014[a]	0.0014	0.99	0.3212					
TRATEPE (Model B)	SUBSTANCE	SLM	0.04321	0.00979	4.41	<0.0000	0.7865	0.0674	11.67	0.300	908.0

[a] Not significant at the 0.05 level

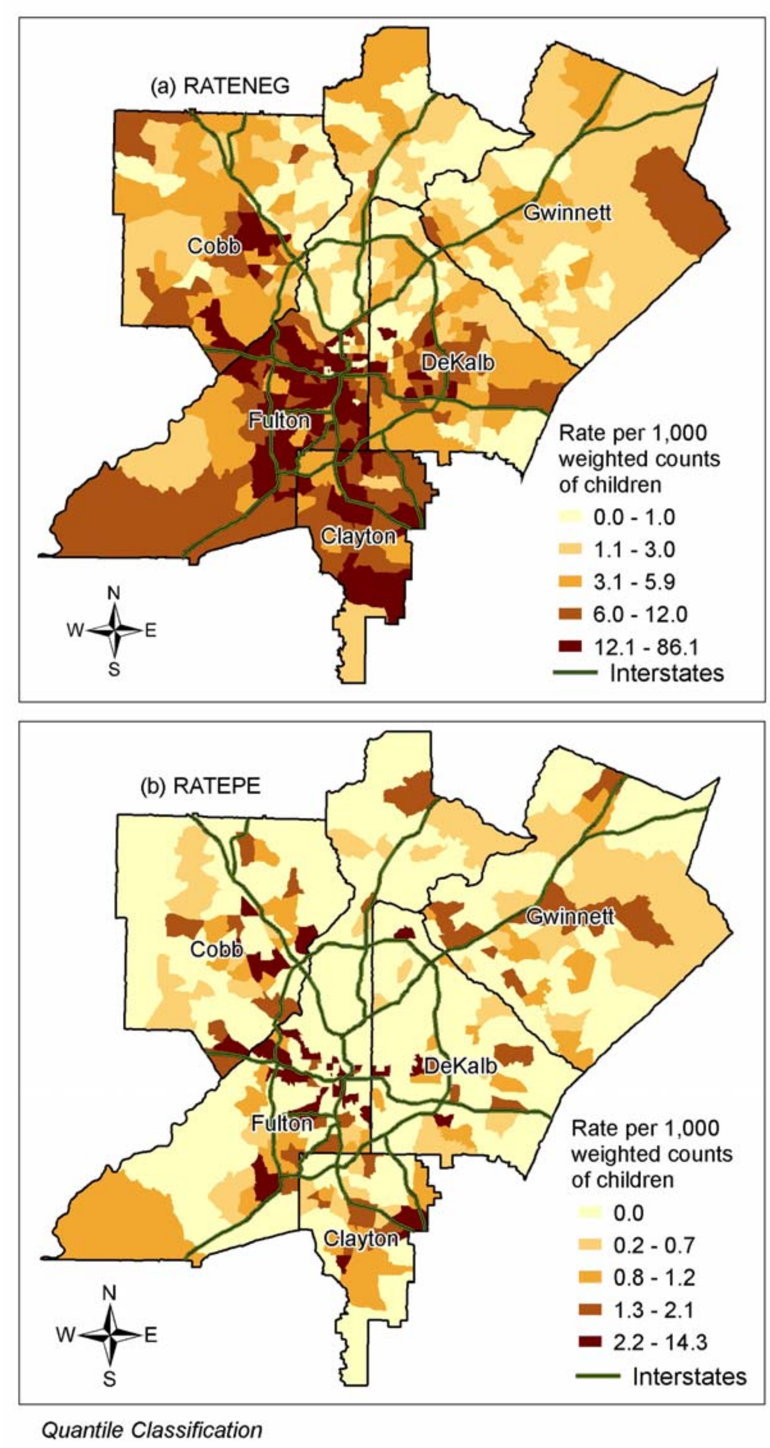


Figure 6.1 Distribution of rates of substantiated neglect, and substantiated physical/emotional abuse, by census tract: (a) rates of substantiated neglect; (b) rates of substantiated physical/emotional abuse.

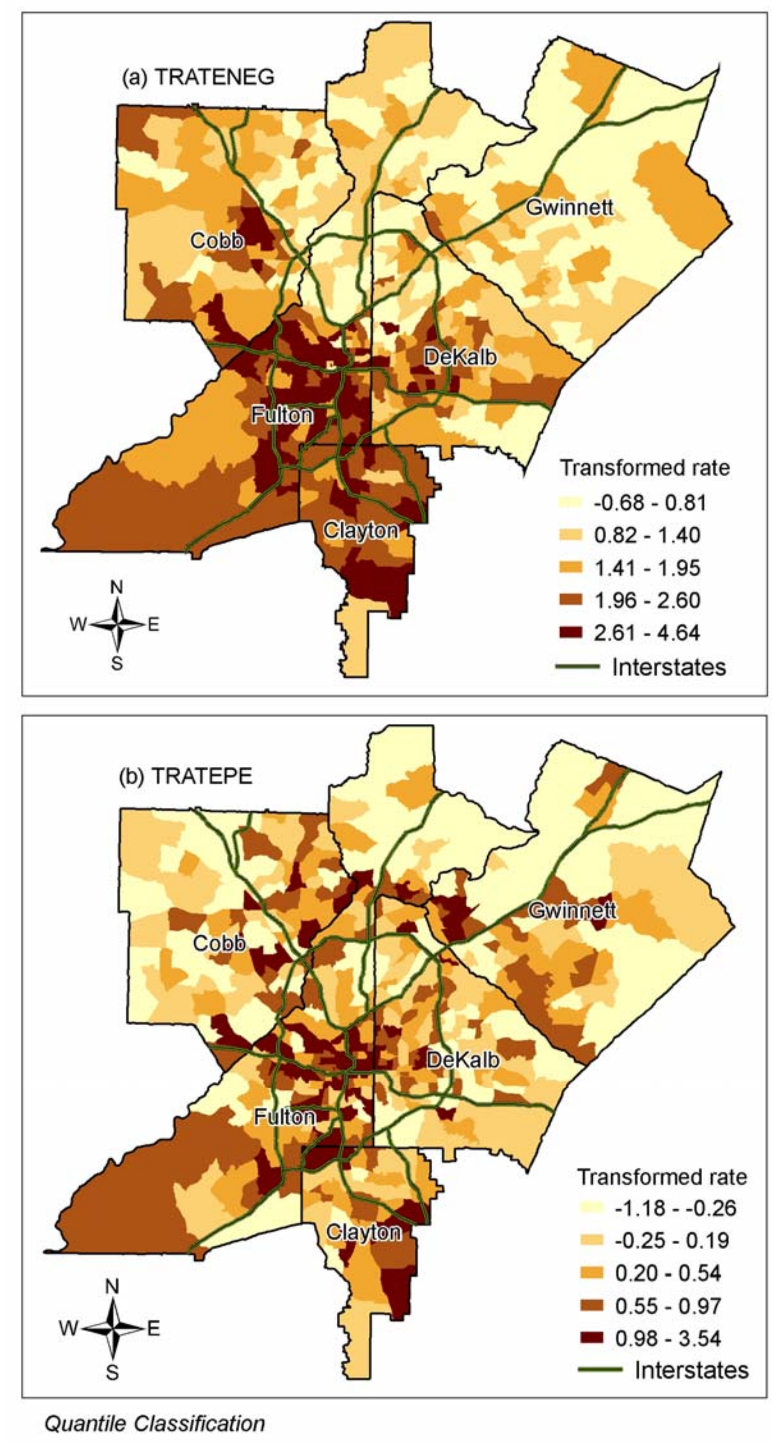


Figure 6.2 Distribution of transformed rates of substantiated neglect, and substantiated physical/emotional abuse, by census tract: (a) transformed rates of substantiated neglect; (b) transformed rates of substantiated physical/emotional abuse.

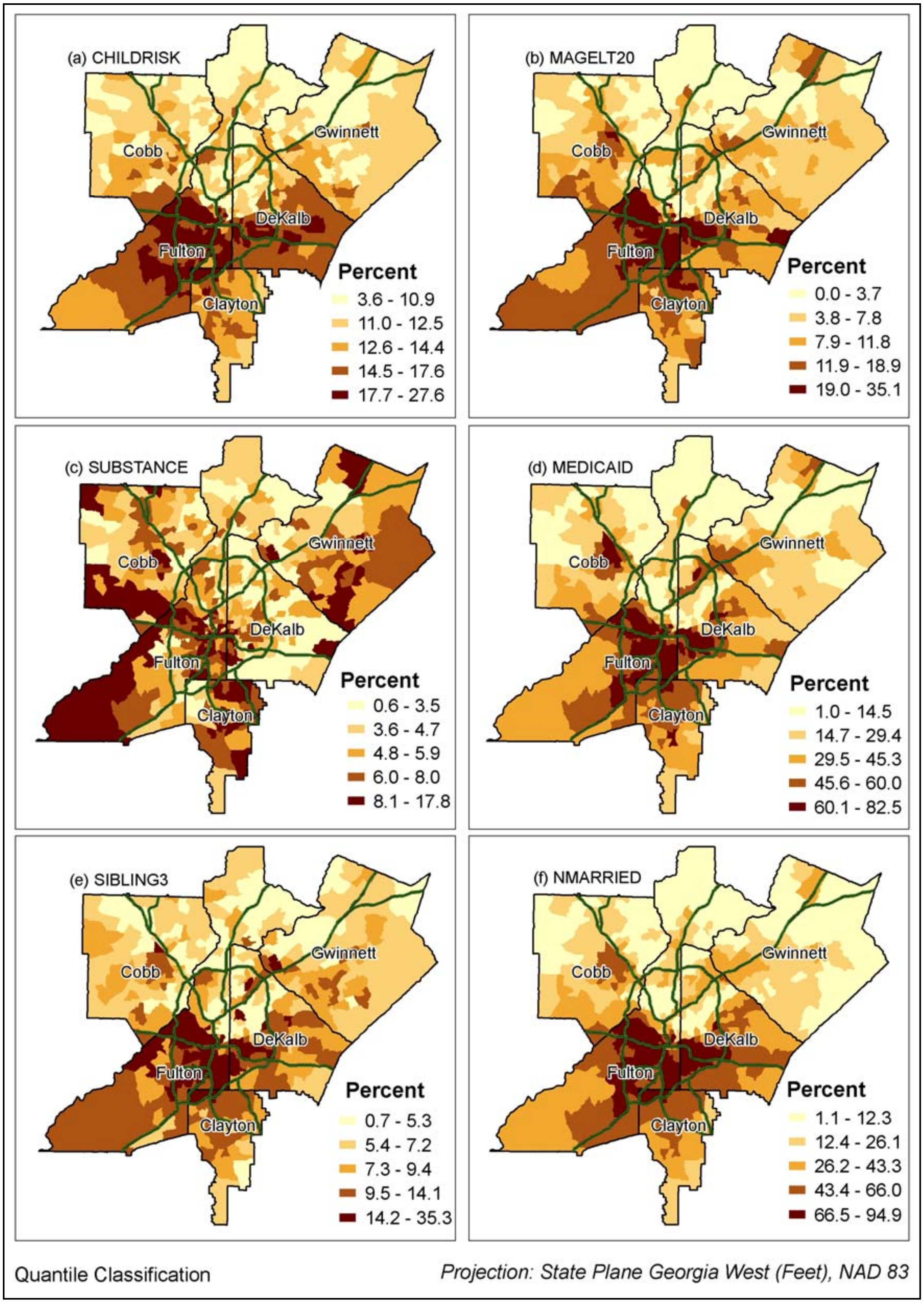


Figure 6.3 Distribution of the microsystem risk variables by census tract.

