DISPARITIES OF COLORECTAL CANCER SURVIVAL IN TEXAS

DISSERTATION

Presented to the Graduate Council of
Texas State University-San Marcos
in Partial Fulfillment
of the Requirements

for the Degree

Doctor of PHILOSOPHY

by

Neng Wan, B.S., M.S.

San Marcos, Texas
August 2011

DISPARITIES OF COLORECTAL CANCER SURVIVAL IN TEXAS

Committee Members Approved:

_____
F. Benjamin Zhan, Chair

_____
Yongmei Lu

_____
John Tiefenbacher

_____
Gerard Rushton

Approved:

_____
J. Michael Willoughby

Dean of the Graduate College

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

Duplication Permission

ACKNOWLEDGEMENTS

I want to express my gratitude to many people for the completion of this dissertation. First, my sincere appreciation goes to my dissertation advisor, Dr. F. Benjamin Zhan. Without his excellent mentorship, consistent support, and kind encouragement, this dissertation would not have been possible. He has challenged me to think broadly in research questions and to implement critically in methodology. What I have learned from him will continue to benefit my future research. I would also like to thank my outstanding committee members: Dr. Gerard Rushton, Dr. John Tiefenbacher, and Dr. Yongmei Lu. I am very grateful for their insightful comments, well-considered suggestions, as well as careful reviewing of my proposal and dissertation.

Also, I want to give my gratitude to the Department of Geography of Texas State. Special thanks go to the graduate staff advisor Allison Glass-Smith, the main office manager Angelika Wahl, the administrative assistant Pat Hell-Jones, and the computer lab coordinator Charles Robinson for their professional assistance for my study at Texas State. I thank Dr. Lawrence E. Estaville for his constant encouragement and help. I also thank my friends at Texas State. It was a pleasure to work and live with them in such a beautiful place.

Furthermore, I want to thank the Texas Department of State Health Services and Mr. Betts Paul, Dr. Brian King, and Dr. David Risser for their excellent work in providing cancer registry data and health professional data. I thank Dr. Qiang Cai of the

National Minority Quality Forum for his assistance in implementing the adaptive spatial filtering method in Chapter 5.

Finally and most importantly, I want to express my deepest appreciation to my family. I thank my dear wife, Haiyan, for her unconditional love, her ever increasing understanding, and her company during these years. I want to thank my parents, Zhongqing Wan, and Mingqing Pan, and my sister, Qiao Wan, for their love and consistent support for my education.

The manuscript was submitted on July 20, 2011.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

ABSTRACT

DISPARITIES OF COLORECTAL CANCER SURVIVAL IN TEXAS

by

Neng Wan, B.S., M.S.

Texas State University-San Marcos

August 2011

SUPERVISING PROFESSOR: F. BENJAMIN ZHAN

Disparities in colorectal cancer (CRC) outcomes in terms of race/ethnicity, socioeconomic status (SES), and geographic location have been widely documented in the United States. However, the mechanism of how different factors influence the disparities remains poorly understood. Previous studies of CRC disparity are limited because (1) they seldom evaluated the joint influence of multiple factors on CRC outcomes, (2) few US studies have investigated the association between potential spatial access to CRC services and CRC outcomes, and (3) the effects of spatial autocorrelation, the small number problem, and the ecological fallacy on analysis results were not fully accounted for. These limitations prevent researchers and health professionals from more accurately understanding the causes of disparities and effectively designing intervention programs. Taking advantage of Geographic Information Science and statistical methods, this dissertation investigates disparities of CRC stage at diagnosis and CRC-specific

survival in Texas during the period from 1995 to 2003. Specifically, it proposes a relative spatial access assessment approach to estimate potential spatial access to CRC prevention and treatment services in Texas (Chapter 4), adopts the generalized equation estimating logistic regression and the adaptive spatial filtering method to evaluate the collective influence of race/ethnicity, SES, geographic location, and potential spatial access to CRC prevention services on CRC stage at diagnosis (Chapter 5), and employs the Kaplan-Meier estimator, the Cox proportional hazard regression, and the spatial scan statistic to uncover the complex disparities of CRC survival (Chapter 6). It is shown that the proposed spatial access approach could effectively overcome the uncertainty problem of gravity-based spatial access models. There were obvious differences in potential spatial access to CRC services by socio-demographic and geographic factors. The investigations into the continuum of CRC reveal systematic disparities in CRC stage at diagnosis and CRC-specific survival by race/ethnicity, SES, and geographic location, with disproportionately unfavorable burdens on non-Hispanic blacks, Hispanics, people from low SES areas, and individuals from specific geographic areas. Potential spatial access to CRC services was found to be associated with CRC stage at diagnosis across the whole state and with CRC-specific survival for non-urban areas. However, the impact of the potential spatial access was minor compared to those of race/ethnicity and SES. This dissertation provides new insights about how CRC disparities accumulate from the diagnosis to mortality in a large and diverse population. The results are useful for CRC disparity elimination and cancer resource allocation in Texas. In addition, this dissertation demonstrates the usefulness of a comprehensive framework of utilizing spatial analysis techniques to complement social epidemiological studies of health disparity.

CHAPTER 1

INTRODUCTION

1.1 Background

Cancer represents a serious threat to the health of the US population. It has claimed many lives and brought about huge economic loss to society. There would be 1,529,560 new cancer cases and 569,490 new cancer-specific deaths in this country in 2010 (American Cancer Society (ACS) 2010). As estimated by the National Institutes of Health (NIH) (2008), in 2008, cancer-related costs totaled $228.2 billion for the United States.

Cancer disparity represents an uneven and unfair distribution of cancer burden among population groups in society. For decades, disparities of cancer incidence, survival, and mortality have been widely documented (Albano et al. 2007; Brawley 2002; Clegg et al. 2002; Correa 2003; El-Serag 2002; Espey et al. 2007; Horner et al. 2009; Singh et al. 2003; Ward et al. 2004). Social groups such as Hispanics, African Americans, and individuals with deprived socioeconomic status (SES) have been systematically experiencing higher cancer risks than others. Cancer incidence and mortality also vary across geographic regions in the United States (Hsu et al. 2007; Lai et al. 2006; Pickens and Orringer 2003; Singh et al. 2003; Sonneveld et al. 1999). These disparities

indicate that there is inequality between disadvantaged groups and other groups in society (Whitehead 1992). Brawley and Freeman (1999) even claimed that this "unacceptable reality" raises "deep ethical and moral questions." Thus, to lessen and to ultimately eliminate cancer disparities have become an important goal of public health administrators and researchers and have been identified by ACS as an overarching theme of its goals (Byers et al. 1999).

Among all types of cancer, colorectal cancer (CRC) deserves special attention. CRC is the type of cancer that originates from the colon or rectum. CRC ranks third in cancer incidence and second in cancer-specific mortality for the US population (ACS 2010). It was estimated that, in 2010, CRC incidence and CRC-specific deaths in the United States would total 142,570 and 51,370, respectively (ACS 2010). In Texas, CRC ranks fourth in cancer occurrence and second in cancer-specific mortality, accounting for about 9,011 new cancer cases and 3,254 cancer-specific deaths each year (ACS, High Plains Division 2008). CRC survival in Texas is below the U.S. average.

CRC survival varies among social groups. Disparities of CRC survival have been widely documented by US researchers over the past two decades. Generally, disadvantaged groups such as some racial/ethnic minorities and individuals with low SES were more probable to experience shorter CRC survival than advantaged groups, even if one controls for stage at diagnosis (Clegg et al. 2002; Elhefni 2006; Govindarajan 2003; Henry et al. 2009a; Kanna et al. 2009; Kirsner et al. 2006; Mayberry et al. 1995; Roetzheim et al. 2000).

1.2 Gaps in Previous Studies

Several limitations of previous studies prevent people from more accurately understanding cancer disparities and designing intervention programs. First, although the influence of multiple factors is important information for cancer disparity elimination, few researchers have investigated the joint impact of SES, race/ethnicity, and geographic location on cancer disparity. Second, although poor access to CRC screening and treatment services has long been suspected associated with low CRC survival rates, few US studies have examined the relation between potential spatial access to medical services and CRC outcomes. Third, there has been little interest in minimizing uncertainties introduced by area-level socioeconomic indicators, spatial dependence, and the small-number problem when analyzing cancer disparities. These limitations will be explicitly explained in Chapter 2.

1.3 Objectives and Research Questions

Utilizing geographic information systems (GISs), spatial analysis methods, and traditional statistical methods, this dissertation aims to comprehensively investigate disparities of CRC survival in Texas. Specifically, it examines disparities of CRC stage at diagnosis and CRC-specific survival from various aspects such as race/ethnicity, SES, geographic location, and potential spatial access to CRC services. This dissertation also seeks to analyze the characteristics of potential spatial access to CRC services in Texas using an improved spatial access model.

The two aims are divided into the following five research questions:

1. Does potential spatial access to CRC-related services differ by race/ethnicity, SES, and geographic location in Texas?

2. Are there any disparities of CRC stage at diagnosis from the aspects of race/ethnicity, SES, geographic location, and potential spatial access to CRC prevention services in Texas?
3. How do race/ethnicity, SES, and potential spatial access to CRC screening services jointly influence CRC stage at diagnosis in Texas?
4. Are there any disparities of CRC-specific survival by race/ethnicity, SES, geographic location, and potential spatial access to CRC treatment service in Texas?
5. How do race/ethnicity, SES, and potential spatial access to CRC treatment services jointly influence CRC-specific survival in Texas?

To answer the research questions, five hypotheses are proposed:

Hypothesis 1: Social groups in Texas have uneven potential spatial accesses to CRC services.

Hypothesis 2: Disparities in CRC stage at diagnosis are statistically significant from the aspects of race/ethnicity, SES, geographic location, and potential spatial access to CRC prevention services in Texas.

Hypothesis 3: SES and race/ethnicity are the major factors influencing disparities in CRC stage at diagnosis in Texas.

Hypothesis 4: Disparities in CRC-specific survival are statistically significant from the aspects of race/ethnicity, SES, geographic location, and potential spatial access to CRC treatment services in Texas.

Hypothesis 5: SES and race/ethnicity are the major factors influencing disparities in CRC-specific survival in Texas.

1.4 Organization of the Dissertation

This dissertation is composed of seven chapters. The first chapter identifies the purposes of this research, outlines the questions that will direct the research, and defines this study's hypotheses. Chapter 2 gives a literature review on cancer disparities and relevant topics such as explanatory factors for cancer disparities, measures of disparities, measures of potential spatial access to medical services, and methods of spatial cluster analysis. The third chapter describes the datasets and methodology used in the entire study. Chapter 4 proposes a relative spatial access assessment approach to estimate potential spatial accesses to CRC services in Texas. Chapter 5 investigates individual as well as joint impacts of race/ethnicity, SES, geographic location, and potential spatial access to CRC prevention services on CRC stage at diagnosis. Chapter 6 analyzes the impacts of various factors on CRC-specific survival. Chapter 7 summarizes the results of the analyses, discusses the contributions and shortcomings of this dissertation, and gives some future research directions.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter reviews previously published works on cancer disparities and some relevant topics. The chapter is composed of five sections. The second section reviews the definition of health disparity. The third section summarizes previous findings about cancer disparities by socio-demographic and geographic factors. This section also reviews explanatory factors for cancer disparities and previous works of cancer disparity elimination. The fourth section describes methodologies of cancer disparity studies. The last section summarizes this chapter and identifies the shortcomings of existing studies.

2.2 Definitions of Health Disparity

Health disparity (also called "health inequality" by some researchers) is an ever-developing concept through which health practitioners, policy makers, and researchers define unjust differences in health among social groups. In its early years during the early 1990s, health disparity was defined as health differences that "are not only unnecessary and avoidable but, in addition, are considered unfair and unjust", based on the assumption that "ideally everyone should have a fair opportunity to attain their full health potential and, more pragmatically, that no one should be disadvantaged from achieving this potential, if it can be avoided" (Whitehead 1992). This definition, initially used in

European countries, referred specifically to health disparities among socioeconomic groups. The definition has been proved effective for researchers to communicate with health administrators and the public in a variety of circumstances (Braveman 2006). In the late 1990s, however, the scope of health disparity research was extended (Anand 2002; Szwarcwald 2002). US researchers began to incorporate other factors (e.g., race/ethnicity, geographic location) in addition to SES to evaluate health disparities. A variety of new definitions (Table 2.1) have been proposed by US health departments and health researchers and these definitions have been systematically reviewed (Braveman 2006; Krieger 2005).