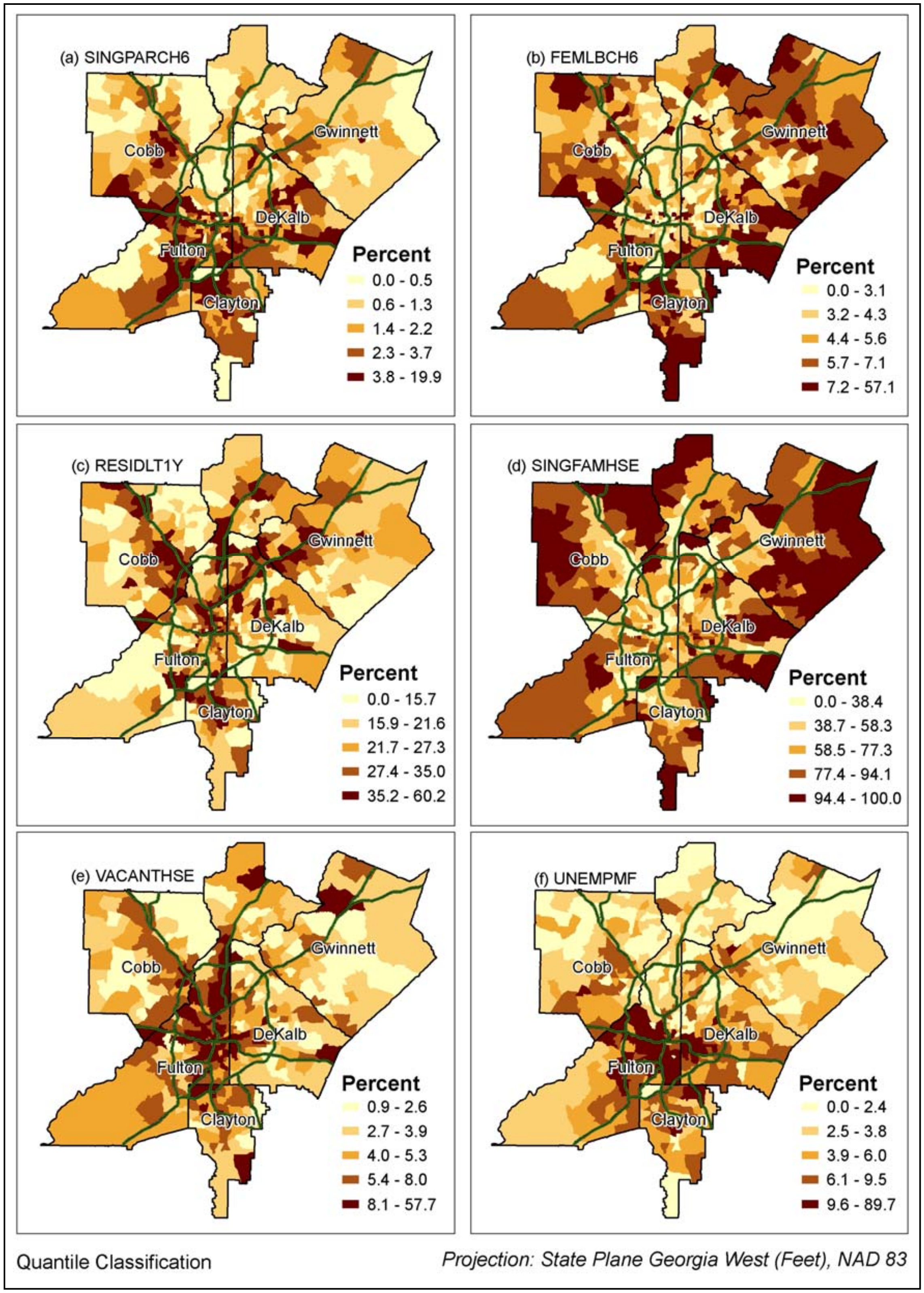
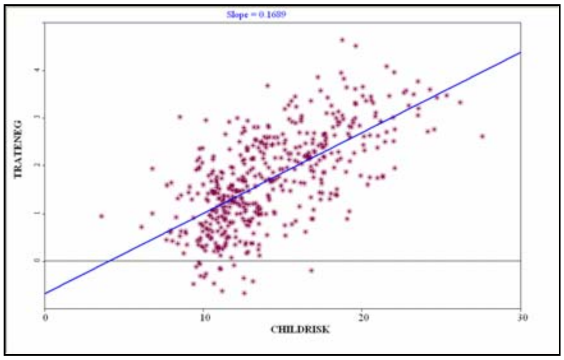
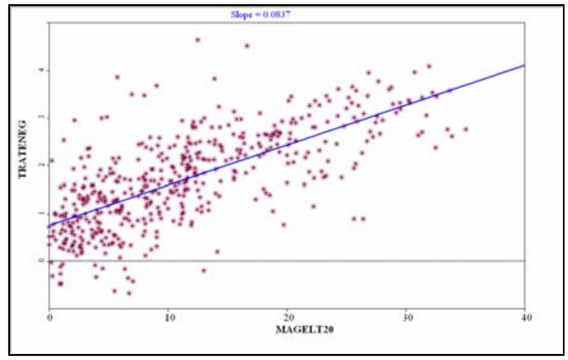
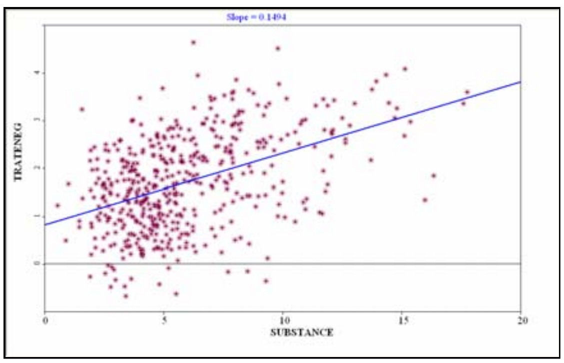


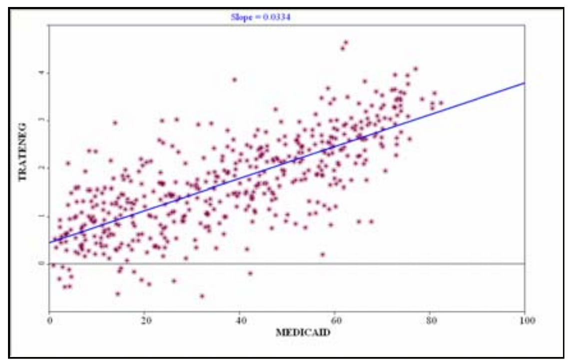
Figure 6.4 Distribution of the exosystem risk variables by census tract.



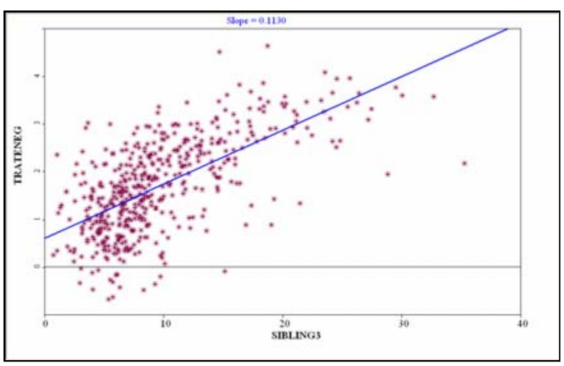


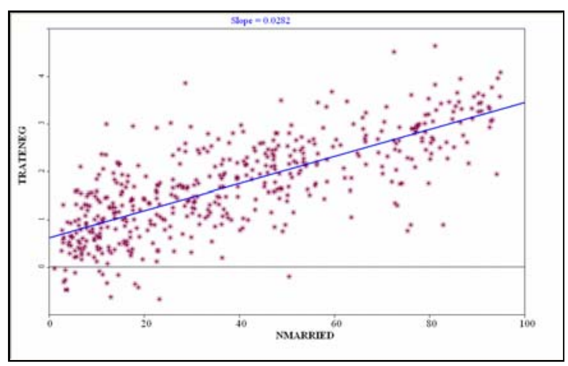
(a) (b)





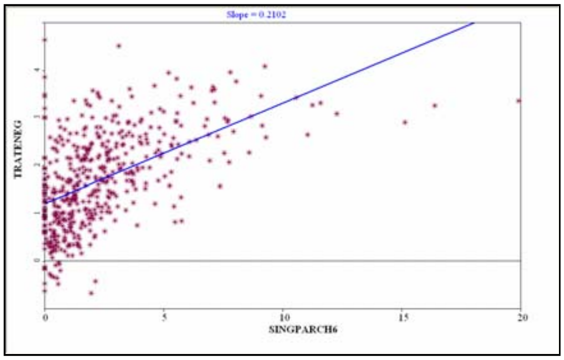
(c) (d)

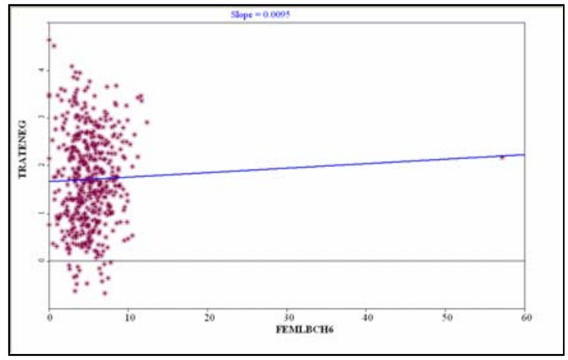




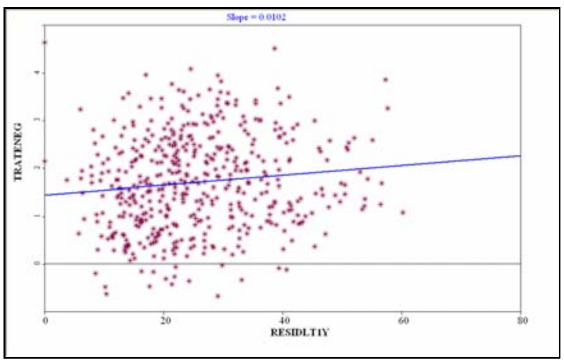
(e) (f)

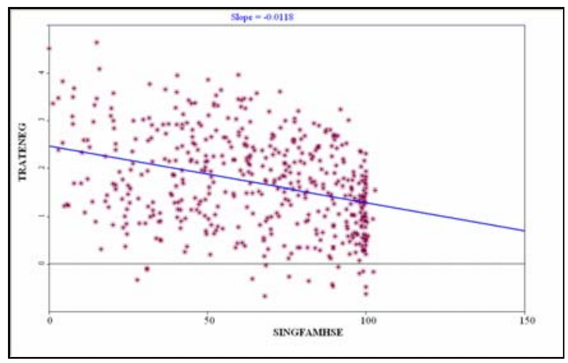
Figure 6.5 Scatterplots of the transformed rates of substantiated neglect (TRATENEG) with the microsystem variables.



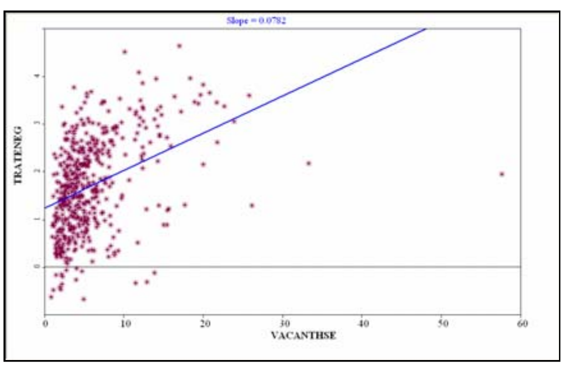


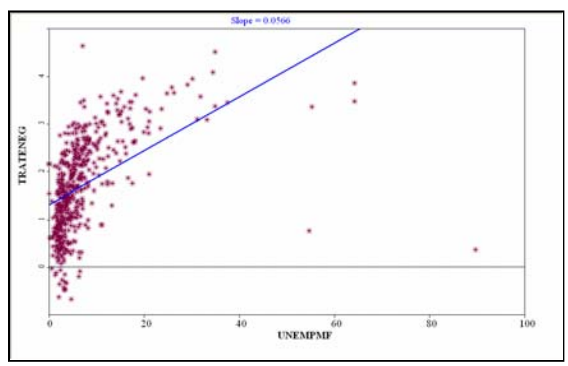
(a) (b)





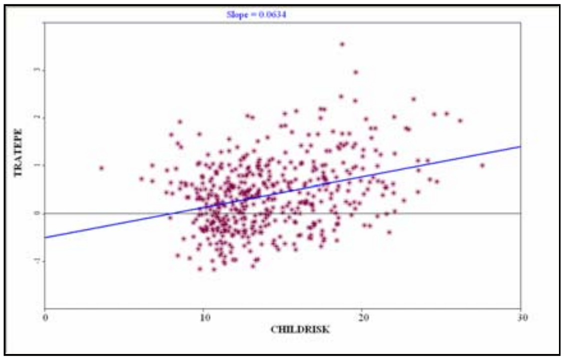
(c) (d)



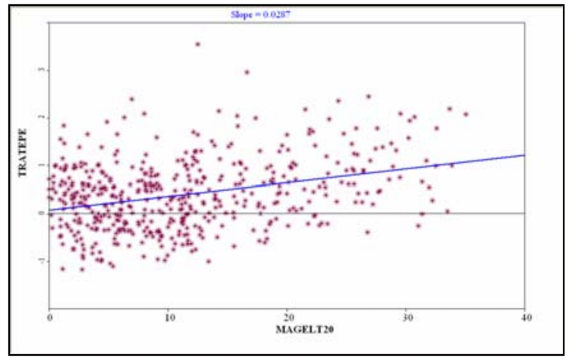


(e) (f)

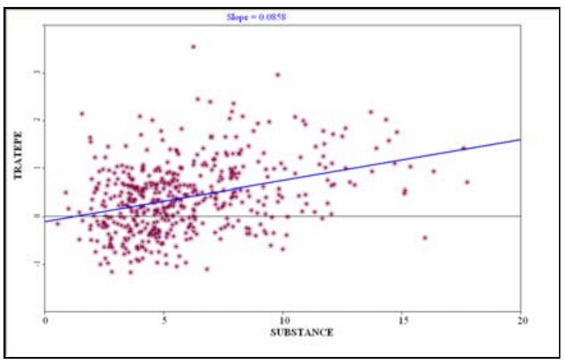
Figure 6.6 Scatterplots of the transformed rates of substantiated neglect (TRATENEG) with the exosystem variables.