Table 2.1 Definitions of health disparity by US health departments

Source	Definition
National Institute of Health (NIH) (2009)	Health disparities are "differences in the incidence, prevalence, mortality, and burden of disease and other adverse health conditions that exist among specific population groups in the United States."
NIH National Center on Minority Health and Health Disparities (2009)	"A population is a health disparity population if there is a significant disparity in the overall rate of disease incidence, prevalence, morbidity, mortality or survival rates in the population as compared to the health status of the general population."
US Department of Health and Human Services, Healthy People 2010 (2000)	"…… to eliminate health disparities among segments of the population, including differences that occur by gender, race or ethnicity, education or income, disability, geographic location, or sexual orientation. "
National Cancer Institute (NIH), National Center to Reduce Cancer Health Disparities (2005)	"Disparities–or inequalities–occur when members of certain population groups do not enjoy the same health status as other groups. Disparities are most often identified along racial and ethnic lines–showing that African Americans, Hispanics, Native Americans, Asian Americans, Alaska Natives and whites have different disease rates and survival rates. But disparities also extend beyond race and ethnicity. . ." and "can be noted on the basis of income and education"

This study defines health disparity as an inequality in which disadvantaged people systematically suffer worse health and less access to health care than advantaged people (Braveman 2006). Disadvantaged social groups can be characterized by race/ethnicity, SES (e.g., income, education, poverty status), geographic location, age, disability, and

sexual orientation. This definition not only stresses the health differences between social groups but also emphasizes the unfair and the preventable nature of these differences (Braveman 2006; Krieger 2005). This characteristic effectively distinguishes health disparity from differences in health. For example, that men can get prostate cancer but women can not only can be classified as a type of "health difference" rather than "health disparity" because the biological differences are neither unfair nor preventable.

2.3 Cancer Disparity

Cancer disparity can therefore be defined as an inequality in which disadvantaged social groups systematically experience higher risks of cancer incidence, survival, and mortality than advantaged social groups. Prostate cancer, CRC, and lung cancer account for 47% of cancer incidence and 48.4% of cancer mortality for American men (ACS 2010). Cancers of the breast, lungs, colon-rectum, and ovary account for 52.1% of cancer incidence and 55.6% of cancer mortality among American women (ACS 2010). Current cancer disparity studies are therefore primarily focusing on the incidence, survival, or mortality of the cancers of prostate, female breast, ovarian, colon-rectum, and lung. A small number of studies also investigated the disparities of other cancers sites such as cervix, stomach, and liver (Correa 2003; Espey et al. 2007).

Generally, most US studies of cancer disparity are focusing on race/ethnicity (racial/ethnic minorities versus whites) and SES (low SES versus high SES). A small number of studies are also examining the impact of geographic location (high-risk areas versus low-risk areas) on cancer occurrence, survival, and mortality. A list of previous cancer disparity studies can be found in Table 2.2.

Table 2.2 Facts of cancer disparities in the United States

Source	Purpose	Cancer Site(s) and Outcomes, Study Period	Geographic Location and Scale	Factors	Conclusion
Albano et al. 2007	To examine the influence of race and education on cancer mortality	Mortality by cancers of the female breast, lung, prostate, and colon, 2001	United States, National level	Race and SES	Cancer mortality varies substantially by education level.
Bradley et al. 2001	To disentangle the influence of race and SES on breast cancer disparities	The incidence and survival for the cancers of female breast, cervix, lung, prostate, and colon, 1996-1998	The state of Michigan, state level	Race and SES	Disparities in cancer outcomes may be greater than expected.
Chu et al. 2007	To examine cancer mortality risks by race/ethnicity, SES, and time	The mortality of carcinomas of female breast, colorectal, cervix, lung, and prostate, 1990-2000	United States, County level	Race and SES	Increases of the racial/ethnic disparity in SES groups vary by cancer sites.
Field et al. 2005	To investigate disparities of breast cancer survival within an insured population	The survival of breast cancer, 1993-1998	United States. National level	Race and SES	Good access to medical services did not reduce the survival disparity among women diagnosed with invasive breast cancer.
Grann et al. 2005	To examine the influence of various factors on breast cancer survival	The survival of breast cancer, 1990-2001	11 geographic areas. National level.	Race, SES, and Geographic location	Breast cancer mortality varies by SES, race, and geographic region.
Haas et al. 2008	To assess how residential segregation mediate racial/ethnic disparities of breast cancer	The mortality of breast cancer, 1992-2002	United States, National level	Race/ethnicity	For seniors, segregation accounts for part of the racial disparity of breast cancer care. However, it did not mediate racial disparities of breast cancer mortality.
Hershman et al. 2005	To evaluate how treatment mediate the racial disparity of breast cancer survival	The survival of breast cancer, 1996-2001	Detroit, MI, city level	Race	Many women diagnosed with early stage discontinued the chemotherapy. This behavior was more common in black women and women from poor areas.
Hsu et al. 2004	To evaluate the geographic	Breast cancer mortality, 1990-	Texas, County level	Race, Geographic	There was no significant hot-spots

Table 2.2-Continued

	variation of breast cancer mortality for different racial groups in Texas	2001		location	nor spatial-temporal patterns of breast cancer mortality.
Hsu et al. 2006	To examine CRC mortality among different Texas demographic groups	Colorectal cancer mortality, 1990-2001	Texas, County level	Race/ethnicity, geographic location	There were significant disparities of CRC mortality among demographic subpopulations in Texas. Counties with consistently higher CRC mortality risks were also detected.
Jemal et al. 2002	To determine clusters of prostate cancer mortality and their association with selected regional characteristics	Prostate cancer mortality, 1970-1994	United States, County level	Race/ethnicity, SES, and geographic location	The clusters were not associated with any of the selected characteristics.
Krieger et al. 1997	To simultaneously examine cancer incidence in relation to social class and race/ethnicity	The incidence of the cancers of lung, prostate, colon, breast, and cervix, 1988-1992	San Francisco Bay Area, CA, census block group level	Race/ethnicity, SES	There is no easy generalization for the racial and socioeconomic disparities of cancer.
Krieger et al. 2006	To test if the socioeconomic gradients of breast cancer incidence is narrowing, and how the decline varies among racial/ethnic groups	Breast cancer incidence, 1978-2002	Los Angeles County and the San Francisco Bay Area, CA and Massachusetts	Race/ethnicity, SES	The socioeconomic disparities of breast cancer incidence changed but the changes vary by race/ethnicity.
Liu et al. 2001	To examine socioeconomic disparities of prostate cancer incidence	Incidence of prostate cancer, 1972-1997	Los Angeles CA, city level	SES	The change of the socioeconomic disparities of prostate cancer incidence might be partly related with the prevalence of PSA screening in high SES social groups.
McDavid et al. 2003	To examine cancer survival disparity based on individual	Survival from the four major cancers, 1995-1998	Lexington, KY, city level	SES	Health insurance status is related with disparities in cancer survival.

Table 2.2-Continued

	health insurance information				
Merkin et al. 2002	To examine the relation between race, SES, and breast cancer diagnosis stage in New York City	Incidence of breast cancer, 1986-1995	New York City, zip code level	Race and SES	Race and SES were the independent factors for the disparities.
Roetzheim et al. 2000	To examine if insurance status and race is related with breast cancer outcomes	Mortality from breast cancer, 1997	Florida, Individual level	Race and SES	Breast cancer patients with limited insurance had higher risk of mortality.
Schwartz et al. 2003	To explore if racial disparities of cancer stage at diagnosis is related with differences in SES	Incidence of the five major cancers, 1988-1992	Detroit, MI. Block group level	Race and SES	SES is the major factor influencing cancer stage, although biological factors partly account for the racial disparities.
Singh et al. 2002	To examine socioeconomic disparities of all-cancer mortality for men in the United States	The mortality of all cancers among men, 1950-1998	United States, County level	SES	Socioeconomic disparities of cancer mortality among US men have reversed during the study period.
Singh et al. 2002	To examine socioeconomic disparities of mortality in lung and colorectal cancer	Mortality from lung and colorectal cancer, 1950-1998	United States, County level	SES	Socioeconomic disparities of CRC mortality have narrowed during the study period.
Singh et al. 2003	To analyze disparities across the whole cancer continuum	Incidence, stage at diagnosis, survival, mortality, and treatment for all cancer sites and for the lungs, colon, breast, cervix, prostate, and melanoma, 1975-1999	United States, county and census tract level	SES	Socioeconomic disparities of the incidence and mortality of cancers differ greatly by gender, race/ethnicity, and study time.
Singh et al. 2004	To analyze socio-temporal inequalities in cervical cancer outcomes in the United	Cervical cancer incidence, stage at diagnosis, survival, and mortality, 1975-2000	United States, County and Census Tract level	SES	Socioeconomic disparities of cervical cancer exist in the US.

Table 2.2-Continued

	States				
Ward et al. 2004	To investigate cancer disparities from race/ethnicity and SES	Incidence, stage at diagnosis, survival, mortality, and treatment of cancer, 1975-2000	United States. National level	Race and SES	There were racial/ethnic and socioeconomic gradients for all cancers combined in the United States.
Yood et al. 1999	To examine if access to health care mediates racial disparities in breast cancer survival	Breast cancer survival, 1986-1996	Detroit, individual level	Racial and SES	Among women with comparable access to healthcare, racial disparities in breast cancer diagnosis state still exist.
Yost et al. 2001	To evaluate socioeconomic disparities of breast cancer incidence for race/ethnic groups	Breast cancer incidence, 1988-1999	California, State level	Race/ethnicity	SES is significantly related with breast cancer incidence, and this association is stronger for Asians and Hispanics than for other racial groups.

2.3.1 Race/ethnicity

Race/ethnicity is an important factor for cancer disparity studies. Cancer disparities between racial/ethnic minorities and white people have been widely documented.

Most US studies have focused on cancer disparities between African Americans (or blacks) and Caucasians (or whites). African Americans are the largest racial minority of this country, comprising more than 12% of the population (US Census Bureau 2000). National studies revealed that African Americans bear the highest cancer risks among all racial/ethnic groups (ACS 2010; Horner et al. 2009; Ward et al. 2004). Compared to white people, African Americans have higher risks of being diagnosed with and dying from the carcinomas of colon-rectum, liver, prostate, and uterine cervix (Ward et al. 2004). The burden of breast cancer mortality is heavier for African American women than for white women. After being diagnosed with prostate cancer, African American men are less likely to attain five-year survival than white men (Clegg et al. 2002).

Cancer disparities between other minority groups and whites are not as serious and long-standing as those between African Americans and whites. However, these disparities still deserve attention because the populations of some minorities are rapidly increasing. Studies revealed that, although Hispanics and Asian Americans have an overall lower incidence rate of cancer, they bear higher incidence rates of liver and stomach cancers than other racial/ethnic groups (El-Serag 2002; Ward et al. 2004). For example, the incidence rate of stomach cancer is 75% higher for Hispanic people than for non-Hispanic white people. Hispanic women are 40% more probable to die from cervical cancer than non-Hispanic white women. In addition, the probability of surviving from female breast cancer is lower for Native Americans than for any other racial/ethnic group (Bach et al. 2002; Chu et al. 2003; Clegg et al. 2002).

Racial/ethnic disparities of cancer have changed over time. For example, age-adjusted mortality rates of female breast cancer were almost equal between black and white women during the period from 1970 to 1985. However, black women began to suffer higher mortality rates of breast cancer since 1985. National data indicate that, although there was diminishing prostate cancer mortality, the rate was decreasing twice as rapidly for whites and Asians than for other racial/ethnic groups (ACS 2008). These findings further demonstrate that cancer disparities are changeable and preventable.

Racial/ethnic disparities of cancer vary when racial/ethnic groups are categorized by sex, age, or immigration status. For example, American women experience a higher lung cancer survival rate than men (Cerfolio et al. 2006). Although black women younger than forty are more likely to develop breast cancer than white women, black women who are forty-nine or older have a lower probability of getting breast cancer than their white

counterparts (Bigby and Holmes 2005). Among the Asian American population, the incidence rate of invasive cervical cancer is four times higher for Vietnamese women than for the rest (ACS 2009). These facts defy a simple and generalized model of racial/ethnic disparities in cancer. A more complicated explanation of how specific factors influence racial/ethnic disparities of cancer is needed.

It has been believed that biological factors are important causes for racial/ethnic disparities of cancer. Some researchers claimed that differences in cancer outcomes stem from genetic differences rather than other factors (Brown 1999; Freeman 1998). However, some other studies indicate that the genetic influence is of negligible importance when analyzing racial/ethnic disparities of cancer (Brawley 2002). In fact, African American women and white women exhibited no obvious differences in the mortality rate of breast cancer before 1980 (Brawley 2002). Clinical trials and institute-specific treatments demonstrate that cancer outcomes are similar for patients in equal medical conditions, regardless of race/ethnicity (Dignam 2000; Yood et al. 1999). These findings have inspired researchers to focus less on genetic issues and more on social factors (e.g., affluence, education, health insurance coverage, and residential area) that may influence cancer (Goel et al. 2003).

2.3.2 Socioeconomic status

SES is the socioeconomic position of a person or a neighborhood in a specific region. SES can be measured from various aspects, including education, occupation, income, wealth, health insurance status, and home ownership (Krieger et al. 1997; Singh et al. 2003). Composite indicators such as the principal component and the Townsend Index have also been proposed to represent SES (Krieger et al. 2002; Krieger et al. 2003; Singh

et al. 2002). Generally, these composite indicators are calculated from multiple single indicators according to a specific algorithm. Associations between SES indicators and cancer outcomes have been observed for different types of cancer.

Affluence

Income, poverty, and employment status have been widely used to represent the affluence of an individual or a social group (Brawley 2002; Horner et al. 2009; Ward et al. 2004). Previous studies have revealed significant associations between these variables and cancer outcomes (Haynes and Smedley 1999; Horner et al. 2009). People from poor areas have been suffering higher risks of late stage diagnosis, lower rates of five-year survival, and higher mortality rates for most cancer sites than people from affluent areas (Brawley et al. 2002; Schwartz et al. 2003; Singh et al. 2003; Ward et al. 2004). The association between affluence and cancer varies according to the specific cancer site, the study region, and race/ethnicity.

Education

Educational attainment represents another important dimension of SES. Education level is also associated with cancer outcomes, although with a much weaker association compared to those between affluence indicators and cancer (Krieger et al. 1999). Lower educational achievement is related with higher mortality rates of lung cancer, breast cancer, and prostate cancer (Albano et al. 2007; Liu et al. 1998; Steenland et al. 2002). Since 1985, the US government has been recording educational attainment in death records. The availability of individual educational information has made education level a popular SES indicator for cancer disparity studies, especially those related with cancer mortality disparity (Albano et al. 2007; Singh et al. 2002).

Health insurance

Health insurance status is often used to approximate SES in cancer disparity studies. Health insurance status has been reported associated with cancer outcomes. Generally, uninsured individuals and people covered by Medicaid have higher probabilities of being diagnosed with a late stage cancer and dying from cancer than those with commercial insurances (ACS 2010; Bradley et al. 2001; Lee-Feldstein et al. 2000; McDavid et al. 2003; Roetzheim et al. 2000).

Using health insurance coverage to represent SES has two primary advantages: the accuracy of the results and the usefulness for cancer disparity elimination. By linking cancer registry cases or death certificates to an insurance enrollment database, one can derive the health insurance information for individual patients. Individual data can yield more accurate results of cancer disparity than community-level data (Grann et al. 2005). In addition, the relation between cancer and health insurance coverage is very useful information for health professionals. Based on this information, health professionals could easily target the vulnerable population and make corresponding policies to benefit uninsured individuals. However, a disadvantage of using health insurance to represent SES is that not all health insurance databases are publicly available. Public databases are only available for Medicaid, Medicare, and a limited number of private insurance providers. As a result, studies based on these datasets are only applicable to specific social groups (i.e., poor people and old people) and cannot be extended to all populations.

2.3.3 Geographic location

Geographic variation represents another critical aspect of cancer disparity. It is clear that cancer risk is not evenly distributed across the space. Studies have revealed

significant geographic disparities of cancer outcomes at various geographic scales (Hsu et al. 2007; Lai et al. 2006; Pickens and Orringer 2003; Singh et al. 2003; Sonneveld et al. 1999).

Knowledge about the geographic distribution of cancer is important because it not only provides a straightforward way for examining cancer disparities but also illuminates factors through which further interventions are possible. Social groups with obvious geographic characteristics (e.g. residential areas) can be more easily targeted than social groups with socioeconomic (e.g. low income) or racial/ethnic (e.g. African Americans or Hispanics) characteristics. After identifying areas with abnormally high cancer risks, researchers can examine the influence of possible factors (racial/ethnic, socioeconomic, environmental, etc.) for the elevated cancer risk in these areas. Furthermore, health administrators can use this information to design strategies for cancer prevention and medical resource allocation (Bradley et al. 2002; Woods et al. 2006).

2.3.4 Cancer disparities by multiple factors

It is well established that cancer disparities are not caused by a single factor. Rather, the incidence, treatment, survival, and mortality of cancers are influenced by interactions among multiple factors (Figure 2.1) (Grann et al. 2005; Ward et al. 2004). This interaction makes it difficult to discern the true reason(s) for disparities. For example, the disadvantageous situation of blacks in cancer outcomes may result from the lower SES of blacks or from the fact that blacks tend to live in areas with high concentrations of carcinogens. Instead of examining cancer disparities from a single factor, more and more researchers are beginning to evaluate the impacts of multiple factors on cancer (Bradley et al. 2002; Brawley 2002; Chu et al. 2007).

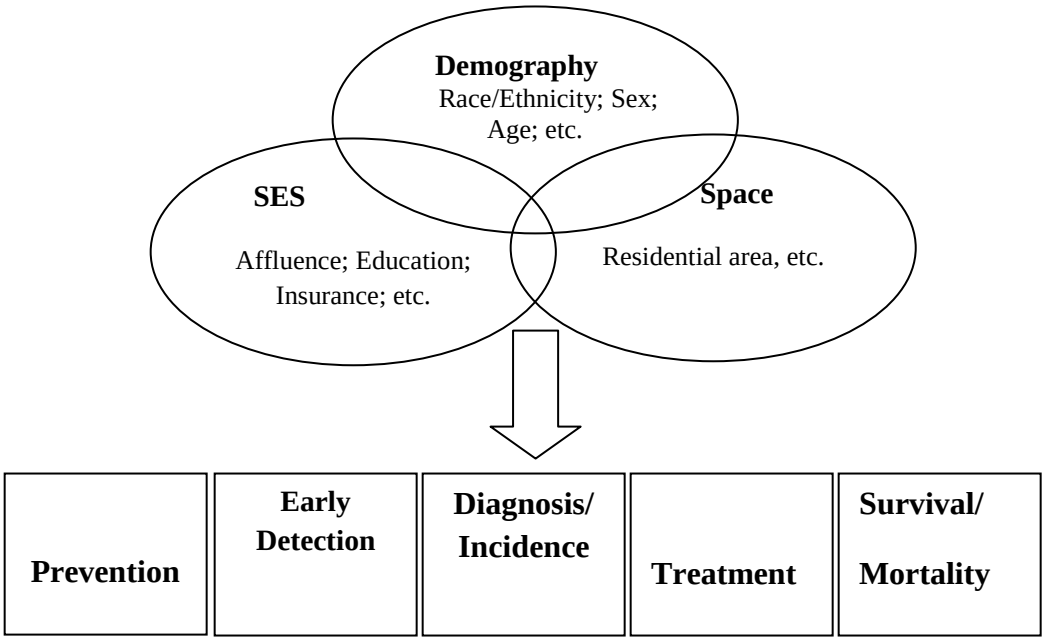

Figure 2.1 Factors that influence the continuum of cancer
(Adapted from Ward et al. (2004))

Cancer disparity studies with the focus on multi-factor interactions primarily investigate the interactive influence of race/ethnicity and SES. Generally, such studies disentangle the effect of SES and race/ethnicity by testing the effect of one factor while controlling for the other (Bradley et al. 2001; Bradley et al. 2002; Brawley 2002; Chu et al. 2007; Dignam 2000; Ward et al. 2004). For example, in an analysis of the disparity of breast cancer survival, Bradley et al. (2002) limited their subjects to black and white women who had similar health insurance status. These studies showed that SES could explain a large part of the racial/ethnic disparity, although race/ethnicity has independent influence on cancer (Brawley 2002; Chu et al. 2007; Grann et al. 2005; Ward et al. 2004).

Despite the prevalence of studies examining the interactions of SES and race/ethnicity, few studies have analyzed how SES, race/ethnicity, geographic location,

and other factors jointly influence cancer outcomes. Jemal et al. (2002) analyzed the national distribution of prostate cancer mortality from 1970-1989 to determine if high mortality rates were related with SES and race/ethnicity. Their results showed that none of the factors could explain the spatial clusters for whites and blacks. This conclusion was weak, however, because their analysis was carried out at a coarse geographic scale (i.e., county-level) and the accuracy of their result could be tainted by ecological fallacy and the modifiable area unit problem. Roche (2002) used a GIS to detect spatial clusters of distant stage breast cancer at a finer scale (census tract) and linked the clusters to local demographic indicators. This study found that areas with higher-than-expected risks of invasive breast cancer were characterized by elevated proportions of minority women and households with language barriers. Henry et al. (2009) examined the geographic disparities of CRC diagnosis stage and CRC survival in New Jersey and found that spatial clusters of CRC late stage diagnosis and CRC survival primarily reflect the geographic distribution of SES and racial/ethnic groups.

Generally, cancer disparity studies of the joint impact of race/ethnicity, SES, and geographic location follow a two-step procedure (Jemal et al. 2002; Roche 2002). The first step determines areas with significantly different cancer risks using cluster detection approaches. This step involves the detection of both areas with higher-than-expected risks and areas with lower-than-expected risks. The second step compares non-spatial factors (e.g., SES, demographic structure, environmental exposure) between the two categories of areas to look for possible explanations for the cancer risk differences. Approaches for cluster detection and for the examination of non-spatial factors may vary with the goal of the researcher, the geographic scale of data observation, and some other issues.

2.3.5 Explanatory factors for cancer disparities

Researchers analyze cancer disparities from race/ethnicity, SES, and geographic location because these factors allow easy categorization of the population and effective communication of the results to the public. Race/ethnicity, SES, and geographic location are not the direct causes for cancer outcomes, but, they influence cancer through some intermediate factors.

Explanatory factors for disparities of cancer incidence

Despite the rapid advancement of medical and biological sciences and technologies today, why some people get cancer and others do not is still unclear. It was hypothesized that cancer incidence could be influenced by behavioral and environmental factors, both of which vary by SES, race/ethnicity, and geographic location (Goy et al. 2007; Ward et al. 2004). Behavioral factors include cigarette smoking, dietary patterns, reproductive and sexual histories, physical inactivity, and mental stress (Krieger et al. 1999). Racial/ethnic minorities and people from low SES areas are more easily impacted by the negative side of these factors than others. For instance, the high rates of lung cancer incidence of some low-SES neighborhoods could be associated with the smoking habits, which are somewhat linked to the marketing strategies of tobacco companies in these neighborhoods (Ward et al. 2004). Over-consumption of animal fat and inadequate physical activity, which could increase the risk of getting certain cancers, are also more common among some racial/ethnic groups than among whites (Ward et al. 2004).