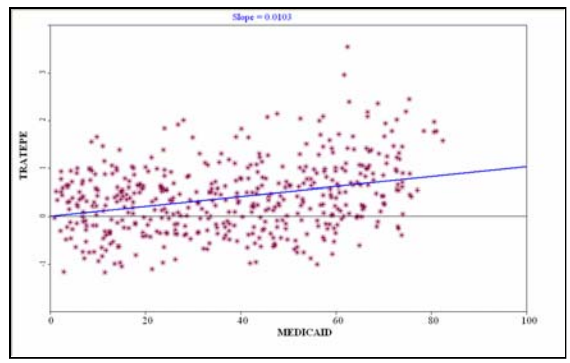
(a)



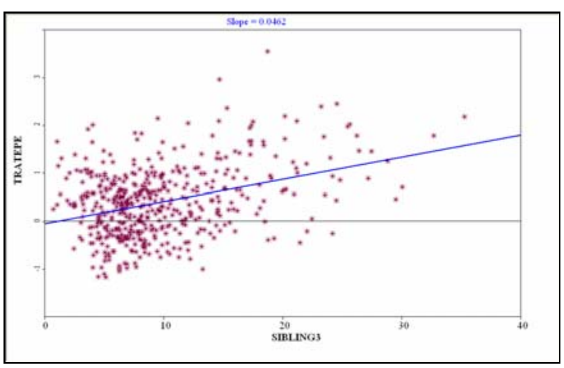
(b)



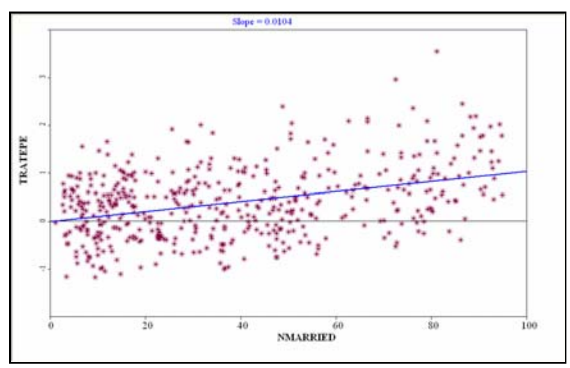
(c)



(d)



(e)



(f)

Figure 6.7 Scatterplots of the transformed rates of substantiated physical/emotional abuse (TRATEPE) with the microsystem variables.

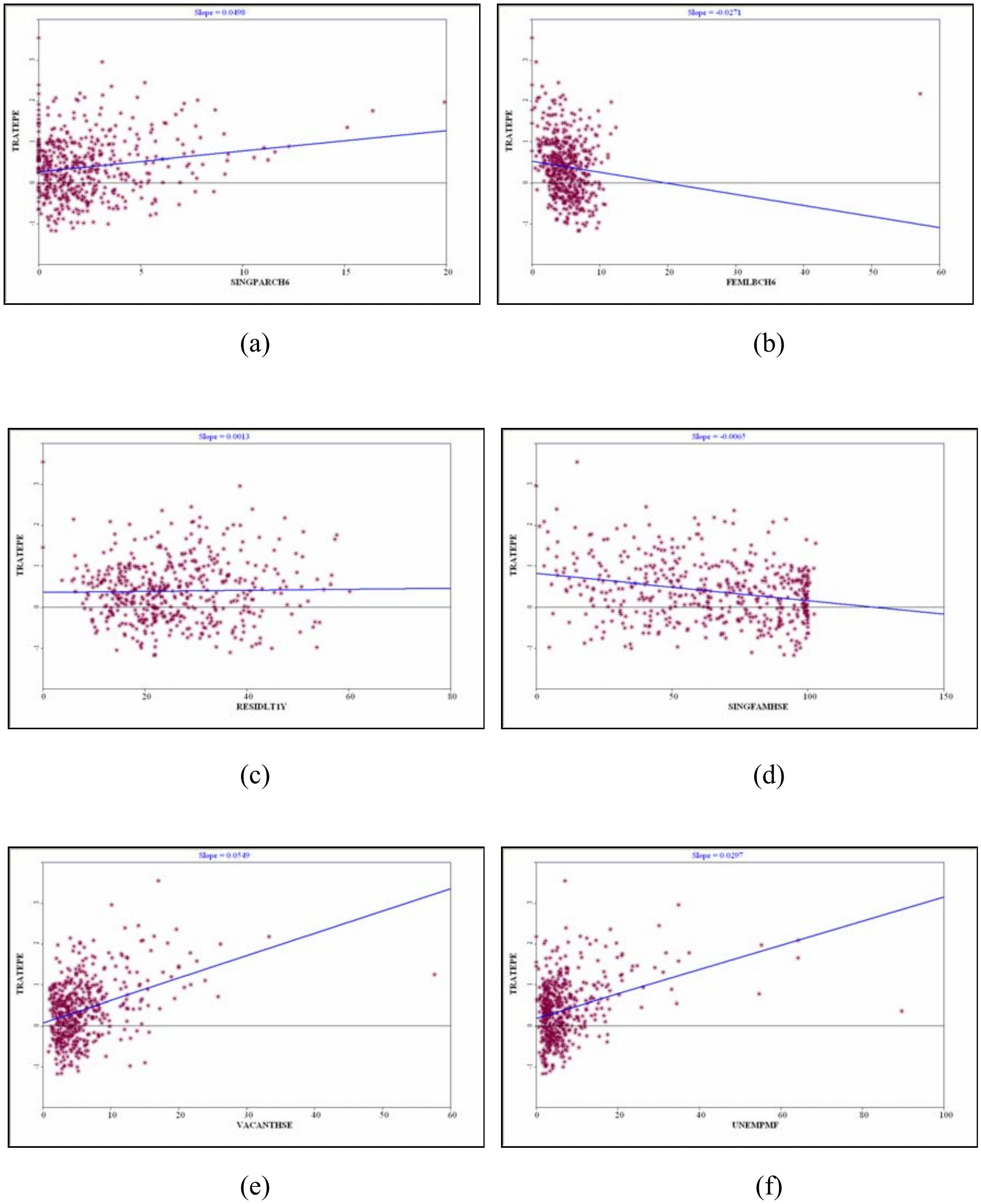


Figure 6.8 Scatterplots of the transformed rates of substantiated physical/emotional abuse (TRATEPE) with the exosystem variables

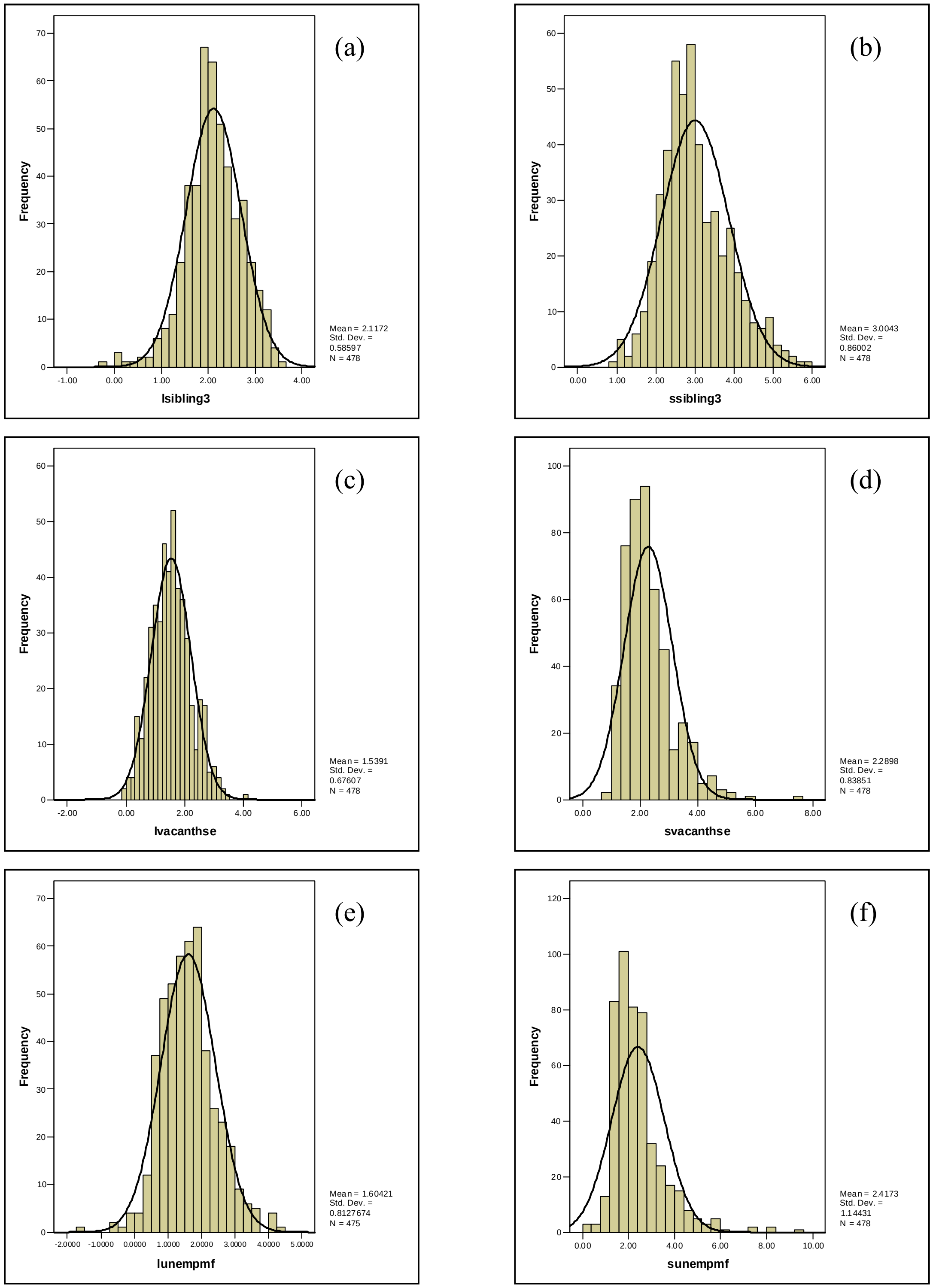


Figure 6.9 Histograms of the transformed risk variables: (a), (c) and (e) corresponding to the transformed variables by the natural logarithmic transformation; (b), (d) and (f) to the transformed variables by the square root transformation.

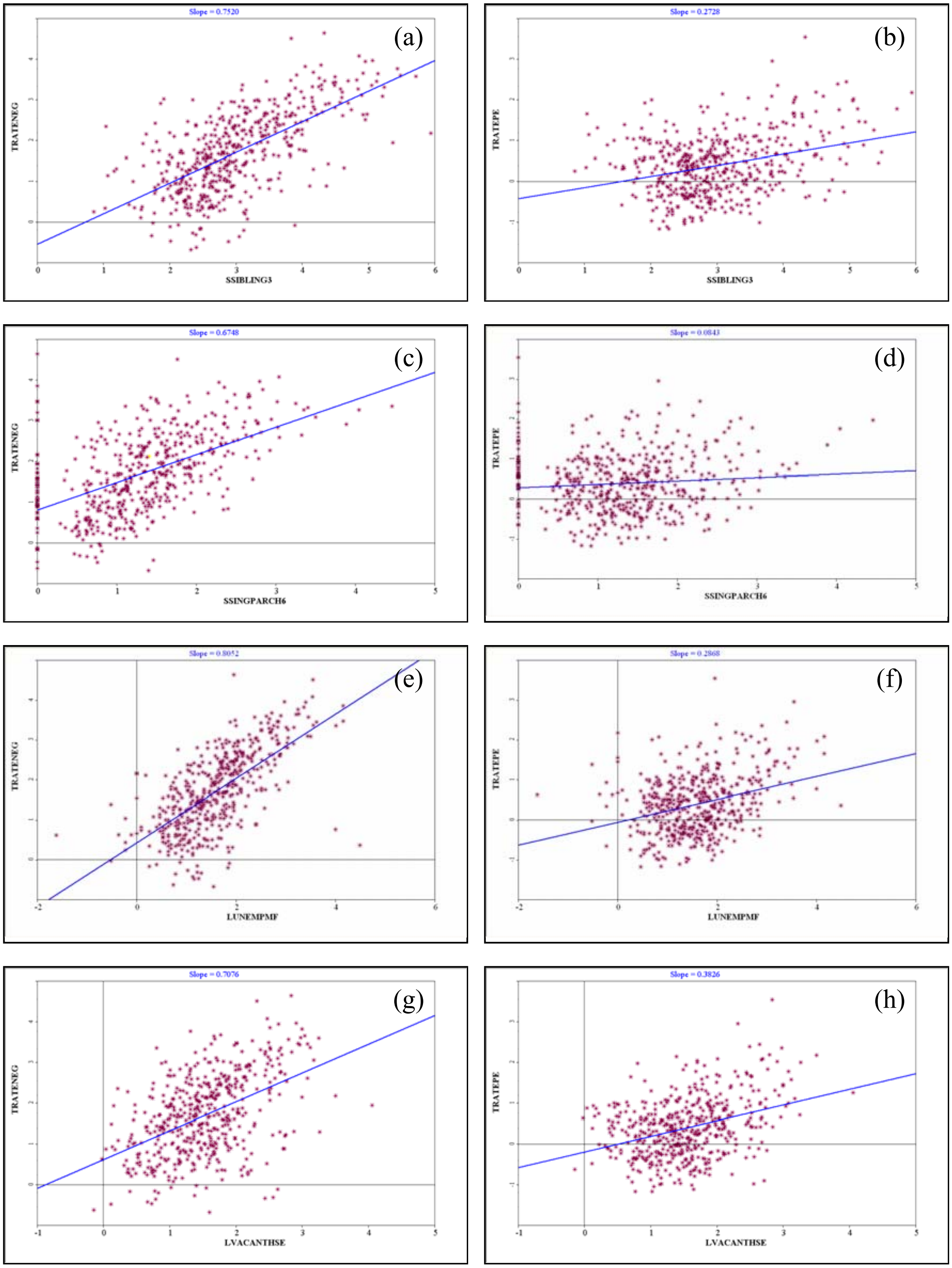


Figure 6.10 Scatterplots of the transformed rates of substantiated neglect (TRATENEG) and of substantiated physical/emotional abuse (TRATEPE) with the transformed risk variables.

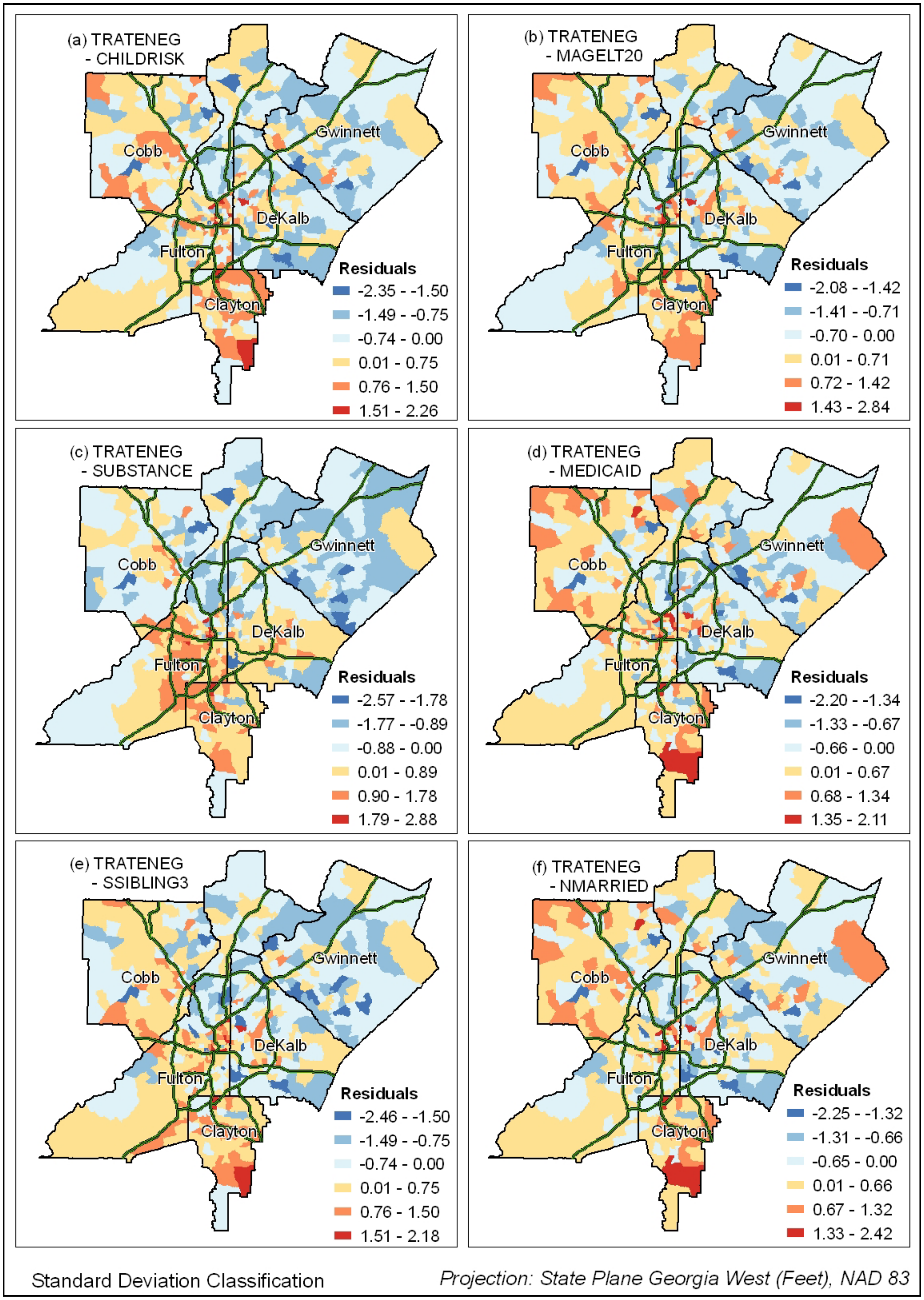


Figure 6.11 Residual maps of the bivariate OLS regression: the transformed rates of substantiated neglect (TRATENEG) on the microsystem risk variables.

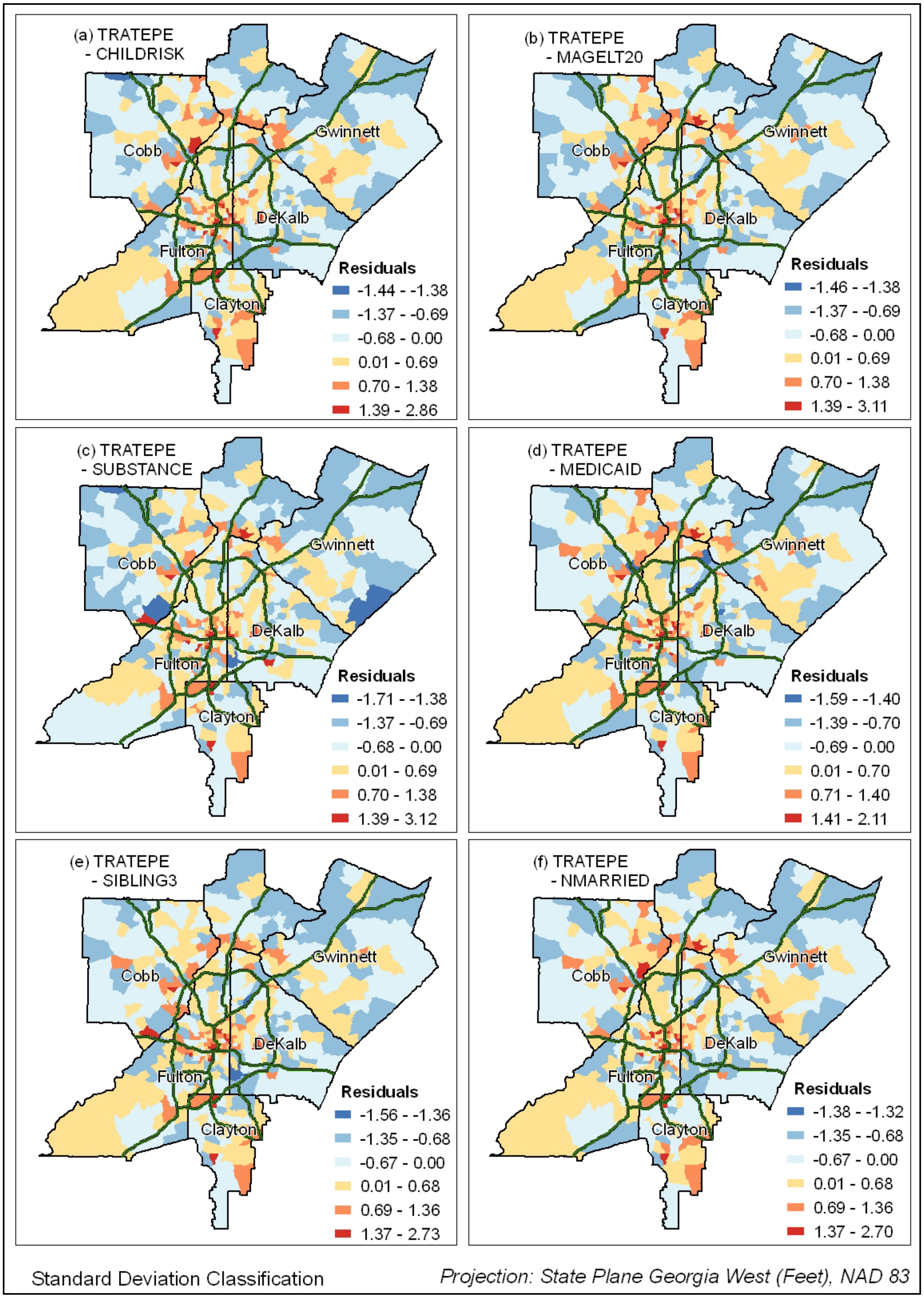


Figure 6.12 Residual maps of the bivariate OLS regression: the transformed rates of substantiated physical/emotional abuse (TRATEPE) on the microsystem risk variables.

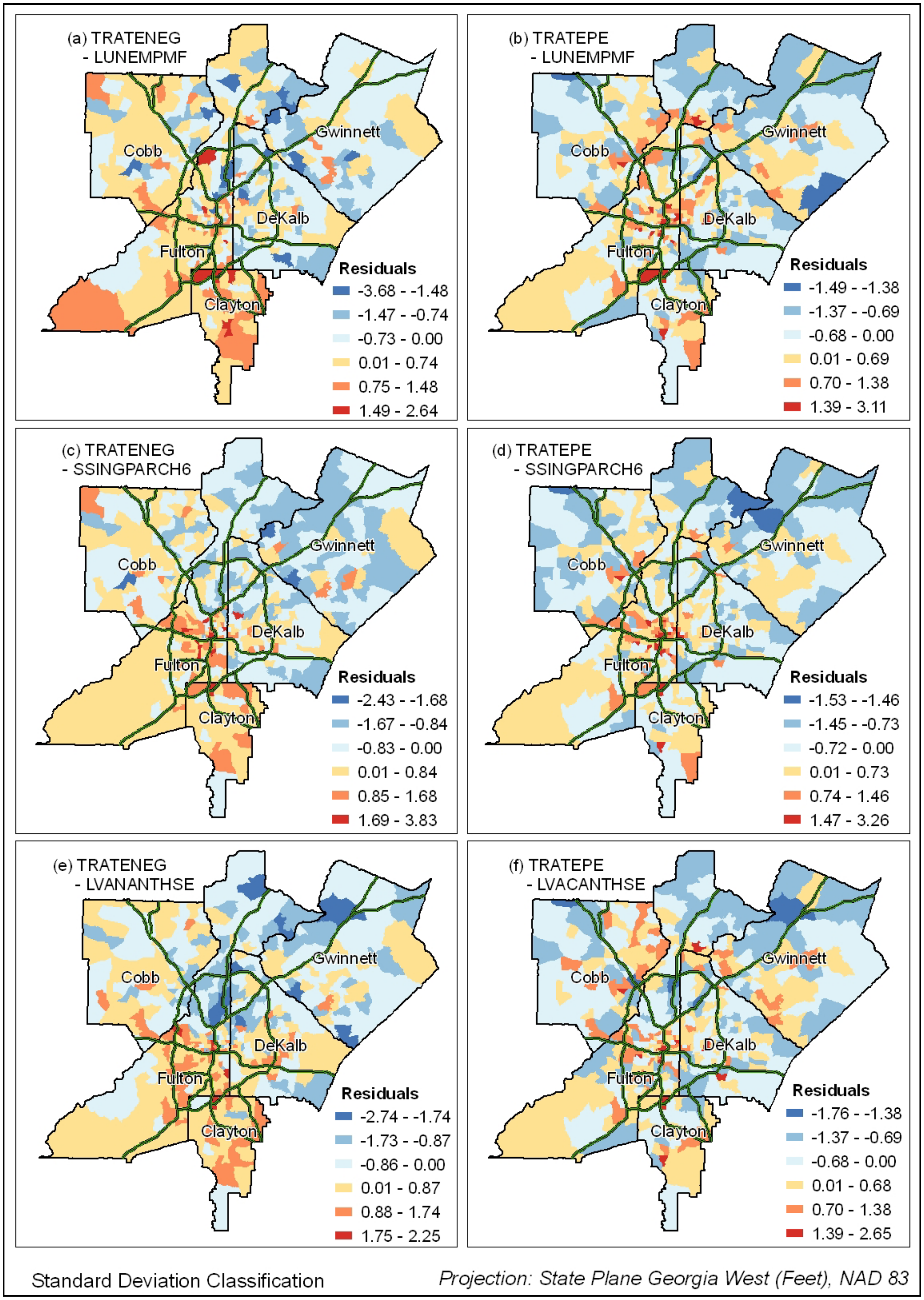


Figure 6.13 Residual maps of the bivariate OLS regression: TRATENEG and TRATEPE on the exosystem risk variables