Environmental factors include exposure to carcinogens and radiation. Communities with low SES or high proportions of racial/ethnic minorities are more likely to be in close proximity to environmentally degraded areas such as landfill sites, industrial areas,

highway-affected areas, or other locations where harmful substances (e.g., some volatile organic compounds (VOCs)) are highly concentrated (Apelberg et al. 2005). Long-term exposure to harmful substances will likely lead to health problems.

Explanatory factors for disparities of cancer survival

Stage at diagnosis and treatment delivery are the two major factors impacting cancer survival (Siminoff and Ross 2005; Wardle et al. 2005). Cancer patients diagnosed at early stages tend to live longer than those diagnosed at late stages, because early-stage cancers are relatively more controllable than advanced ones (Wardle et al. 2005). High-quality and in-time treatment can also help extend a patient's life and lower the chance of dying from cancer (Siminoff and Ross 2005).

Stage at diagnosis and medical treatment could be influenced by access to medical services (Baldwin et al. 2005; Chandra and Skinner 2003; Ciccone et al. 2000). Access to medical services refers to a person's ease of obtaining health-care services that can bring about the best possible health outcome (Cockerham 1998; US Department of Health and Human Services 2000). Joseph and Philips (1984) classified access to health services into potential access and realized access (or utilization). Potential access to health service refers to a population group or individual's ease of accessing the service based on existing conditions but does not ensure use of the service (Joseph and Phillips 1984). Realized access, based on potential access, reveals how people actually use the service. One can further classify both types of access into spatial and non-spatial access based on how health-care accessibility is influenced by spatial factors (for example, spatial location and travel distance) and non-spatial factors (for example, SES, health insurance status, and cultural background) (Aday and Andersen 1974).

Access to medical services may affect cancer stage at diagnosis and treatment in two aspects. On one hand, adequate access to cancer screening services can increase the frequency of screening among vulnerable populations, thus increasing the probability of early-stage diagnosis of cancer (Lee-Feldstein et al. 2000; McDavid et al. 2003). On the other hand, access to cancer treatment services can determine the type, quality, and duration of cancer treatment, thus influencing the treatment outcome.

Access to cancer services can be influenced by structural and personal factors. Structural factors, such as the spatial locations of medical facilities, are determined by the medical system and health-care providers (Shavers et al. 2002). Personal factors include patients' capability to pay for medical treatment, transportation, cultural background, and English proficiency (Shavers et al. 2002; Siminoff and Ross 2005).

Prior studies revealed substantial differences in potential and realized access to cancer services by SES, race/ethnicity, and geographic location in the United States (Baldwin et al. 2005; Chandra and Skinner 2003; Ciccone et al. 2000; Onega et al. 2008; Wang et al. 2008). For example, Onega et al. (2008) found significant regional differences in potential spatial accesses to cancer services among US regions. Ciccone et al. (2000) found that Hispanics, blacks, and people from poor areas were less likely to undergo screening tests than advantaged groups (e.g., whites, people with high SES). Disadvantaged groups also tend not to receive timely and appropriate treatment for their illnesses (Shavers and Brown 2002).

The associations between potential spatial access to cancer-related services and cancer survival have also been analyzed (Dejardin et al. 2005; Dejardin et al. 2008; Huang et al. 2009; Jones et al. 2008; Wang et al. 2008). However, most researchers were

focusing on cancer stage at diagnosis rather than the actual survival from cancer. For example, Jones et al. (2008) analyzed the relation between travel time to medical service locations and cancer diagnosis stage in Northern England and found a positive correlation between the two. Wang et al. (2008) also observed a significant relation between potential spatial access to primary health care and breast cancer diagnosis stage in some areas of Illinois. Huang et al. (2009) found that increasing distance to mammography facilities could raise the probability of being diagnosed with late stage breast cancer in rural Kentucky. These studies indicated a complicated relation between potential spatial access to medical services and cancer survival. This relation is influenced by multiple factors such as the measure of accessibility (e.g., travel time, travel distance, or gravity-based measure), the type of medical service (e.g., general practitioner, primary care, or screening facilities), the type of cancer, and characteristics of the study area (e.g., urban or rural).

2.3.6 Studies of cancer disparity elimination

Efforts to eliminate cancer disparities have increased over the past decades. Generally, such efforts try to enhance the use of cancer screening services and appropriate treatment for disadvantaged social groups and patients. Models, such as the Markov model (Schleinitz et al. 2005) and the behavioral model for vulnerable populations (Bazargan et al. 2004; Jandorf et al. 2005), have been proposed to predict the use of cancer-related services using social, economic, geographic, and cultural variables. Community intervention programs such as the Patient Navigator (Lisovicz et al. 2006) have also been developed to address cancer survival disparities. Most programs have been implemented in small areas to eliminate personal and clinical barriers rather than

more extensive structural barriers. To overcome the structural problems, however, national- or state-level interventions are needed.

Compared to efforts to eliminate cancer survival disparities, efforts to address cancer incidence disparities are rare. One reason for this situation is that the relation between various factors (e.g., biological, behavioral, emotional, environmental) and cancer is still poorly understood. Projects to address disparities in cancer incidence are limited to cancer-awareness education and vaccination programs among high-risk populations (Bigby et al. 2003; Gilewski et al. 2000; Lisovicz et al. 2006; Ma et al. 2006).

2.3.7 Disparities of colorectal cancer survival

Most studies of CRC survival disparity examined the disparity from either SES or race/ethnicity, but not both. Few studies have tried to assess the relative importance of race and SES on CRC survival (Du et al. 2007; Gomez 2007). A study of a large cohort of senior CRC patients revealed that differences in SES accounted for a large part of racial disparities of CRC survival (Du et al. 2007). Medical experiments on CRC cancer treatment supported this by demonstrating that equal treatment lead to equal outcomes in CRC (Dignam et al. 2000; Hassan 2009).

While most studies focus on race/ethnicity and SES, few have examined the geographic disparities of CRC survival. Only two US studies have examined the geographic patterns of CRC survival (Henry et al. 2009a; Huang et al. 2007). The first study was conducted in California (Huang et al. 2007) and the next was conducted in New Jersey (Henry et al. 2009a). Both of them found disproportionate distribution of CRC survival in the study regions.

2.4 Methodologies for Cancer Disparity Research

2.4.1 Measuring racial/ethnic and socioeconomic disparities of cancer

How can one accurately measure the disparity? Methods (Table 2.3) have been proposed to calculate health disparities based on different assumptions of the disparity. Generally, measuring health disparity involves comparing a health indicator (e.g. incidence rate) between a disadvantaged group and a reference group. The reference group can be represented by the most-advantaged group such as the racial/ethnic majority and people with the highest SES.

Table 2.3 Health disparity measures used by past studies

Disparity Measure	**Definition**
Rate Ratio	$RR = h_1/h_2$, where h_1 and h_2 are the health indicators for the least and the most advantaged groups, respectively.
Index of Disparity	$IDisp=\sum \left(\frac{\lvert h_i - h_{ref}\rvert}{I} - 1\right)/h_{ref} \times 100$, where h_i and h_{ref} are the health indicators of the i^{th} group and the reference group, respectively
Relative Concentration Index	$RCI=\left(\frac{2}{\mu}\right) \times [\sum p_i h_i X_i] - 1$, where μ represents the mean indicator of health for the study population, p_i represents the share of population group *i*, h_i denotes the health indicator of group *i*, and Xi is the rank of group *i*.
Theil Index (TI) and Mean Deviation (MD)	$TI=\sum p_i r_i \ln r_i$ and $MD=\sum p_i[-\ln r_i]$ where p_i is the share of population group *i* and r_i is the ratio of group *i*'s health indicator relative to the population average indicator.
Rate Difference	$RD = h_1 - h_2$, where h_1 and h_2 are the health indicators of the least advantaged and the most advantaged groups, respectively.
Between-Group Variance	$BGV=\sum p_i(h_i - \mu)^2$, where μ represents the mean indicator of health for the study population, p_i denotes the share of population group *i*, and h_i represents the health indicator of group *i*.
Absolute Concentration Index	ACI=μRCI, where μ represents the mean indicator of health for the study population and RCI denotes the relative concentration index.
Individual Mean Difference (IMD)	$IMD(\alpha, \beta) = \frac{\sum(h_i - \mu)^c}{n\mu^\beta}$, where h_i denotes the health indicator of group *i*, μ represents the mean indicator of health for the study population, and α and β denote the significance parameters.
Inter-Individual Difference (IID)	$IID(\alpha, \beta)=\frac{\sum\sum(h_i - h_j)^2}{2\alpha^2\mu^\beta}$, where h_i and h_j denote the health indicators of groups *i* and *j*, respectively, μ represents the mean indicator of health for the study population, and α and β denote the significance parameters.

Rate ratio and rate difference are two common measures for quantifying the disparity between two groups. However, both measures are limited in that they only reflect the disparity between two groups and cannot be expanded to more than two groups. Complex methods have therefore been proposed to measure the health disparity among multiple groups (Mackenbach and Kunst 1997; Wagstaff et al. 1991).

Health disparities revealed by different measures are inconsistent. Previous studies have found substantial disagreement among different measures in the magnitude and direction of health gradients (Harper et al. 2008; Mackenbach and Kunst 1997; Wagstaff et al. 1991). This disagreement might stem from the varying assumptions of the disparity among different measures. It is vital to know the advantages and shortcomings of each measure before adopting one for health disparity analysis. However, few studies have comprehensively compared the performances of all health disparity measures.

Health disparity can also be measured using statistical inference. Regression has been widely employed to investigate the influence of different factors on health. Using regression methods, one can estimate how cancer mortality and incidence (used as the dependent variable), for example, is related to SES (used as the independent variable). Compared to the numerical methods mentioned above, regression methods can provide information regarding the statistical significance of their results. For example, user-defined confidential intervals (e.g., 95%) are always employed to assess the significance of logistic regression results.

2.4.2 Measuring geographic disparities of cancer

Researchers can analyze the geographic disparities of health events using two categories of methods: cluster-detection methods and clustering-detection methods.

Cluster-detection methods differ from clustering-detection methods in that the former tries to detect areas with unusually high likelihood of an event (i.e. disease incidence or mortality) (Besag and Newell 1991), whereas the latter aims to find tendencies of clustering across the whole study area. Since the former fits the purpose of this research, only cluster-detection methods are discussed here.

A cluster-detection method can be either area-based or point-based, depending on the scale at which the disease event is observed. Area-based methods are preferred when health events are aggregated at area units. Commonly-used area-based cluster-detection methods include the Poisson model of the spatial scan statistic (Kulldorff 1997), local Moran's method (Anselin 1995), the Besag-York-Mollie model (Besag et al. 1991), and the generalized additive method (GAM) (Hastie and Tibshirani 1987). Most area-based methods determine the presence of clusters by statistical inference. For example, if conterminous areas are inferred to have significantly higher disease risk than their neighbors, these areas are defined as a local cluster. Area-based cluster-detection methods are suitable for ecological studies because clusters can be then conveniently linked with area-level socio-demographic information. This linkage enables analysis of the influence of SES, demographic factor, and spatial location on health.

Point-based methods determine the presence of clusters based on the locations of individual cases and their background population. These methods identify clusters by addressing two issues: to measure the regional density of the event and to adjust for the background population which is unevenly distributed. The local density of the event can be measured in two ways: using kernel-based methods or distance-based methods (O'Sullivan and Unwin 2003). Kernel-based methods estimate the intensity of events

within a predefined areal unit, or a kernel. The kernel is moved over the study region in a controlled way to evaluate the distribution of disease intensity across the whole region. Typically, three factors need to be considered when controlling for background population for kernel-based methods: the kernel size, the background population size, and the number of case events. Researchers can adjust the kernel size and test the relation between case number and the size of the background population (Fotheringham and Zhan 1996; Openshaw et al. 1987; Rushton and Lolonis 1996). Distance-based methods determine clusters by comparing the distances between cases and controls. Examples of distance-based methods include the nearest neighbor model (Cuzick and Edward 1990) and the Bernoulli-based spatial scan statistic (Kuldorff and Nagarwalla 1995).