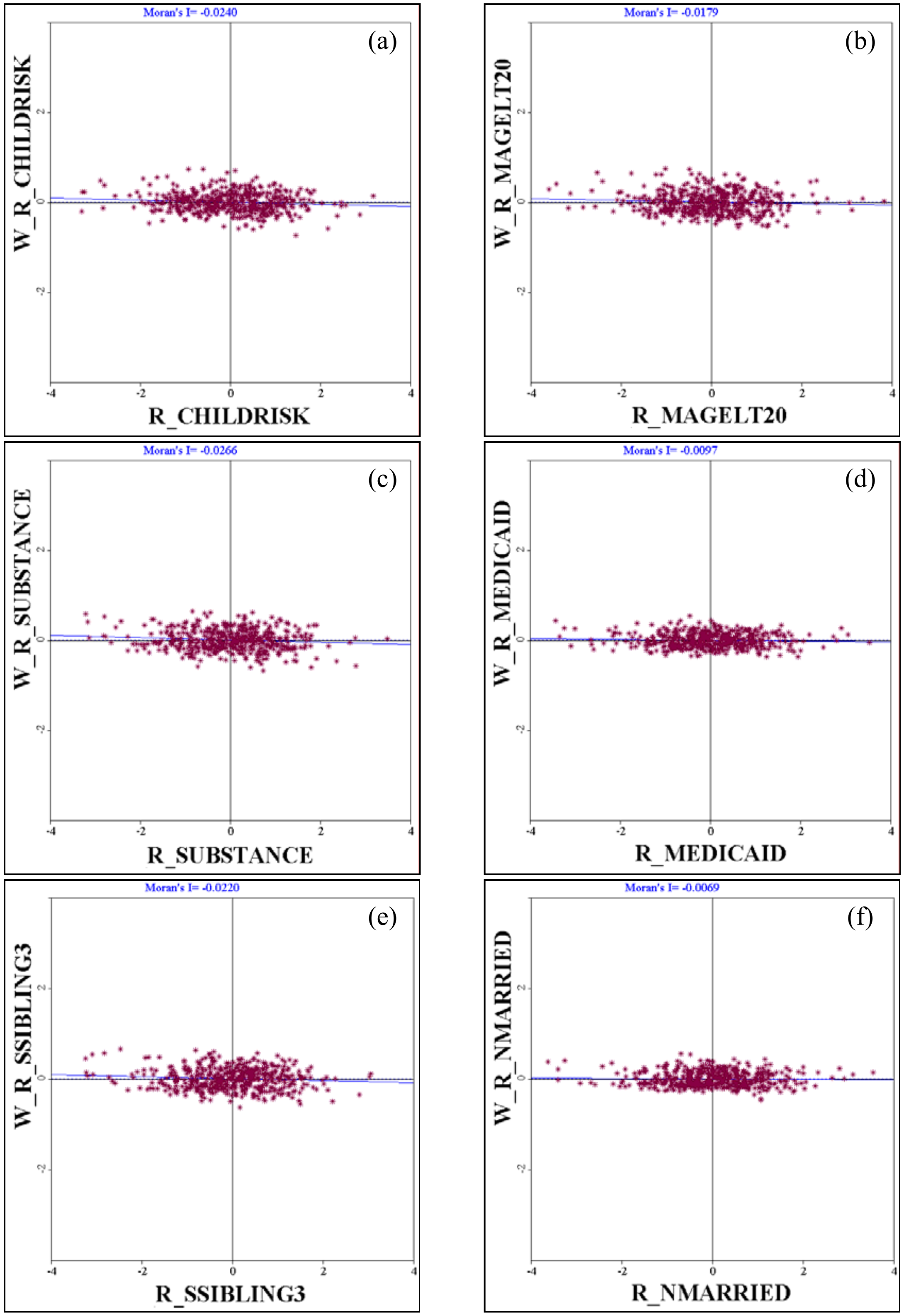


Figure 6.14 Moran scatter plots of the bivariate spatial regression residuals in the regression of TRATENEG on the microsystem variables.

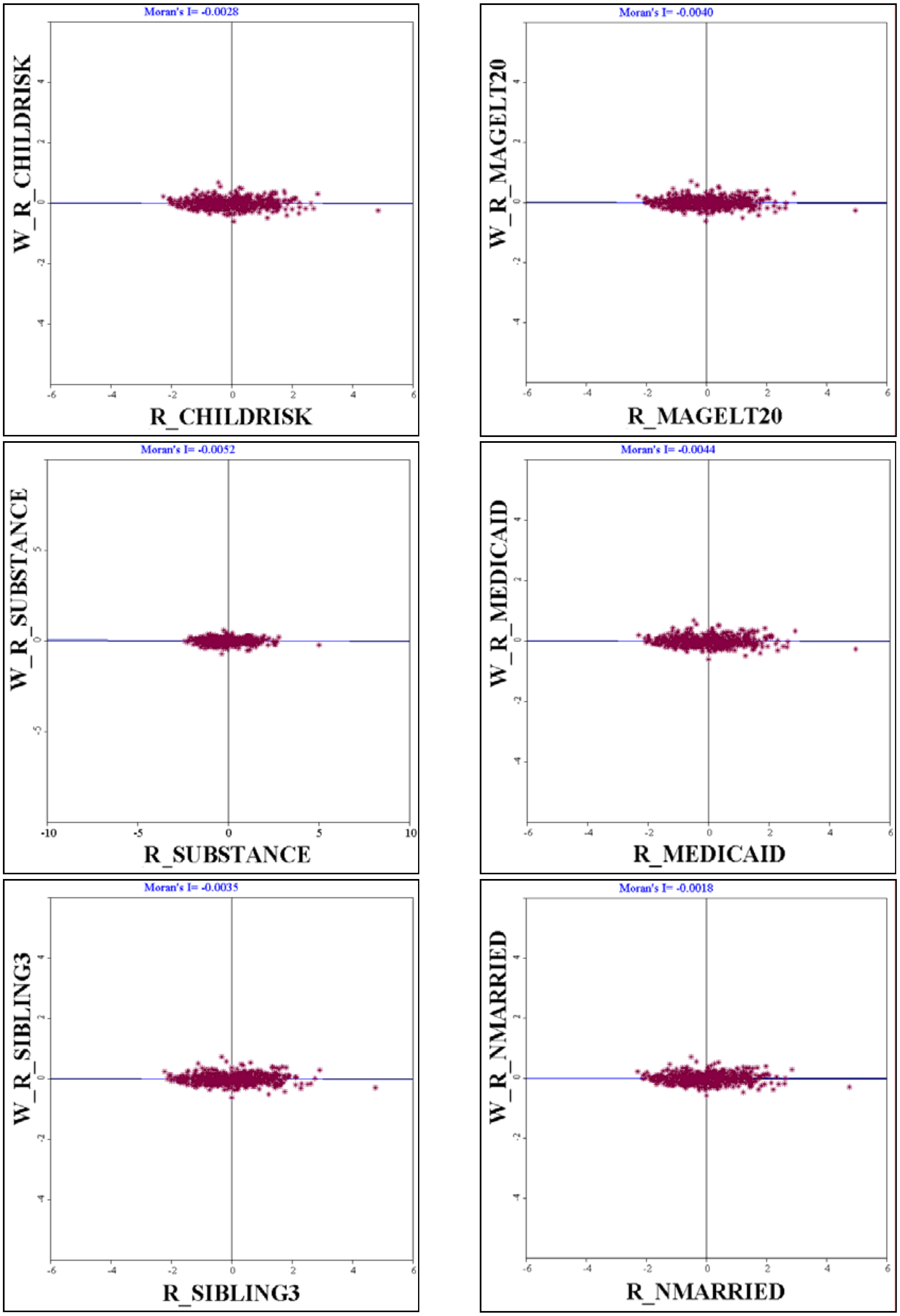


Figure 6.15 Moran scatter plots of the bivariate spatial regression residuals in the regression of TRATEPE on the microsystem variables.

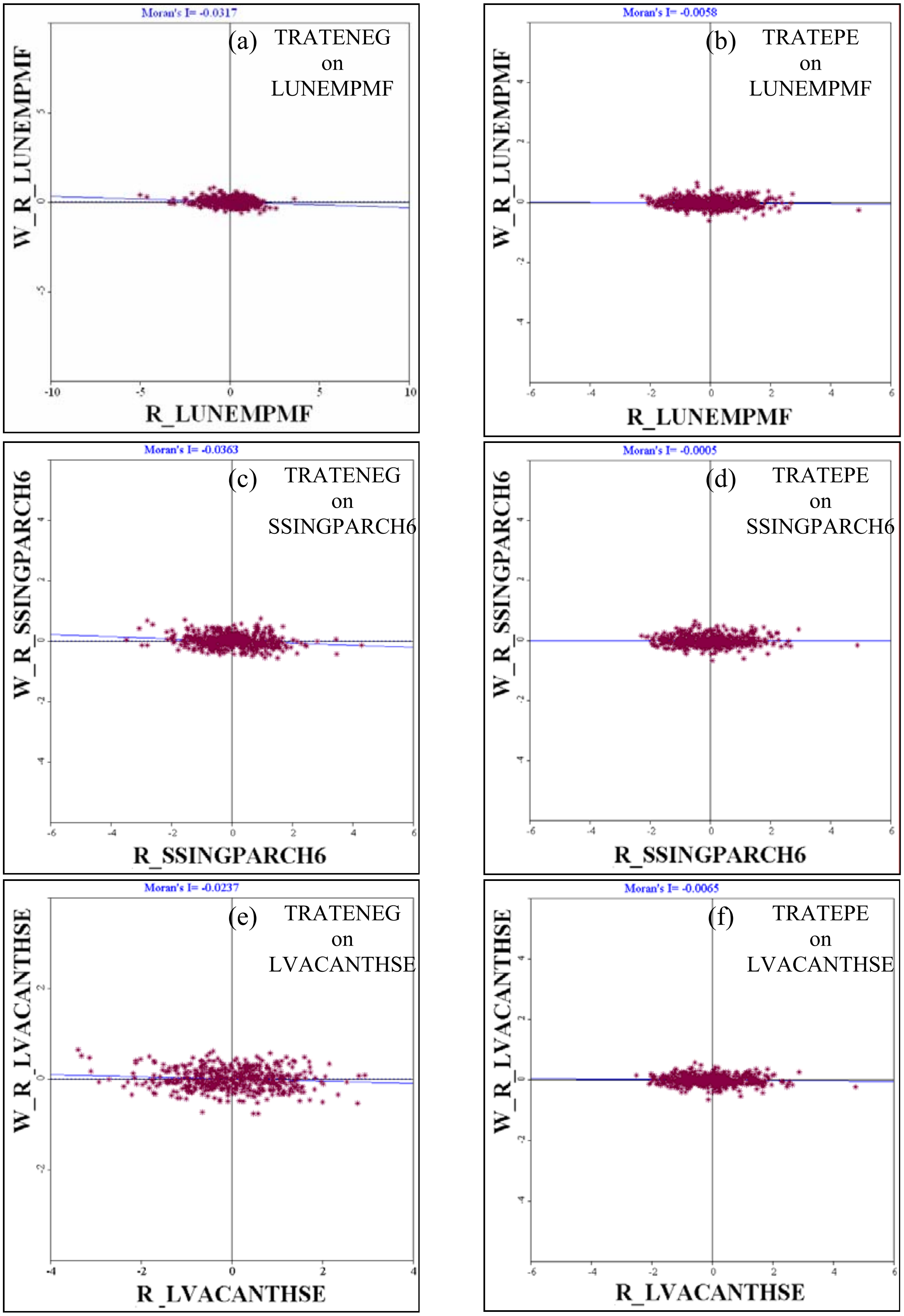


Figure 6.16 Moran scatter plots of the bivariate spatial regression residuals in the regression of TRATENEG and TRATEPE on the exosystem variables.