2.4.3 Measuring potential spatial access to healthcare

As discussed in Section 2.3.5, potential spatial access to cancer care services is the basis for cancer service utilization. Potential spatial access to medical services is primarily influenced by three factors: supply of medical services, population demand for the services, and travel costs between the demanding populations and medical sites. Common potential spatial access measurements include the regional availability method (Khan 1992), kernel density models (Guagliardo 2004; Silverman 1986), and gravity-based models (Joseph and Bantock 1982; Luo and Qi 2009; Luo and Wang 2003a; Schuurman et al. 2010).

The regional availability method

A basic method for evaluating potential spatial access to health services is the regional availability method, which compares the sum of health care capacity and the population demand within an area. This method provides a straight-forward view of

spatial access and was adopted by US health departments as a critical criterion for identifying whether or not a social group or geographic area is medically underserved (Lee 1991; Ricketts et al. 2007; US General Accounting Office (GAO) 1995). The regional availability measure, however, has long been criticized for its two problematic assumptions: (1) people are restricted to one area and do not go beyond the area to seek health care, and (2) all individuals within an area have equal access to the service, regardless of how far away they live or work from healthcare sites. These assumptions are hardly valid in real situations (GAO 1995; Wing and Reynolds 1988).

The kernel density model

The kernel density model is based on a kernel density function (e.g., Gaussian function) which can estimate a smooth density surface for a point in the two-dimensional space (Silverman 1986). Generally, a kernel density model is composed of two steps. The first step generates a supply surface from the supply sites and a demand surface from the population centroids by using the kernel function. The second step divides the supply surface by the demand surface to derive a supply-to-demand ratio for each pixel within the two-dimensional space. The supply-to-demand ratio is then used to represent the potential spatial access.

The kernel density model is better than the regional availability measure in estimating potential spatial access to medical services because it considers both the distance-decay effect and the cross-boundary health care seeking behaviors. However, it is still limited in two aspects. First, it uses the straight line distance, that is, the radius of circles, in the calculation of density surfaces, which is an inferior indicator of travel cost compared to the road network distance or road network travel time (Wang and Minor

2002). Second, the kernel density function is not a good model for estimating medical supply and demand. For example, it may mistakenly "assign" medical services to non-populated areas such as big lakes and forests (Yang et al. 2006). In addition, the kernel density function assumes the population density peaks at the centroid point and attenuates along the radius. However, the population distribution does not necessarily follow that way. For example, the population might be homogeneous across the whole region or the peak might not locate at the centroid at all. The first problem of the kernel density model can be solved by introducing a road network kernel density scheme which distributes the service over a road network instead of over the two dimensional space (Borruso 2005). However, the other problems remain.

The gravity model

The gravity model estimates potential spatial access to medical services according to the Law of Gravitation (Joseph and Bantock 1982). The basic gravity model can be stated as

$$A_i^G = \sum_{j=1}^{n} \frac{S_j f(d_{ij})}{\sum_{k=1}^{m} P_k f(d_{kj})} \qquad 2.1$$

where A_i^G is the gravity-based spatial access for population location *i*, S_j represents the medical capacity of any medical site *j* , P_k denotes the population size of any population location *k*, and d_{ij} is the travel cost from *i* to *j*. *m* and *n* represent the numbers of population locations and supply sites, respectively. The geographic impedance function, $f(d)$, determines how travel distance influences the accessibility. According to Kwan (1998), the most common forms of $f(d)$ are the exponential function ($f(d) = e^{-\beta d}$), the inverse-power function ($f(d) = d^{-\beta}$), and the Gaussian function ($f(d) = e^{-d^2/\beta}$), where β is the impedance coefficient indicating the extent of distance-decay.

The basic gravity model assumes the potential spatial access to health service of a population location equals the sum of impedance-weighted supply-to-demand ratios of all nearby medical sites. This model is conceptually more complete than previous methods but difficult to understand (Luo and Qi 2009).

The floating catchment area methods

An improvement of the basic gravity model is the two-step floating catchment area method (2SFCA), which was initially designed by Radke and Mu (2000), subsequently improved by Luo and Wang (2003b), and recently enhanced by Luo and Qi (2009). The basic 2SFCA model works in two steps. The first step is to generate a driving time zone (or catchment) with a pre-defined threshold travel time (d_0) for each medical service site *j*, searching all area units inside the catchment, and computing the supply-to-demand ratio, R_j, for *j* by

$$R_j = \frac{S_j}{\sum_{k \in (d_{jk} \leq d_0)} P_k} \qquad 2.2$$

where S_j is the medical capacity of medical site *j*, P_k represents population size of area unit *k* whose centroid falls inside the catchment of *j* ($d_{kj} \leq d_0$), and d_{jk} represents the traveling cost between *j* and *k*. The second step is to generate the catchment for each area unit *i* with d_0 as the threshold travel time, searching all medical service sites inside the catchment, and adding up the supply-to-demand ratios of these service sites by

$$A_i^F = \sum_{l \in (d_{il} \leq d_0)} R_l \qquad 2.3$$

where A_i^F is the spatial access of population at area unit *i*, R_l represents the supply-to-demand ratio of service site *l*, and d_{il} represents the traveling cost between *i* and *l*.

The 2SFCA method stems from the basic gravity model but expresses the basic model in a more intuitive way. It first estimates the demand for each medical site and

calculates the supply-to-demand ratio according to its medical capacity and local demand. The second step summarizes the supply-to-demand ratios of nearby medical sites for each population. Both steps are easy to interpret and convenient to realize using GIS. The 2SFCA method has been employed to estimate potential spatial access to a variety of medical services (Albert and Butar 2005; Cervigni et al. 2008; Guagliardo 2004; Langford and Higgs 2006; Wang 2007; Wang et al. 2008; Yang et al. 2006). This method, however, is still limited because it characterizes all area units inside a same catchment with equal access and disregards the distance impedance within the catchment (Luo and Wang 2003).

An improved version of 2SFCA (Luo and Qi 2009), the enhanced 2-step floating catchment area (E2SFCA) method, has been proposed to solve the limitation of the basic 2SFCA method. Briefly, E2SFCA divides the catchments into several subzones and assigns a distance-based weight for each subzone to simulate the within-catchment distance impedance. The E2SFCA method will be described in detail in Chapter 3.

The modified gravity model

In a study of spatial accessibility of Primary Health Care (PHC) in Canada, Schuurman et al. (2010) proposed a modified gravity model in which two improvements were made to the basic gravity model. First, they use travel time, instead of travel distance, to represent the travel cost. Second, the distance impedance is captured by a segmented inverse-power function which can be expressed by

$$f(t_{ij}) = \begin{cases} 1, & t_{ij} < 10 \\ 10/t_{ij}, & 10 < t_{ij} < 120 \\ 0, & t_{ij} > 120 \end{cases} \qquad 2.4$$

where t_{ij} represents the shortest traveling time (in minutes) between population location i and medical service site j. The modified gravity model works in a similar way as the 2SFCA method. The only difference is that the former uses a continuous, inverse-power impedance function but the latter uses discretized Gaussian weights. Compared to continuous weights, discretized weights are more suitable for representing the distance impedance in health care studies because people do not mind a few minutes' difference when driving for medical services (Luo and Qi 2009).

The major drawback of gravity-based models, as indicated by Schuurman et al. (2010), is the impedance coefficient (β). Since β reflects people's willingness to access a medical service when considering travel cost alone, it should be estimated from actual physician-visiting data and healthcare utilization surveys using regression methods. However, these data are generally not available. Instead, researchers tend to use arbitrarily-determined impedance coefficients when computing potential spatial access to medical services. This may be problematic because the values of spatial access index may vary substantially when different values of impedance coefficient are used (Luo and Wang 2003). Thus, gravity-based spatial access models may bring significant uncertainties to the analysis results.

To evaluate the uncertainty of spatial access index as stated above, Luo and Wang (2003) tested five consecutive β values on the generic gravity model, but their investigation was limited to standard deviation of the potential spatial access, not the mean value. Luo and Qi (2009) compared analysis results of potential spatial access between two sets of Gaussian weights corresponding to sharp and slow distance-decay. But their study did not specify the β values of the weights. In real-world applications, the

determination of weights is fairly subjective. Therefore, it is still problematic to thoroughly understand the influence of impedance coefficient on the values of potential spatial access calculated by gravity-based models.

2.5 Limitations of Previous Studies of Cancer Disparity

There has been increasing interest in cancer disparity research over recent decades. Researchers have revealed substantial disparities in cancer outcomes by SES, race/ethnicity, and geographic location. Some strategies have been proposed and some implemented to address these disparities by health professionals. Despite progress, some limitations still remain.

2.5.1 The joint impact of multiple factors on cancer disparity

The first challenge of cancer disparity research is to accurately assess the independent and joint influence of SES, race/ethnicity, and geographic location. Cancer disparities result from complicated interactions among many factors, and no single factor can fully explain the disparities. Analysis based on only one or two of these factors would be incomplete. Instead, information about the interactions among these factors would be much more valuable for interpreting the disparity and designing effective interventions.

In order to assess the independent impact of a factor on cancer disparity, one needs to completely control for other factors. However, this has been proven to be difficult for several reasons. First, the lack of relevant data prevents researchers from examining the independent impact of single factors (Brawley 2002). For example, when assessing the separate influence of SES and race on cancer incidence, most studies failed to enroll enough rich African Americans or impoverished whites to prove that SES is independent

of race in influencing cancer incidence. Conclusions based on incomplete datasets are not only weak but also potentially misleading. In order to gain an accurate image of how several factors contribute to cancer disparities, detailed information about both patients and their background population is needed.

Second, researchers did not show enough interest in analyzing the impact of geographic location on cancer disparities. Space is important because it provides a context within which SES and race/ethnicity can be linked to explanatory factors (e.g., environmental exposures and regional behavior risks). In addition, the spatial patterns of cancer and potential spatial access to cancer services are important information for cancer prevention and treatment. The reason for the lack of "geographic thinking" in current studies might lie in the fact that health disparity has long been considered a social problem rather than a geographic problem. Therefore, the methods for analyzing cancer disparities have been greatly influenced by social epidemiology. Being different from health/medical geographers who focus on the place-dependent nature of health problems, social epidemiologists emphasize the societal generalization of health phenomena (Cutchin 2007), leading them to categorize patients and at-risk population by race/ethnicity and SES, regardless of the places they live or work.

2.5.2 Relation between CRC outcome and potential spatial access to CRC services

Frequent screening can increase the probability of early stage CRC diagnosis and good-quality CRC treatment can help extend patients' lives. Therefore, potential spatial access to CRC screening and treatment services may influence CRC outcomes. The connection between CRC survival and potential spatial access to CRC services is valued

by health administrators because it has important implications for allocating CRC-related resources and reducing CRC disparities.

However, despite the growing interest in CRC survival and potential spatial access to CRC services, few studies have analyzed the associations between them. Recent studies have highlighted substantial disparities in CRC diagnosis by race/ethnicity, SES, and geographic location; however, few of them have assessed whether these disparities could be explained by potential spatial access to medical services or not. Other studies have investigated the influence of non-spatial access factors on CRC diagnosis. For example, Ananthakrishnan et al. (2007) analyzed different patterns of CRC screening usage within a Medicare population, but they did not investigate whether the utilization of the service was associated with spatial factors (e.g., distance, location) or not. Henry et al. (2009) examined the impact of health insurance coverage on CRC diagnosis in New Jersey, but did not consider the locations of medical service sites.

In the broader field of cancer epidemiology, Wang et al. (2008) investigated the associations between potential spatial access to breast cancer prevention services and breast cancer stage at diagnosis in Illinois. However, their investigation was conducted at the zip code level, which some have argued may be too coarse of a geographic scale for cancer disparity research (Krieger et al. 2002). In addition, the regression analysis did not consider patient age, a factor that might be closely associated with cancer stage at diagnosis. In another study in rural Kentucky, Huang et al. (2009) found a positive association between distance to mammography facilities and the probability of being diagnosed with late stage breast cancer. But their investigation was limited to screening facilities and did not consider other cancer prevention services such as primary care.

Only two studies have assessed the associations between potential spatial access to medical service and CRC survival. Neither of them was conducted in the United States. For example, Jones et al. (2008) observed no significant relation between CRC survival and travel time to general practitioners or hospitals in Northern England. Dejardin et al. (2008) investigated the association between CRC survival and distance to cancer care in France. They found that CRC survival was influenced by road distance from the nearest referent care centre even when adjusting for stage at diagnosis. However, the results are limited because the access measure used (i.e., the distance to the nearest care center) is too simple and might not be able to capture the important characteristics of either local population or the care centers.

2.5.3 Uncertainties caused by various factors

Uncertainties brought about by area level SES indicators

A variety of SES indicators (e.g., education attainment, income, poverty, occupation, and house ownership) have been used to assess socioeconomic disparities of cancer. Since individual level socioeconomic data are always unavailable, most researchers use area-level SES indicators which are aggregated at different geographic scales (e.g., zip code area, census units). Area-level SES indicators are readily available and can be conveniently linked to demographic indicators (e.g., racial/ethnic composition, age). In addition, area SES indicators can be effectively applied to all people, regardless of their race/ethnicity, sex, and residential area.

However, using area SES to investigate socioeconomic disparities of cancer has two potential drawbacks. First, researchers have not reached a consensus on the most appropriate SES indicator for socioeconomic disparity analysis. Different indicators

represent different dimensions of SES and no single indicator can capture all socioeconomic characteristics related to cancer (Krieger et al. 1999; Krieger et al. 2003). Although composite indicators, synthesized from multiple single indicators, have been adopted to represent the principal aspects of SES (Krieger et al. 2006; Liu et al. 1998; Singh et al. 2002; Yost et al. 2001), there is no evidence that these composite indicators can outperform single ones in revealing socioeconomic disparities of cancer (Krieger et al. 2006; Singh et al. 2002). A second drawback is that using area-based indicators to approximate the SES of individual people may lead to "ecological fallacy", a common source of error in ecological studies. Although assessing the impact of ecological fallacy on health disparity is difficult, it is believed that the impact is associated with the geographic scale at which the study is carried out (Roux et al. 2003). Logically, the impact of ecological fallacy may decrease when smaller units are used because people in smaller units are usually assumed to be more homogeneous in SES than those in larger ones.

Uncertainties brought about by spatial dependence

In addition to area-level socioeconomic indicators, spatial dependence could also impact the result of cancer disparity analysis. Spatial dependence occurs when observations of nearby locations are similar to each other. It can be explained by the first law of geography which claims higher correlations among nearby things than among far things (Tobler 1970). The existence of spatial dependence of the phenomenon being analyzed violates the randomization assumption of most traditional regression methods. Thus, using these traditional regression methods in a spatial context may not reveal the real relation between variables involved in the regression. Although the impact of spatial

dependence has been increasingly recognized, few cancer disparity studies have managed to address the uncertainty problem brought about by it.

Uncertainties caused by the small-number problem

When analyzing cancer disparities, it is necessary to account for the effect of some basic factors (e.g., sex and age). A straightforward way to do this is to apply a direct age-sex adjustment procedure by, for example, adjusting cancer rates according to the standard Census 2000 age-sex groups. The premise of a valid direct-age-sex adjustment is that there is a sufficient population in each subgroup. However, this premise is always difficult to fulfill in practice, especially when small areas (e.g., census tracts, zip code areas) or racial/ethnic minorities (e.g., Asians, African Americans) are involved in the analysis (Pickle and White 1995). Biases and errors will be introduced if these rates are used in subsequent mapping or regression analysis.

One solution to the small-number problem is to use an indirect adjustment procedure. The indirect adjustment works in two steps. The first step is to apply the national or regional level rates of age-sex groups to the local age-sex groups to generate an expected number of cases for an area unit. The second step is to compute an observed-to-expected ratio to represent the unit. The indirect adjustment does not rely on the large-number assumption and has been proved to be more stable in estimating disease rates for small areas than the direct adjustment method (Aylin et al. 1999). Meanwhile, the results of indirect adjustments were very similar to those of direct adjustments, as revealed by a simulation study (Goldman and Brender 2000).

Another solution to the small number problem is to aggregate adjacent units to larger ones to avoid small numbers (Mu and Wang 2008; Wang and O'Brien 2005). These

methods, however, are at the cost of spatial resolution. Degraded spatial resolution may reduce the accuracy of regression analysis (Krieger et al. 2003; Roux 2000).

Other methods have also been designed to address the small-number problem. One example is the empirical Bayes estimator, which has been applied and modified in different studies to estimate age-standardized relative risk of diseases (Clayton and Kaldor 1987; Marshall 1991). This method uses a Poisson distribution to characterize the observed events and assumes the number of observed events within an area unit to be conditional on the relative risk of a larger district and the expected event number of this area unit. Based on this assumption, a maximum likelihood estimator is used to calculate the posterior expectation of the relative risk according to a random-effect model. Another example is the spatial smoothing technique (Kafadar 1996), which smoothes neighboring areal units to calculate stable disease rates. In addition, the Poisson Krieger method, which estimates the disease risk of an area unit as a linear function of neighboring risks, was also introduced in health studies to overcome the small number problem (Ali et al. 2006; Goovaerts 2005; Goovaerts 2006).

A common characteristic of the methods mentioned in the last paragraph is that they use neighborhood information to obtain a smoothed map of health events, thus avoiding unreliable values caused by the small-number problem. However, these methods are only suitable for mapping purposes. It is not appropriate to use these methods to analyze racial/ethnic and socioeconomic disparities of cancer for two reasons. First, cancer rates are generally used as the dependent variable in statistical regressions. In this case, adjusting the dependent variable (i.e., the cancer rate) with the neighborhood data without adjusting the independent variables (i.e., socio-demographic indicators) in the same way

may lead to incorrect results of cancer disparity. Second, neighborhood-based adjustment may increase the extent of spatial dependence, thus bringing more uncertainties to the regression analysis (Kafadar 1996).

CHAPTER 3

DATA AND STUDY DESIGN

3.1 Introduction

This chapter depicts the study area, the data sources, and the study design of this dissertation. The chapter is made up of four sections. Section 3.2 introduces the study area and data sources. It also depicts how the datasets were pre-processed before subsequent analyses. Section 3.3 describes how the permission to perform human-subject research was obtained and how the rights and safety of human subjects were protected. Section 3.4 offers an overview of the research scheme and detailed descriptions of the methods.

3.2 Study Area and Data Sources

3.2.1 Study area

The state of Texas was selected as the study area of this research. Texas lies at the southern corner of the southwestern region of the United States. It has the second largest area and population among all states of the country. The total population of Texas was 20,851,820 in 2000, among which 51.4% (n=10,719,670) were non-Hispanic whites, 11.3% (n=2,364,255) were non-Hispanic blacks, 32.0% (n=6,669,666) were Hispanics, 0.6% (n=118,362) were Native Americans, and 2.7% (n=562,319) were Asians (US Census Bureau 2001a). Rural residents accounted for 13.9% (n=2,907,272) of the total

population. The poverty rate was 15.4% and the median household income was $39,927 in 2000 (US Census Bureau 2001b).

3.2.2 Data sources

Three types of data were used in this research: CRC incidence data, CRC-service data, and socio-demographic data.

CRC incidence data

Individual-level CRC incidence data of Texas from 1995 to 2003 were collected from the Texas Cancer Registry (TCR). TCR is the most comprehensive cancer-data collection system in Texas and has been recording and compiling cancer incidence information of Texas residents since 1979. The case ascertainment rate of TCR incidence data was greater than 98.3% (Risser et al. 2009). The selection codes of CRC are 153.0-154.1 for the 9th revision of the International Classification of Diseases (ICD-9) and C18-C20 for the 10th Revision (ICD-10). The recorded information of each case includes race/ethnicity, stage at diagnosis, sex, date of diagnosis, age at diagnosis, residential address, survival status (by March 1, 2010), age at death, date of death (if deceased before March 1, 2010), cause of death (if deceased before March 1, 2010), date of last update (if still alive after March 1, 2010), and death certificate number (if deceased before March 1, 2010). According to the datasets, there were 77,667 CRC incidence cases in Texas during the period from 1995 to 2003.

CRC stage at diagnosis was measured according to the classification of the Surveillance Epidemiology and End Result (SEER) program, which categorizes cancers into in-situ, localized, regional, and distant stages. This classification provides the basic information about the extent of cancer development according to clinical and pathologic

test results (Young et al. 2000). In this research, the four SEER stages were further combined into two groups: the early-stage group and the late-stage group, with the former including in-situ and localized stages and the latter including regional and distant stages. This pairing is reasonable because, compared to CRC patients diagnosed at regional or distant stages, those diagnosed at in-situ or localized stages are mostly asymptomatic. Therefore, people with early-stage CRC are largely identified through CRC screening rather than through specific symptoms (Read and Kodner 1999; Skibber et al. 2001).

Geocoding was implemented to transfer the address information of CRC cases to two-dimensional points in the GIS environment. The geocoding was implemented in the Geocoding module of ArcGIS 9.3 (Environmental Systems Research Institute (ESRI) 2009) with the census 2000 street file (US Census Bureau 2001c) as the reference data. The geocoding successfully located 69,701 CRC cases. The rest were not geocoded due to either the unmatched residential address or the low geocoding score (the threshold value was set to 60%). The ungeocoded cases would be excluded from subsequent analyses. The ungeocoded cases are evenly distributed across the whole study region, as indicated by their zip code information. As shown in Table 3.1, the ungeocoded cases and the geocoded cases did not differ much in distributions by sex, age, stage at diagnosis, and race/ethnicity.

Table 3.1 Characteristics of geocoded and ungeocoded colorectal cancer cases

	Study Cases	**%**	**Ungeocoded Cases**	**%**
Distribution by Sex				
Male	35,345	50.7	4,102	51.5
Female	34,356	49.3	3,864	48.5
Total	69,701	100	7,966	100
Distribution by Age				
<50	7,391	10.6	804	10.1
50-60	12,176	17.5	1,418	17.8
61-70	16,836	24.2	1,952	24.5
71-80	19,779	28.4	2,223	27.9
>80	13,519	19.4	1,569	19.7

Table 3.1-Continued

Total	69,701	100	7,966	100
Distribution by Race/Ethnicity				
Non-Hispanic White	50,544	72.5	5,788	72.6
Non-Hispanic Black	8,160	11.7	948	11.9
Hispanic	9,592	13.8	1,051	13.2
Asian	857	1.2	80	1.0
Native American	41	0.1	6	0.1
Other	507	0.7	93	1.1
Total	69,701	100	7,966	100
Distribution by Stage at Diagnosis				
In-situ	3,513	5.0	391	4.9
Localized	21,458	30.8	2,324	29.2
Regional	25,044	36.0	2,918	36.6
Distant	10,687	15.3	1,214	15.2
Unknown	9,001	12.9	1,119	14.0
Total	69,701	100	7,966	100

CRC service data

This research focuses on potential spatial access to CRC prevention and treatment services in Texas in 2000. CRC prevention services were decomposed into CRC screening facilities and primary care physicians (PCPs). Oncologists were used to denote the CRC treatment service.

The addresses of CRC screening facilities were collected from the Texas Cancer Prevention and Research Institute (TCPRI), which has been surveying Texas hospitals and clinics about CRC screening services since 2000. The feedback information from survey respondents include the name, the address, the website, the phone number, and the type of available CRC screening facility (e.g., traditional colonoscopy, virtual colonoscopy, flexible sigmoidoscopy, Fecal-occult Blood test, helical CT scanning, and double contrast barium enema) of the hospital (or clinic). By May, 2010, TCPRI had identified 277 hospitals and 13 freestanding cancer centers/community cancer treatment centers that had been equipped with at least one type of CRC screening facility.