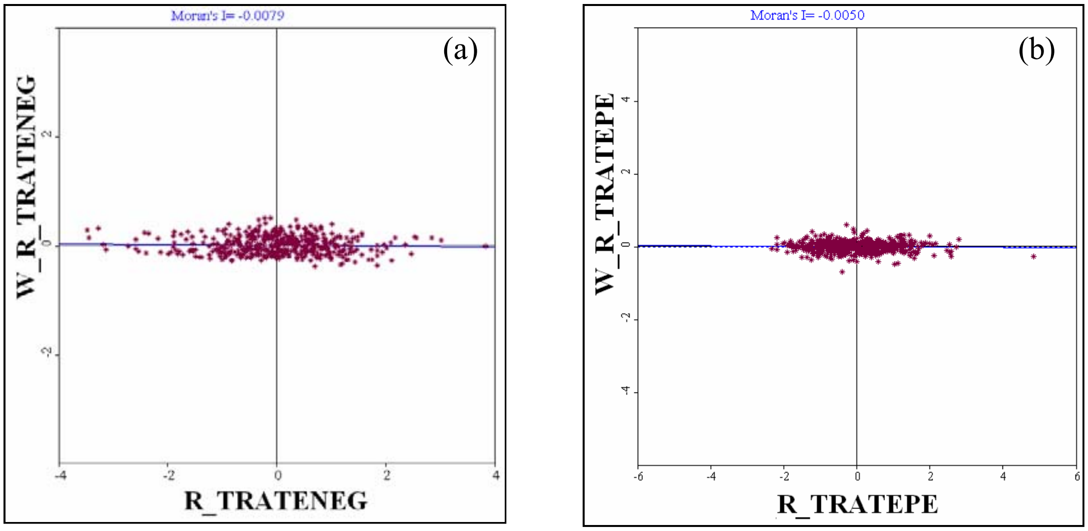


Figure 6.17 Moran scatter plots of the multivariate spatial regression residuals: (a) regression of TRATENEG; (b) regression of TRATEPE

CHAPTER SEVEN

DISCUSSION

7.1 ISSUES TO NOTE

There are several issues to note in order to properly interpret the results: 1) ecological design, 2) difficulty in interpreting the spatial regression results, and 3) unequal number of observations.

The first issue is the ecological study design. This study confirms that, at the census tract level, some examined risk variables are significantly, positively related to the rates of substantiated neglect and physical/emotional abuse. Relationships found in the ecological studies, however, may suffer from two problems: the ecological fallacy and modifiable areal unit problem (MAUP) (Wong, 1996; Waller & Gotway, 2004). The ecological fallacy refers to the logical fallacy inherent in making causal inferences from aggregated data to individual characteristics or behavior. This problem was first documented by Robinson (1950) who stressed the difference between the ecological correlations based on the aggregated data and the individual correlations based on the individual data. The examples provided in this article showed that the correlation coefficients based on the data aggregated by areas such as state were quite different in strength and even in signs from those based on the individual data. To avoid the ecological fallacy, any conclusions drawn from the analyses at the ecological level should not be inferred to the individuals.

For the same reason, relationships found at the census tract level in the present study should not be inferred to individuals who lived in the census tracts. For example, it is found that variable SUBSTANCE (percent of mothers who smoked or drank alcohol during pregnancy) is significantly related to the transformed rates both of substantiated neglect and of substantiated physical/emotional abuse. However, it cannot be concluded based only on the ecological study results that children under the age of four whose mothers smoked during pregnancy are at higher risk of being maltreated than those children whose mothers did not smoke during pregnancy.

The MAUP refers to the inconsistency of analytical or statistical results derived from data recorded or aggregated at different levels of partitioning (referred to as the scale effect) or aggregated to areas partitioned in different ways (referred to as the zonal effect) for the same geographic domain (Wong, 1996). The scale effect was recognized as early as in the 1930s, e.g., by Gehlke and Biehl (1934). In ecological correlations, the scale effect means the correlation coefficient tends to increase when the size of the areal units for which the data are aggregated increases while the number of units becomes smaller, or vice versa.

Openshaw and Taylor (1979) exemplified the zonal effect. They agglomerated the 99 counties in Iowa in different ways into fewer larger zones and aggregated the number of elderly and those voting Republican. They then calculated the correlation coefficient between the percent of elderly and the percent of those voting Republican, and obtained a large variety of correlation coefficients, e.g., ranging from –0.811 to 0.979 when the number of zones was 24.

Fotheringham and Wong (1991) investigated both the scale and zonal effects on multivariate regression and logistic regression models. They found the parameter estimates of regression models derived from different scales or different configuration schemes at a given scale were widely dispersed and even had opposite signs.

For the characteristics of data used in the present study, there is no doubt that the results derived are scale and configuration dependent, that is, they are pertinent only to the current configuration of the 2000 census tract partitioning. Therefore, they should not be interpreted for the county or census block group level. It is likely that different findings would be derived from the analyses performed at the census block group level or for different portioning schemes of the 478 census tracts. For example, different results might have been derived if the data used for analyses were aggregated for the 1990 census tracts.

Another implication of the MAUP to the present study is that if the current configuration of the 2000 census tract does not capture the actual child maltreatment phenomenon, the results may not be reliable. Although the spatial regression method used in the present study does account for the scale effect (Anselin, 1988), it is unclear if the zonal effect is taken into account as well.

The second issue is the difficulty in interpreting spatial regression results. Spatial regression is used to control for spatial autocorrelation effects. However, introducing the spatial autocorrelation variable may complicate parameter interpretations (Waller & Gotway, 2004). This can be seen from the following example.

Assume spatial autocorrelation is caused by omitting important explanatory variables. Let $X_1 = (x_1, x_2, \cdots, x_p, x_{p+1}, \cdots x_k)$ denote the complete set of explanatory variables, and $\beta_1 = (b_1, b_2, \cdots, b_p, b_{p+1}, \cdots b_k)$ corresponding coefficients; $X_2 = (x_1, x_2, \cdots, x_p)$ are the subset of variables included in the regression model, and $\beta_2 = (b_1, b_2, \cdots, b_p)$ corresponding to X_2; e_1 and e_2 are the error terms associated with $X_1\beta_1$ and $X_2\beta_2$, respectively. Under the above assumption, e_1 is independent error term with a mean of zero, and e_2 is correlated due to the omission of variables $(x_{p+1}, \cdots, x_k)$. It is hoped that the estimates of $\beta_2 = (b_1, b_2, \cdots, b_p)$ are the same as the estimates by the OLS method if the complete set of variables were included. But in fact, they may be quite different although the two models are both valid and may be comparable in terms of the measures of goodness-of-fit (i.e., the AIC values). This is because the spatial autocorrelation variable capturing the effects caused by omitting $(x_{p+1}, \cdots, x_k)$ is spatially correlated with the error term in the spatial regression model (Anselin & Bera, 1998). Thus, the influences of the omitted variables $(x_{p+1}, \cdots, x_k)$ may not be completely captured by the estimate of the spatial autocorrelation parameter; part of the influences may be imposed on the estimates of β_2.

Therefore, spatial regression should be used as only the last resort to give more reasonable results than the classic OLS method whenever spatial autocorrelation is inevitable.

The third issue is unequal number of observations. Both visual and statistical analyses find that, in general, the examined risk variables have stronger relationships with the transformed rates of substantiated neglect than with the transformed rates of substantiated physical/emotional abuse. However, these findings may have resulted from unequal number of observations. As noted earlier, there are 3,526 substantiated neglect victims in 405 census tracts, but only 313 victims of substantiated physical/emotional abuse in 167 census tracts. The accuracy of the calculation of rates is directly related to the number of observations. When the number of observations is too small, the estimates of rates are subject to substantial random variation. As a consequence, any relationship that actually exists may not be demonstrated by the data. Therefore, while the number of substantiated neglect victims may be large enough to show the relationships between the response variable and the risk variables, the number of substantiated physical/emotional abuse victims might be too small to show similar relationships.

7.2 SOURCES OF SPATIAL AUTOCORRELATION

Spatial autocorrelation effects are found statistically significant in both bivariate and multivariate regression models. These effects may result from several sources described in Chapter Five (Subsection 5.3.1). First, the response variables may be inherently spatially dependent. Since the observations of the response variable are not acquired through a strict sampling design but a collection of data arranged by the geographic unit, i.e., the census tract, the interdependence between observations of neighboring census tracts may be the rule rather than the exception (Anselin & Bera, 1998). This has been recognized long before, as stated in “the first law of geography” that “everything is related to everything else, but near things are more related than distant things.” (Tobler, 1970)

Second, the unit of analysis may not match the unit of actual phenomena. The present study uses the census tract as a surrogate for the community. The rationale is that census tracts are designed to be demographically homogeneous; if the outcomes of child maltreatment are related to similar demographic characteristics, then tracts may offer some value in natural groupings of individuals. However, there is no compelling reason at this time to believe that child maltreatment conforms to the configuration of census tracts.

Third, important explanatory variables might be missing. It is apparent that many variables are not included in any of the bivariate regression models. Even in the multivariate regression models, there are probably important variables missed. This is because child maltreatment is multiply determined by multiple forces at multiple levels, as suggested by the ecological theories of child maltreatment (Garbarino, 1977; Belsky, 1980), and the variables examined in the present study are only part of the factors.

Lastly, the observations of the response and/or explanatory variables may not be free of errors. The geocoding of child maltreatment records and vital birth records may be a source of measurement errors. For example, at least 8.3% of maltreatment records are not geocoded to the accuracy of the census tract level due to lack of appropriate address information. For these records, latitudes and longitudes are randomly assigned within tracts where the victims had the highest probability to live. Therefore, a record is more likely to be located in a census tract that is close to the correct tract than in a tract farther away.

CHAPTER EIGHT

CONCLUSIONS

8.1 STUDY FINDINGS

The ecological theory of child maltreatment considers child maltreatment the product of a multiplicity of risk factors in the abuser's ontogenic development, the family (microsystem), the community (exosystem), and the culture (macrosystem). Under this framework, this ecological study examined, at the census tract level, the rates of substantiated neglect and substantiated physical/emotional abuse among children under four years old by their biological parents, and their relationships with six risk variables in the microsystem: (1) percent of births experiencing neonatal difficulties (premature birth, low birth weight, or low five-minute Apgar score), (2) percent of births to mothers less than 20 years old, (3) percent of births to mothers who smoked or drank alcohol during pregnancy, (4) percent of births to Medicaid beneficiaries, (5) percent of births having three or more siblings, and (6) percent of births to non-married mothers, as well as six risk variables in the exosystem: (1) percentage of single parent families with children under six years old, (2) percentage of females 16 and older (with children under six years old) in the labor force outside the home, (3) percentage of families living in the current residence less than one year, (4) percentage of single-family housing units, (5) percentage of vacant housing units, and (6) percentage of males and females 16 years and older in the labor force who are unemployed. The microsystem variables reflected the characteristics of the child, the mother, or the family of a birth. The first five exosystem variables were chosen to indicate inadequate social support from the community, and the last variable to indicate socioeconomic resource stress. The hypothesis was that the rates of substantiated neglect and of substantiated physical/emotional abuse were positively related to the risk variables. The hypothesis was first examined through visual analyses including reviewing maps and investigating scatter plots, and then tested using spatial regression methods, which controlled for the effect of spatial autocorrelation.

Findings from visual analyses of maltreatment rates in relation to three variables: percentage of females 16 years and older (with children under six years old) in the labor force outside the home, percentage of families living in the current residence less than one year, and percentage of single-family housing units did not support the hypothesis. Neither the first nor the second variable was related to the rates of either type of maltreatment. The third variable was related to the rates of both types, but the relationships were opposite to the hypothesized direction. These three variables were dropped from the regression analysis.

Findings from bivariate spatial regression analyses of the transformed rates of substantiated neglect on nine risk variables supported the hypothesis: the transformed rates of substantiated neglect were significantly, positively related to each of the nine variables. The top four variables that give smaller AIC values are: percent of births to Medicaid beneficiaries, percent of births to non-married mothers, percentage of males and females 16 years and older in the labor force who are unemployed, and percent of births to mothers less than 20 years old.

Bivariate spatial regression analyses of the transformed rates of substantiated physical/emotional abuse on nine risk variables suggested that the transformed rates of substantiated physical/emotional abuse were significantly, positively related to seven explanatory variables at least at the 0.05 level, but not related to the other two variables: percent of births to Medicaid beneficiaries and percentage of single parent families with children under six years old. The top four variables that give smaller AIC values are: percent of births to mothers who smoked or drank alcohol during pregnancy, percent of births having three or more siblings, percentage of vacant housing units, and percent of births to non-married mothers.

Four variables, percent of births to non-married mothers, percentage of males and females 16 years and older in the labor force who are unemployed, percent of births to mothers who smoked or drank alcohol during pregnancy, and percentage of single parent families with children under six years old, were identified through multivariate spatial regression as the set of independent variables that best predicted the transformed rates of substantiated neglect. All four variables were significantly contributive. The model had a moderate predictive ability. The set of independent variables that best predicted the transformed rates of substantiated physical/emotional abuse included single variable: percent of births to mothers who smoked or drank alcohol during pregnancy. The model had a low predictive ability. Variable “percent of births to mothers who smoked or drank alcohol during pregnancy” was the only variable that was significantly predictive of the rates of substantiated neglect and the rates of substantiated physical/emotional abuse.