However, one shortcoming of the TCPRI data is that it only contains the most recent update time for each service site and does not retain any time information of

previous updates. This limitation makes it difficult for one to determine whether a hospital offered CRC screening service in 2000 even if it did it in 2010. To overcome this shortcoming, a back-tracing procedure was implemented to ascertain the availability of CRC screening facilities of the 290 hospitals and clinics in 2000. The procedure included both telephone and email surveys. The major question of the surveys was: did your hospital or clinic have any facilities of CRC screening in 2000? Sites with the answers of "yes" were kept in the list while sites with the answers of "no" were excluded from the list. Three rounds of survey were implemented with each round focusing on the undetermined sites remained from the last round. According to the back-tracing, among the 290 candidate screening sites, 111 offered CRC-screening services in 2000 and 170 did not. Nine respondents could not give a clear answer due to the loss of previous years' records. According to the address information, the undetermined sites were primarily dispersed in the eastern part of Texas.

PCP is an important type of medical resource because it provides individuals the frontier of contact with the healthcare system. Access to PCPs has been proved critical for disease prevention and medical cost reduction (Lee 1995; Luo 2004). Previous research also suggests that, in some situations, potential spatial access to PCPs is more important for cancer prevention than potential spatial access to screening facilities (Wang et al. 2008). The practicing addresses of Texas PCPs in 2000 were collected from the Center for Health Statistics in Texas Department of State Health Services (DSHS). PCPs were limited to general practice physicians, pediatrics, family physicians, obstetrician-gynecologists, and general internists (Cooper 1994). According to the data, there were 14,268 PCPs working in 6,372 practicing addresses in Texas in 2000.

The oncologist information was collected from the Center for Health Statistics in Texas DSHS, too. Oncologists were limited to those who have postgraduate medical education and have been spending the majority of their time in office or practicing in the hospital, in research, or in teaching. According to the data, there were 205 oncologists located in 121 practicing addresses in Texas in 2000.

To determine the geographic locations of CRC screening sites, PCP sites, and oncologist sites, geocoding was implemented in ArcGIS 9.3 (ESRI 2009) with the US 2000 street map used as the reference layer. About 99% (230 out of 232) of the oncologist sites and the back-traced CRC screening facility sites were successfully geocoded. About 97% (13,816 out of 14,268) of PCPs were successfully geocoded. The high geocoding rates of these sites might be due to the sound address information provided by health professionals. The ungeocoded PCPs were primarily located in eastern Texas, as indicated by the zip code information.

Socio-demographic data

The socio-demographic data include census tract level poverty rate and age-sex structure information for racial/ethnic groups. The poverty rate data was derived from the Summary File 3 (SF3) of census 2000 datasets (US Census Bureau 2001b). The age-sex structure information for racial/ethnic groups was obtained from the Summary File 1 (SF1) of census 2000 datasets (US Census Bureau 2001a).

3.3 Protection of Human Subjects

When human subjects are involved in research, it is necessary to protect their safety and rights. Federal regulation and rules have been developed to guarantee the rights for both human subjects and researchers (International Science and Engineering

Fair 2006). The human subjects involved in this research were CRC patients. Although the research did not include any direct or indirect interactions with CRC patients, some steps were still necessary to protect some confidential information such as the residential address and some other identifiable variables of the cases.

The use and analysis of the CRC incidence data in this research have been approved by Texas DSHS Institutional Review Board (IRB) after a reviewing process. The IRB review involved an agreement between TCR and the data users to ensure the confidentiality of the human subjects. According to the agreement, the following provisions were required during the processing and analyses of the CRC incidence data.

a). The cancer registry data are treated as strictly confidential.

b). During the study, a password-protected computer with up-to-date antivirus software is used to store and analyze the confidential data. A cabinet with access limited only to the data users is used to lock up the computer when not in use.

c). The presentation and publication of results do not include specific information of individual cases or make any case identifiable.

d). The confidential dataset will be destroyed one year after the research is finished. A non-confidential dataset will be created and maintained.

3.4 Study Design

3.4.1 Methodology overview

The general scheme of this dissertation is illustrated in Figure 3.1. Three parts constitute the whole research. The first part investigates potential spatial access to CRC services in Texas and examines if the access varied by race/ethnicity, SES, and geographic location. A relative spatial access approach is proposed in this part to

overcome the limitation of previous models. The other two parts analyze the disparities of CRC stage at diagnosis and CRC-specific survival in Texas by following a similar procedure. Specifically, the second part examines if there were any disparities of CRC stage at diagnosis from the aspects of SES, race/ethnicity, geographic location, and potential spatial access to CRC prevention services in Texas. Bivariate analyses and generalized estimating equation logistic regressions are employed to analyze the individual and joint impact of these factors on CRC stage at diagnosis. The overall spatial pattern and the spatial clusters of CRC stage at diagnosis are examined by an adaptive spatial filtering method. The third part of this research investigates if there were any disparities in CRC-specific survival by race/ethnicity, SES, geographic location, and potential spatial access to CRC treatment service in Texas. Kaplan-Meier estimators are used to examine characteristics of CRC-specific survival by single factors. Cox proportional hazard regression models are employed to examine how different factors jointly influence CRC-specific survival. Spatial clusters of CRC-specific survival are identified by the spatial scan statistic (exponential model). A further comparison of racial/ethnic and socioeconomic characteristics is made between high risk and low risk clusters.

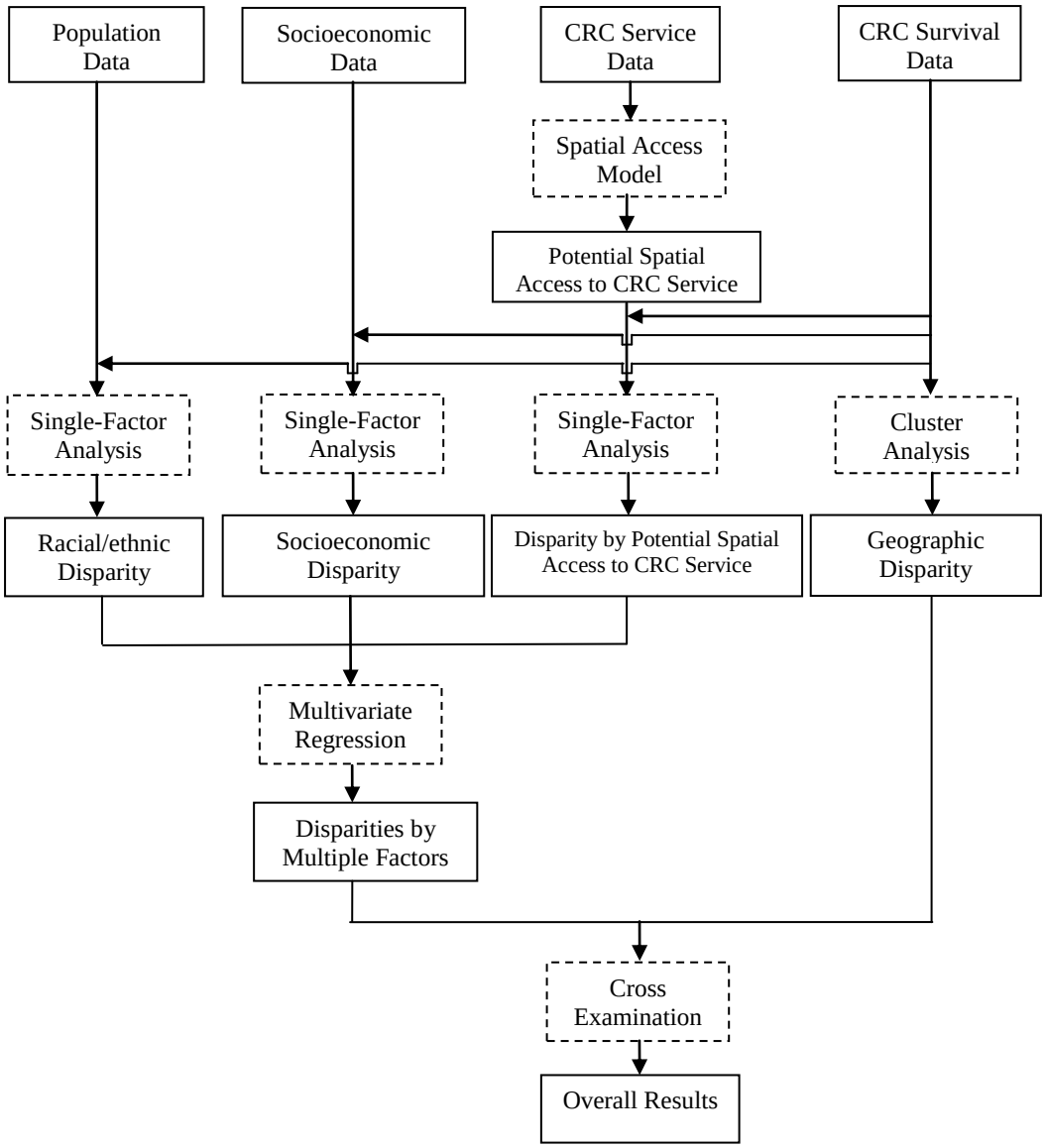

Figure 3.1 The general scheme of the research

3.4.2 A review of methods used in the research

This section introduces the methods for analyzing potential spatial access to CRC services, disparities of CRC stage at diagnosis, and disparities of CRC-specific survival in this research. Specifically, it describes the E2SFCA method, the generalized equation estimation logistic regression, the Kaplan-Meier estimator, the Cox proportional hazard

regression model, the adaptive spatial filtering method, and the exponential spatial scan statistic method.

The enhanced 2-step floating catchment area (E2SFCA) method

The E2SFCA method is an improved version of the 2SFCA method for characterizing potential spatial access to medical services. Like 2SFCA, E2SFCA also works in two steps. The first step is to generate a 30-minutes catchment for each medical service site *j*, dividing the catchment into three sub-zones at the intervals of 10 and 20 minutes, and calculating the supply-to-demand ratio, R_j, by

$$R_j = \frac{S_j}{\sum_{j\in(d_{kj}\in D_r)} P_k W_r} = \frac{S_j}{\sum_{j\in(d_{kj}\in D_1)} P_k W_1 + \sum_{j\in(d_{kj}\in D_2)} P_k W_2 + \sum_{j\in(d_{kj}\in D_3)} P_k W_3} \qquad 3.1$$

where S_j represents the medical capacity of medical site *j*, P_k denotes the population of area unit *k* inside the catchment, d_{kj} is the traveling cost between *j* and *k*, D_r is the r^{th} sub-zone of the catchment, and W_r is a predefined Gaussian-weight for D_r. The second step is to calculate the spatial access index of location *i* as the sum of weighted supply-to-demand ratios of all medical sites within the catchment of *i*:

$$A_i^F = \sum_{l\in(d_{il}\in D_r)} R_l W_r = \sum_{l\in(d_{il}\in D_1)} R_l W_1 + \sum_{j\in(d_{il}\in D_2)} R_l W_2 + \sum_{l\in(d_{il}\in D_3)} R_l W_3 \qquad 3.2$$

where A_i^F is the spatial access index for area unit *i*, R_l is the supply-to-demand ratio of medical site *l* that falls inside the catchment of *i*, and d_{il} is the traveling cost between *l* and *i*.

E2SFCA assumes that a population location's potential spatial access to medical services decreases with the increase of traveling time within the catchment. This assumption is more reasonable than that of the basic 2SFCA model (Luo and Qi 2009).