Comparisons of the above findings suggested that the relationships between substantiated maltreatment rates and the examined risk variables differed by type of maltreatment. Findings also suggested that the combination of risk variables that best predicted the rates of substantiated maltreatment differed by type of maltreatment.

Spatial autocorrelation effects were found statistically significant in the OLS residuals of both bivariate and multivariate regression models. Results of this study supported the idea that when spatial autocorrelation is present in the OLS regression residuals, the absolute values of the test statistic are upward biased. The relationships of the transformed rates of substantiated physical/emotional abuse to two variables, percent of births to Medicaid beneficiaries and percentage of single parent families with children under six years old, were found significant in the bivariate OLS regression models but not significant in the bivariate spatial regression models. Similarly, the transformed rates of substantiated physical/emotional abuse were found significantly related to two variables, percent of births to non-married mothers and percentage of vacant housing units, in the multivariate OLS regression; but the relationships were found not significant in the multivariate spatial regression.

8.2 IMPLICATIONS

The present study has two implications for public health. First, it may help design community-based, proactive child maltreatment intervention programs. This proactive approach may help not only prevent young children from experiencing negative developmental outcomes, but also effectively allocate scarce resources. The risk variables examined in the present study were directly computed from the birth variables defined by NCHS for birth certificates and the US decennial census data. Through routinely assessing the variables identified as significant predictors of maltreatment rates, it is easy to identify high-risk areas for maltreatment among young children. These high-risk areas may be targeted for community-based interventions before maltreatment occurs.

For example, it was found that "percent of births to mothers who smoked or drank alcohol during pregnancy" was a significant predictor both for the rates of substantiated neglect and for the rates of substantiated physical/emotional abuse. This finding has at least two potential public health implications. First, allocating resources in the census tracts with high percentage of births to mothers who smoked or drank alcohol during pregnancy to provide social support to the needed families is a priority in order to prevent child maltreatment occurrence. Second, great emphasis should be put on community-based programs aiming at reducing the use of tobacco and alcohol among pregnant women and addressing the underlying stressful conditions under which smoking and drinking alcohol take place.

Second, spatial autocorrelation must be taken into account in the area-based ecological models of public health research to provide more reliable results. Ignoring the presence of spatial autocorrelation in analyzing spatially aggregated data, using traditional methods with nonspatial data, may lead to false significant relationships. Demonstrated in the present study, one would have been led to believe that high percentages of births to non-married mothers and vacant housing units in the community are significant predictors for substantiated physical/emotional abuse if the effects of spatial dependency among neighboring communities have not been controlled.

8.3 RECOMMENDATIONS FOR FUTURE STUDY

Based upon the present study, several recommendations are readily made for future studies. First, it may be of benefit in future studies to examine the interactions between the microsystem variables and the exosystem variables, and include the interactions in the multivariate regression models to determine if including them can improve predictive abilities. The present study only examined individual risk variables and a combination of individual variables. In fact, the essence of the ecological approach is that it focuses on not only individual variables, but also interactions among variables across systems.

Second, it is recommended in future studies to examine the protective or supportive factors that may decrease the probability of maltreatment. The present study only examined the risk variables under the ecological framework. The models designed through multivariate spatial regression had moderate to low abilities to predict community substantiated maltreatment rates. Missing important individual risk variables and the interactions between variables might be one possible reason. However, it has been recognized in previous studies that the ecological theories alone do not help explain why maltreatment rates vary across areas with similar risks. There exist other factors that play roles in decreasing the probability of maltreatment, as suggested by the ecological/transactional theory and the ecological/developmental theory. Identifying protective factors may be beneficial not only to improve the model's predictive abilities, but also to provide recommendations for implementing community-based prevention programs to strengthen those protective factors.

Last but not least, future studies using a multi-level approach may be beneficial. At this time, a barrier to the use of the multi-level approach is the unavailability of child maltreatment data linked to birth records. However, this is not an unsolvable problem. In fact, some US states such as Florida and California have implemented the data collection systems that automatically link the maltreatment cases to birth records. Similar systems may be implemented in Georgia as well. When data are in place, the multi-level modeling is considered more suited to model the concepts of the ecological framework that emphasizes the nested arrangement of the individuals, the families, the communities, and the societies.

REFERENCES

Anda, R. F., Croft, J. B., Felitti, V. J., Nordenberg, D., Giles, W. H., Willimson, D. F., et al. (1999). Adverse childhood experiences and smoking during adolescence and adulthood. *Journal of the American Medical Association, 282*, 1652-1658.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.

Anselin, L. (2002). Under the hood: Issues in the specification and interpretation of spatial regression models. *Agricultural Economics, 27*(3), 247-267.

Anselin, L. (2003). *GeoDa 0.9.3 user's guide*. Retrieved October 1, 2005, from http://www.sal.uiuc.edu/stuff/stuff-sum/pdf/geoda093.pdf

Anselin, L. (2004). *GeoDa 0.9.5-i release notes*. Retrieved October 1, 2005, from http://www.sal.uiuc.edu/stuff/stuff-sum/pdf/geoda095i.pdf

Anselin, L. (2005). *Exploring spatial data with GeoDa: a workbook*. Retrieved October 1, 2005, from http://www.geoda.uiuc.edu/pdf/geodaworkbook.pdf

Anselin, L., & Bera, A. K. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ullah, and D. E. A. Giles (Ed.), *Handbook of Applied Economic Statistics* (pp. 237-289). New York: Marcel Dekker Inc.

Anselin, L., & Griffith, D. A. (1988). Do spatial effects really matter in regression analysis? *Papers of the Regional Science Association, 65*, 11-34.

Belsky, J. (1980). Child maltreatment: an ecological integration. *American Psychologist, 35*(4), 320-335.

Belsky, J. (1993). Etiology of child maltreatment: a developmental-ecological analysis. *Psychological Bulletin, 114*(3), 413-434.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.

Benirschka, M., & Binkley, J. K. (1994). Land price volatility in a geographically dispersed market. *American Journal of Agricultural Economics, 96*, 185-195.

Bifulco, A., Bernazzani, O., Moran, P. M., & Ball, C. (2000). Lifetime stressors and recurrent depression: preliminary findings of the Adult Life Phase Interview (ALPHI). *Social Psychiatry & Psychiatric Epidemiology, 35*(6), 264-275.

Bronfenbrenner, U. (1974). Developmental research, public policy, and the ecology of childhood. *Child Development, 45*, 1-5.

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist, 32*, 513-531.

Bugental, D. B., & Happaney, K. (2004). Predicting infant maltreatment in low-income families: the interactive effects of maternal attributions and child status at birth. *Developmental Psychology, 40*(2), 234-243.

The Child Abuse and Prevention and Treatment Act as Amended by the Keeping Children and Families Safe Act of 2003, Pub. L. No. 108-36, Sec. 111 (2004).

Chatterjee, S., Hadi, A. S., & Price, B. (2000). *Regression Analysis by Example* (3rd ed.). New York: Wiley.

Cicchetti, D., & Lynch, M. (1993). Toward an ecological/transactional model of community violence and child maltreatment: consequences for children's development. *Psychiatry, 56*, 96-118.

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

Dembo, R., Williams, L., Berry, E., Getreu, A., Washburn, M., Wish, E., et al. (1988). The relationship between physical and sexual abuse and illicit drug use: a replication among a new sample of youths entering a juvenile detention center. *International Journal of the Addictions, 23*(11), 1101-1123.

DFCS. (2004). *Protective services data system annual report 2003*. Retrieved February 2, 2006, from http://dfcs.dhr.georgia.gov/DHR-DFCS/DHR-DFCS_CommonFiles/28822306PSDS_2003_Report_Draft.pdf

DHHS. (2001). *Child maltreatment 1999*. Retrieved January 19, 2006, from http://www.acf.hhs.gov/programs/cb/pubs/cm99/index.htm

DHHS. (2003). *Child maltreatment 2001*. Retrieved January 19, 2006, from http://www.acf.hhs.gov/programs/cb/pubs/cm01/index.htm

DHHS. (2005a). *Child maltreatment 2003*. Retrieved January 19, 2006, from http://www.acf.hhs.gov/programs/cb/pubs/cm03/index.htm

DHHS. (2005b). *NCANDS survey instrument*. Retrieved February 10, 2006, from http://www.acf.hhs.gov/programs/cb/systems/ncands/survey.htm

DHHS. (2006). *Statistics & Research*. Retrieved January 19, 2006, from http://www.acf.hhs.gov/programs/cb/stats_research/index.htm#can

Drake, B., & Pandey, S. (1996). Understanding the relationship between neighborhood poverty and specific types of child maltreatment. *Child Abuse & Neglect, 20*(11), 1003-1018.

Egeland, B., Yates, T., Appleyard, K., & van Dulmen, M. (2002). The long-term consequences of maltreatment in the early years: a developmental pathway model to antisocial behavior. *Children's Services: Social Policy, Research, and Practice, 5*(4), 249-260.

Ernst, J. S. (2000). Mapping child maltreatment: looking at neighborhoods in a suburban county. *Child Welfare, 79*(5), 555-572.

Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., et al. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: the adverse childhood experiences (ACE) study. *American Journal of Preventive Medicine 14*(4), 245-258

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically Weighted Regression: the Analysis of Spatially Varying Relationships*. Chichester, England: John Wiley & Sons Ltd.

Fotheringham, A. S., & Wong, D. W. S. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A, 23*, 1025-1044.

Freisthler, B. (2004). A spatial analysis of social disorganization, alcohol access, and rates of child maltreatment in neighborhoods. *Children and Youth Services Review, 26*, 803-819.

Frodi, A. M., Lamb, M. E., Leavitt, L. A., Donovan, W. L., Neff, C., & Sherry, D. (1978). Fathers' and mothers' responses to the faces and cries of normal and premature infants. *Developmental Psychology, 14*(5), 490-498.

Fromm, S. (2001). *Total estimated cost of child abuse and neglect in the United States: statistical evidence*. Retrieved October 30, 2006, from http://www.preventchildabuse.org/learn_more/research_docs/cost_analysis.pdf

Garbarino, J. (1976). A preliminary study of some ecological correlates of child abuse: the impact of socioeconomic stress on mothers. *Child Development, 47*, 178-185.

Garbarino, J. (1977). The human ecology of child maltreatment: a conceptual model for research. *Journal of Marriage and the Family, 39*(4), 721-735.

Garbarino, J., & Crouter, A. (1978). Defining the community context for parent-child relations: the correlates of child maltreatment. *Child Development, 49*, 604-616.

Gehlke, C. E., & Biehl, K. (1934). Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association Supplement, 29*(185), 169-170.

Child Abuse and Neglect Prevention Act, Georgia General Assembly Ch. 19-14 Sec. 2.

Children and Youth Act, Georgia General Assembly Ch. 49-5, Sec. 8, 180 and 181.

Getis, A. (1990). Screening for spatial dependence in regression analysis. *Papers of the Regional Science Association, 69*, 69-81.

Group 1 Software. (2003). Centrus Desktop (Version 4.01.01). Mountain View, CA: Group 1 Software Inc.

Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge, UK: University Press.

Hallisey, E. (2005). Cartographic visualization: an assessment and epistemological review. *Professional Geographer, 57*(3), 350-364.

Kempe, C. H., Silverman, F. N., Steele, B. F., Droegemueller, W., & Silver, H. K. (1962). The battered child syndrome. *Journal of the American Medical Association, 181*(1), 17-24.

Kessler, R. C., & Magee, W. J. (1994). Childhood family violence and adult recurrent depression. *Journal of Health and Social Behavior, 35*(1), 13-27.

Kotch, J. B., Browne, D. C., Dufort, V., Winsor, J., & Catellier, D. (1999). Predicting child maltreatment in the first 4 years of life from characteristics assessed in the neonatal period. *Child Abuse & Neglect, 23*(4), 305-319.

Kotch, J. B., Browne, D. C., Ringwalt, C. L., Dufort, V., Ruina, E., Stewart, P. W., et al. (1997). Stress, social support, and substantiated maltreatment in the second and third years of life. *Child Abuse & Neglect, 21*, 1025-1037.

Kotch, J. B., Browne, D. C., Ringwalt, C. L., Stewart, P. W., Ruina, E., Holt, K., et al. (1995). Risk of child abuse or neglect in a cohort of low-income children. *Child Abuse & Neglect, 19*(9), 1115-1130.

Krieger, N., Waterman, P. D., Chen, J. T., Rehkopf, D. H., & Subramanian, S. V. (2004). *Geocoding and monitoring US socioeconomic inequalities in health: an introduction to using area-based socioeconomic measures-- The Public Health Disparities Geocoding Project monograph*. Retrieved September 28, 2005, from http://www.hsph.harvard.edu/thegeocodingproject/

Krishnan, V., & Morrison, K. B. (1995). An ecological model of child maltreatment in a Canadian province. *Child Abuse & Neglect, 19*(1), 101-113.

Millard, F., & Freymann, G. (2001). Spatial imputation method for birth and fetal death records: assigning birth and fetal death records that fail address-matching criteria to census tracts within a county. Atlanta: Georgia Department of Human Resources, Division of Public Health, Office of Health Information and Policy.

Mitchell, A. (2005). *The ESRI Guide to GIS Analysis, Volume 2: Spatial Measurements and Statistics*. Redlands, CA: ESRI Press.

NACC. (2005). *Children and the law*. Retrieved January 30, 2006, from http://www.naccchildlaw.org/childrenlaw/childmaltreatment.html

NCCANI. (2004). *About CAPTA: a legislative history*. Retrieved January 30, 2006, from http://www.childwelfare.gov/pubs/factsheets/about.cfm

Openshaw, S., & Taylor, P. J. (1979). A million or so correlation coefficients: three experiments on the modifiable areal unit problem. In N. Wrigley (Ed.), *Statistical applications in the spatial sciences* (pp. 127-144). London: Pion Limited.

Robison, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*(3), 351-357.

Rogerson, P. A. (2006). *Statistical Methods for Geography: a Student's Guide* (2nd ed.). London: SAGE Publications.

Rovi, S., Chen, P.-H., & Johnson, M. S. (2004). The economic burden of hospitalization associated with child abuse and neglect. *American Journal of Public Health, 94*(4), 586-590.

Scannapieco, M., & Connell-Carrick, K. (2005). *Understanding Child Maltreatment: An Ecological and Developmental Perspective*. New York: Oxford University Press.

Sedlak, A. J. (2001). *A history of the national incidence study of child abuse and neglect*. Retrieved February 12, 2006, from https://www.nis4.org/NIS_History.pdf

Sedlak, A. J., & Broadhurst, D. D. (1996). *Executive summary of the third national incidence study of child abuse and neglect*. Retrieved February 2, 2006, from http://nccanch.acf.hhs.gov/pubs/statsinfo/nis3.cfm

Simpson, D. G., Imrey, P. B., Geling, O., & Butkus, S. (2000). Statistical estimation of child abuse rates from administrative databases. *Children and Youth Services Review, 22*(11/12), 951-971.

Springer, K. W., Sheridan, J., Kuo, D., & Carnes, M. (2003). The long-term health outcomes of childhood abuse. *Journal of General Internal Medicine, 18*, 864-870.

Timmreck, T. C. (2002). *An Introduction to Epidemiology* (3rd ed.). Sudbury, MA: Jones and Bartlett.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*(2), 234-240.

Tomison, A. M., & Wise, S. (1999). *Community-based approaches in preventing child maltreatment*. Retrieved March 27, 2006, from http://www.aifs.gov.au/nch/issues11.html

Tzeng, O. C. S., Jackson, J. W., & Karlson, H. C. (1991). *Theories of Child Abuse and Neglect: Differential Perspectives, Summaries, and Evaluations*. New York: Praeger.

U.S. Census Bureau. (2001). *Glossary of basic geographic and related terms – Census 2000*. Retrieved January 5, 2006, from http://www.census.gov/geo/www/tiger/glossary.html#glossary

UNICEF. (2003). *A league table of child maltreatment deaths in rich nations*. Retrieved January 1, 2006, from http://www.unicef-icdc.org/publications/pdf/repcard5e.pdf

Waller, L. A., & Gotway, C. A. (2004). *Applied Spatial Statistics for Public Health Data*. Hoboken, NJ: Wiley.

Weissman, A. M., Jogerst, G. J., & Dawson, J. D. (2003). Community characteristics associated with child abuse in Iowa. *Child Abuse & Neglect, 27*, 1145-1159.

Westat. (2004). *NIS 4*. Retrieved February 12, 2006, from https://www.nis4.org/nishome.asp

WHO. (1999). *Report of the consultation on child Abuse prevention*. Retrieved January 5, 2006, from http://www.who.int/mipfiles/2017/childabuse.pdf

Widom, C. S. (1999). Posttraumatic stress disorder in abused and neglected children grown up. *American Journal of Psychiatry, 156*, 1223-1229.

Wong, D. W. S. (1996). Aggregation effects in geo-referenced data. In S. L. Arlinghaus, D. A. Griffith, W. C. Arlinghaus, W. D. Drake & J. D. Nystuen (Eds.), *Practical Handbook of Spatial Statistics* (pp. 83-106). Boca Raton, FL: CRS Press Inc.

Wu, S. S., Ma, C.-X., Carter, R. L., Ariet, M., Feaver, E. A., Resnick, M. B., et al. (2004). Risk factors for infant maltreatment: a population-based study. *Child Abuse & Neglect, 28*, 1253-1264.

Young, G., & Gately, T. (1988). Neighborhood impoverishment and child maltreatment: an analysis from the ecological perspective. *Journal of Family Issues, 9*(2), 240-254.

Zhou, Y., Hallisey, E., & Freymann, G. (2006). Identifying perinatal risk factors for infant maltreatment: an ecological approach Atlanta: Division of Public Health, Georgia Department of Human Resources.

Zuravin, S. J. (1989). The ecology of child abuse and neglect: review of the literature and presentation of data. *Violence and Victims, 4*(2), 102-120.

APPENDICES

Appendix A Paradigms of research on child maltreatment theories

Paradigms	Units of Analysis	Major Theories	Variables to be Studied
Individual determinants	• Individual maltreaters	• Psychiatric • Psychoanalytic • Intrapsychic • Humanistic	*Physical abuse* • Traumatic experiences in early childhood (being the victim or witness of abuse) • Abnormal characteristics (psychopathology, personality defects, poor impulse control, and substance abuse) • Affective processes (inappropriate or blunt emotions, negative affect toward the child, poor self-esteem) • Distorted cognitive processes (rationalizations for abusive behavior, inaccurate beliefs about the child and child discipline) • Reinforcement (being relieved of intrapsychic tension and the quieting of the child) *Sexual abuse* • Traumatic experiences in early childhood (being the victim of abuse) • Abnormal characteristics (excessive hostility, anxiety, mental illness, alcoholism, and psychosexual disorders) • Lack of personal resources (poor self-esteem, inadequate social skills) • Short-term stressors (fights, work-related problems, and substance abuse) • Cognitive processes (rationalizations and irresponsibility in decision and choice making)
Offender typology	• Individual maltreaters	• Typology of physical abusers • Typology of sexual abusers	*Physical abuse* • Socially or parentally incompetent • Acting out of frustration or displacement • Generally neglectful • Limited in cognitive abilities (having low intelligence and/or poor judgment) • Mentally ill *Sexual abuse* • Regressed offenders associated with stress • Fixated offenders associated with sexual attraction to children
Family systems	• Family system • Interactions with individual, community, and cultural systems	• Family systems	*Physical abuse* • Personal characteristics of each family member • Personal stressors • Cognitive processes (beliefs concerning the use of punishment) • Family structure (single/both parents, family size) • Family values (goals and level of acceptance of violence) • Family dynamics (feedback mechanisms and interactions among family members) • Interaction between the family system and other systems (formal community organizations and neighbors)

Sociocultural	• Social • Cultural • Economic • Political	• Social systems • Social isolation • Patriarchal	*Physical abuse* • Social stressors (unemployment, low income, large family size, poor education, social isolation, and low social class) • Mismanagement of national resources • High degree of competition for jobs • Formal and informal social isolation factors • Social ideologies that teach selfishness and disconcern for others • An established inegalitarian and abusive social order • Symbolic social violence against families *Sexual abuse* • Social isolation • Fixated offenders associated with sexual attraction to children
Individual-environmental interaction	• Individual maltreaters • Sociological	• Resource • Three-component • Social psychological • Symbiosis • Social interaction • Three-factor • Exchange /Control • General stress	*Physical abuse* • Personality traits (authorizatarianism, dependency needs, impulsiveness, and psychopathology) • Personal resources (self-esteem, parenting skills, and stress-coping mechanisms) • Personal stressors (family conflicts, illness, and disruptive child behavior) • Cognitive processes (perceiving the child as being difficult, having a negative attitude toward the child, and preconventional cognitive development level of moral reasoning) • Characteristics of the family (adverse marital relationship, norms for punishment, and family dynamics) • Community values and norms (subcultural acceptance of violence, childrearing practices, and community isolation) • Sociocultural variables (socioeconomic status, cultural scriptings, and social controls of behavior) • Characteristics of the child (prematurity, hyperactivity, and low birth weight) *Sexual abuse* • Being motivated from internal reason • Internal inhibitors being lacking or weakened (alcohol, stress, learned rationalizations, personal disorders) • External inhibitors being lacking or weakened (poor supervision of the child, isolation, or crowded housing conditions) • Child's resistance being overcome

Parent-child interaction	• Parent, child • Parent-child relationship • Environmental factors	• Transactional • Encounter • Cognitive, behavioral and developmental	*Physical abuse* • Disturbed parent-child relationship • Characteristics of the parent (disturbances in impulse control, cognitive dysfunctions, and emotional needs) • Characteristics of the child (resemblance to a disliked person, hyperactivity, too much or too little self-confidence, refusal to accept authority, and deviance) • Environmental factors (family/social stressors, social help networks, and contextual situations)
Sociobiological	• Genetic factors	• Socio-biological	*Physical abuse* • Weak parent-child bonding • Inadequate resources (poverty, large family size, single parenthood) • Premature or defective children
Learning/ situational	• Parent-child • Social • Situational	• Social learning • Situational analysis • Coercion	*Physical abuse* • Frustration (a child's interference with a parent's need for tranquility by crying) • Aggressive cues (environmental stimuli) • Aggression-produced rewards (quilting of a child or release of tension)
Ecological	• Individual (child and maltreaters) • Family • Community • Society	• Ecological • Ecological context • Family breakup	*Physical and psychological/emotional abuse* • Individual (social isolation and cognitive processes) • Family (family values, childrearing practices, stress, interactions among family members) • Community (support systems, social isolation, stressors) • Cultural (cultural attitudes and beliefs toward the child and child discipline)

Appendix B Area-based ecological studies of child maltreatment

Study	Geographic Location	Spatial Unit of Analysis	Types of Maltreatment	Calculation of Maltreatment Rates	Socioeconomic and Demographic Variables Examined
(Young and Gately, 1988)	EL Paso City, Texas	Block group	Abuse and neglect combined	Number of reported incidences per 1,000 families with children	1) % households with income more than $25,000 a year 2) % households with income less than $10,000 a year 3) % households headed by females 4) % females in the labor force with children under 6 years of age 5) % residents who moved to the current residence within the last five years
(Zuravin, 1989)	Baltimore City, Maryland	Census tract	Abuse; neglect	Number of families reported to CPSs per 1,000 families with children	1) % families with annual income greater than 400% of the poverty line 2) % families with annual income less than 200% of the poverty line 3) % families headed by females 4) % married women (with children under 6 years old) in the work force outside the home 5) % families living in current residence less than 1 year 6) % single family dwellings 7) % vacant housing
Coulton, Korbin, Su, and Chow, 1995)	Cleveland, Ohio	Census tract	Abuse and neglect combined	Number of children who experienced one or more confirmed instances per 1,000 children	1) % persons below poverty level 2) % residents unemployed 3) % vacant housing 4) % population change between 1980 and 1990 5) % residents who moved between 1985 and 1990 6) % households in current residence less than 10 years 7) % households that moved in 1 year 8) % households with children that are female-headed 9) Contiguous to poor tracts (more than 40% residents below poverty) (1=True) 10) ....
(Krishnan and Morrison, 1995)	Alberta, Canada	District office	Abuse and neglect combined	Number of maltreatment reports per 1,000 children (0-19 years old)	1) % population change between 1981 and 1986 2) % population 0-19 years old 3) % people unemployed 4) % females in labor farce 5) % Native people 6) % single-parent families

(Ernst, 2000)	Montgomery County, Maryland	Census tract	Physical abuse; sexual abuse; neglect	Number of families investigated for maltreatment per 1,000 families with children	1) % families with income below 200% of poverty line 2) % families with income above 400% of poverty line 3) % renters who pay more than 35% of income in rent 4) Median residential property value 5) % families female-headed 6) % females in labor force 7) % single-family dwellings 8) % movement 1989-1990 9) % movement 1985-1990 10) % arrivals 1985-1990
(Weissman, Jogerst, and Dawson, 2003)	State of Iowa	County	Abuse and neglect combined	Number of reported incidences per 1,000 children under 18; Number of substantiated incidences per 1,000 children under 18	1) % population unemployed 2) Median family income 3) % children under age 6 in poverty 4) Marriage dissolution rate 5) % singles with children under 18 6) Mean family size 7) ....
(Freisthler, 2004)	Three counties (Sacramento, Alameda, and Santa Clara) in California	Census tract	Abuse and neglect combined	Number of substantiated reports per 1,000 population	1) % female-headed families 2) % persons living in poverty 3) % persons unemployed 4) % vacant housing units 5) % population change between 1990 and 2000 6) % African American residents 7) % elderly person 8) Ratio of children ≤12 to adults ≥21 9) % persons moved last 5 years 10) % Hispanic residents 11) ....