Generalized estimating equation (GEE) logistic regression

Generalized estimation equation (GEE) is an improvement of the basic regression models for data with unknown correlations (Zeger et al. 1988; Zorn 2001). It was designed to facilitate the analysis of multi-level or temporal observations. Supposing Y_{it} is the dependent variable of the t^{th} case (or repeated measurement) within the i^{th} strata (or group) and X_{it} is the corresponding covariate, GEE expresses the relation between Y and X as

$$E(Y_i) = \mu_i = h(X_i\beta) \qquad 3.3$$

where h represents the reverse linking function and β denotes the parameter vector. The variance (V_i) of Y_i is specified as

$$V_i = \frac{(A_i)^{\frac{1}{2}} R_i (A_i)^{\frac{1}{2}}}{\phi} \qquad 3.4$$

where A_i is a diagonal matrix determined by the reverse link function, R_i is a working correlation matrix for Y_i, and ϕ is a scale parameter. Then, one can estimate the value of β by solving the "quasi-likelihood" equations of V_i:

$$U_k(\beta) = \sum_{i=1}^{N} D_i' V_i^{-1} (Y_i - \mu_i) = 0 \qquad 3.5$$

where $D_i = \mu_i/\beta$. One can either use a generalized weighted least-square approach or an iterative approach to solve the functions (Zeger et al. 1988). GEE is applicable for both linear and logistic regression models. A GEE logistic regression can be implemented by adopting the logit function as the reverse linking function.

An attractive attribute of GEE regression is that it does not require any explicit definition of the origin of correlations. Using a GEE model, investigators have great flexibility in specifying the working correlation matrix, which could come in a variety of forms (Zorn 2001).

The Kaplan-Meier estimator

An important characteristic of survival analysis is that it involves case censoring. Censored observation occurs when (1) the event of interest (e.g., cancer-specific death) has not occurred during the study period, (2) the case withdrew from the follow-up, or (3) the event occurred due to other causes. When calculating cause-specific survival rates, it is critical to involve both censored and uncensored cases in the computation.

The Kaplan-Meier estimator was designed for analyzing survival data that involve censored cases (Kaplan and Meier 1958). Let T represent the survival time and $S(t)$ represent the probability that a case from a background population will have a lifetime (or survival time) longer than *t* ($S(t) = \Pr(T > t)$). For a random sample of the population, let the observed time be $0 \leq t_1 \leq t_2 \leq t_3 \leq \cdots \leq t_N$, $S(t)$ can be estimated by a maximum likelihood approach

$$\hat{S}(t) = \prod_{t_i \leq t} \frac{n_i - d_i}{n_i} \qquad 3.6$$

where n_i and d_i represent the number of alive samples prior to t_i and the number of deaths at time t_i, respectively. When there are no censored cases, n_i represents the number of survivors prior to t_i. When censoring occurs, n_i equals the number of survivors minus censored cases. The statistical variance of the estimation could be assessed by estimators such as Greenwood's formula (Kaplan and Meier 1958).

Cox proportional hazard regression

The Kaplan-Meier estimator could only analyze the survival based on single factors. In order to trace the survival from multiple factors, advanced regression methods are needed. Cox proportional hazard regression, first proposed by Cox (1972), is the most widely used model for multivariate survival analysis.

Let $h_i(t)$ be the hazard function showing the instantaneous risk of demise for the i^{th} observation at time *t*, the Cox proportional hazard regression models $h_i(t)$ as a function of covariates by

$$h_i(t) = h_0(t)e^{X_i\beta} \quad 3.7$$

where $h_0(t)$ represents the baseline hazard at *t*, X_i represents the covariate vector corresponding to the i^{th} individual, and β is the parameter vector. For two observations *k* and *k'* that differ in their covariate values, the hazard ratio can be calculated by

$$\frac{h_k(t)}{h_{k'}(t)} = \frac{h_0(t)e^{X_k\beta}}{h_0(t)e^{X_{k'}\beta}} = \frac{e^{X_k\beta}}{e^{X_{k'}\beta}}. \quad 3.8$$

The parameter vector β can be estimated by a partial likelihood estimation approach introduced by Cox (1972).

The major advantage of Cox proportional hazard regression lies in that it does not make any assumptions of the hazard function and the baseline hazard. Using Cox proportional hazard regression model, investigators could focus more on the data instead of the specific form of hazard functions.

The adaptive spatial filtering method

The adaptive spatial filtering method represents an improvement of the basic spatial filtering method for analyzing the geographic patterns of health phenomena (Rushton and Lolonis 1996; Talbot et al. 2000). This method can be implemented in three steps. In the first step, a fine resolution (e.g., 10 miles) regular grid lattice is generated across the study area. The second step is to draw an adaptive distance circle, or spatial filter, around each grid point to achieve a constant number of observations within the circle. This constant number is determined by testing the relation between significance level and sample size using standard statistical testing procedures (Cai 2007; Talbot et al.

2000). The third step is to randomly generate simulated cases from the background population. The simulation is run many times (for example, 999 times). For each simulation, the second step is repeated while replacing the observed cases with simulated cases. A reference distribution of disease rates can be obtained through this simulation. Based on the simulation result, an empirical *p*-value can be computed for the observed disease rates.

The exponential model of spatial scan statistic

The exponential model of spatial scan statistic was proposed to facilitate the spatial analysis of survival data (Huang et al. 2007). This method can determine if there are any significant geographic variations of survival times. Based on the assumption that survival times follow an exponential model, this method compares the mean survival times within a geographic area (θ_{in}) with those outside the area (θ_{out}). Therefore, the null hypothesis is H0: $\theta_{in} = \theta_{out}$ and the alternative is either Ha: $\theta_{in} < \theta_{out}$ or Hb: $\theta_{in} > \theta_{out}$, depending on whether the researcher wants to detect low or high risk areas. A circular kernel whose size ranges from two cases to a half of the total cases is used to scan the entire study area. The Monte Carlo permutation method is employed to test the statistical significance of "hot" or "cold" spots and to adjust for multiple testing (Huang et al. 2007).

Sometimes it is necessary to adjust the survival time for some important covariates to see how spatial clusters are influenced by these covariates. One can implement the covariate adjustment by following the suggestions of Huang et al. (2007). Briefly, the lifetime is modeled as an exponential linear function of covariates by

$$\ln(T_i) = x_i'\beta + W_i \qquad 3.9$$

where T_i is the lifetime of case *i*, x_i is the indicator vector associated with *i*, β represents the coefficient vector, and W_i is the error term which follows a density distribution function $f_W(\omega) = e^{-\omega}e^{-e^{\omega}}, \omega \in (-\infty, +\infty)$. T_i follows a density function of $f(t|x) = \exp(-x'\beta)\exp(-t \times \exp(-x'\beta))$ which represents an exponential distribution. Based on this model, one can estimate the parameter vector, $\hat{\beta}$, by applying all the observed cases in the regression. Then, the estimated survival time for the i^{th} case can be calculated as $x_i'\hat{\beta}$. Finally, given a reference indicator vector *k* (for example, $k_t = \min(x_{it})$), the adjusted survival time for case *i* is estimated as

$$t_i^{adj} = t_i \times \exp\{-\sum_{j=2}^{p} \hat{\beta}_j \times (x_{ij} - k_j)\} \qquad 3.10$$

where t_i^{adj} is the adjusted survival time, t_i is the original survival time, and *p* is the total number of regression coefficients (the size of $\hat{\beta}$).

CHAPTER 4

A RELATIVE ASSESSMENT APPROACH FOR ANALYZING POTENTIAL SPATIAL ACCESS TO COLORECTAL CANCER SERVICES IN TEXAS

4.1 Introduction

Access to medical services is critical for CRC prevention and treatment. Poor access to CRC screening services may lead to low frequency of screening among at-risk populations and increase the risk of late stage diagnosis. Less access to CRC treatment services may decrease the quality of treatment and lower the chance of survival. Ensuring adequate and equitable access to CRC-related services for all population groups has become a prerequisite for CRC disparity elimination.

As noted in Chapter 2, access to medical services is influenced by both spatial and non-spatial factors. Potential spatial access to medical service is critical for improving health and eliminating health disparities because it provides the basis for medical service utilization (Higgs 2004). Previous US studies have identified substantial differences of potential spatial accesses to a variety of medical services by race/ethnicity, SES, and rurality (Onega et al. 2008; Wang et al. 2008). However, few studies have thoroughly examined potential spatial accesses to CRC services.

A methodological limitation of gravity-based spatial access models is the uncertainty problem brought about by the arbitrarily-determined impedance coefficient. However, as revealed in Chapter 2, few studies have assessed how the results of gravity-based models vary with the impedance coefficient and how to better present the spatial access results.

This chapter first implements an uncertainty assessment of the E2SFCA method, the most recent development of gravity models, and then proposes a relative spatial access approach to minimize the influence of the uncertainty problem. The proposed approach is then applied to investigate potential spatial accesses to CRC prevention and treatment services in Texas and to examine their socio-demographic and geographic characteristics. The remainder of this chapter is composed of four parts. A detailed description of the data sources and methodology is provided in the second section. The third section presents the results of the uncertainty assessment and potential spatial accesses to CRC services. The fourth section concludes the chapter with a summary and discussion of the results.

4.2 Data and Methodology

4.2.1 Data

In this study, primary care physicians (PCPs) and CRC screening facilities were used to represent CRC prevention services. Oncologists were used to represent CRC treatment service. The datasets of PCPs, CRC screening facilities, and oncologists of Texas in 2000 have been collected from different sources and pre-processed by procedures described in Chapter 3. The geocoding successfully located 13,816 PCPs

(6,369 practicing sites), 110 CRC screening facility sites, and 204 oncologists (120 practicing sites).

Socio-demographic factors include census tract level population size, racial/ethnic distribution, poverty rate, and rural/urban status. Population data and racial/ethnic distribution data were collected from the Summary File 1 of census 2000 datasets (US Census Bureau 2001a). Census tract poverty rate data were compiled from the Summary File 3 of census 2000 datasets (US Census Bureau 2001b). The rural/urban information of census tracts was derived from the Rural Urban Commuting Area (RUCA) codes (Hart 2006), which integrates the census-defined urban/rural classifications and work commuting data to characterize the urban and rural status of census tracts. The RUCA scheme classifies census tracts into four groups: metropolitan (urban core), micropolitan (large rural town), small rural, and isolated rural. This classification scheme has been widely used in healthcare research due to its ability to better characterize rural areas at the sub-county level (Washington State Department of Health 2009).

4.2.2 Methodology

4.2.2.1 Methodology overview

Two different methods were used to compute potential spatial accesses to different CRC services, with the shortest travel time for CRC screening facilities and the E2SFCA method for PCPs and oncologists. This differentiation is reasonable because of the different characteristics of screening facilities and the other two medical resources. Since a cancer screening only takes a short time, the supply-to-demand ratio may not influence the accessibility of screening facilities much. In this case, the shortest travel time represents the potential spatial access better (Dejardin et al. 2006; Wang et al. 2008).

However, since PCPs and oncologists are human resources of the medical system, their accessibilities can be easily influenced by the supply of medical professionals and the magnitude of potential demand. For example, a long waiting-list of a physician may significantly decrease people's access to that physician. Therefore, potential spatial accesses to PCPs and oncologists are better characterized by gravity-based methods.

4.2.2.2 Sensitivity assessment of the E2SFCA method

The aim of the sensitivity assessment in this chapter is to examine how the results from E2SFCA would be sensitive to varying degrees of distance impedances. In addition to the spatial access index, the spatial access ratio was introduced to represent the relative spatial access as an extension of the E2SFCA method. The spatial access ratio was calculated as the ratio between a census tract's spatial access index and the mean spatial access index of all census tracts. Both spatial access index and spatial access ratio were computed and a comparison analysis was conducted to evaluate their performances.

The assessment was implemented by measuring potential spatial access to PCPs in the Austin-San Antonio corridor area instead of the whole state of Texas. The corridor area (Figure 4.1) contains nine counties, including Williamson, Travis, Bastrop, Hays, Caldwell, Comal, Guadalupe, Bexar, and Wilson counties. The area is composed of both metropolitan areas (e.g., Austin and San Antonio metropolitan areas) with highly concentrated PCP sites and rural areas (e.g., Wilson and Caldwell counties) where PCP sites are sparse. The total population of the corridor area in 2000 was 2,842,146. To account for the "edge effect," a buffer zone of 60-minute travel time was extended from the borders of the corridor area. Both the buffer zone and the corridor area were

incorporated in the computation but only the results of the corridor area were used in the discussions.

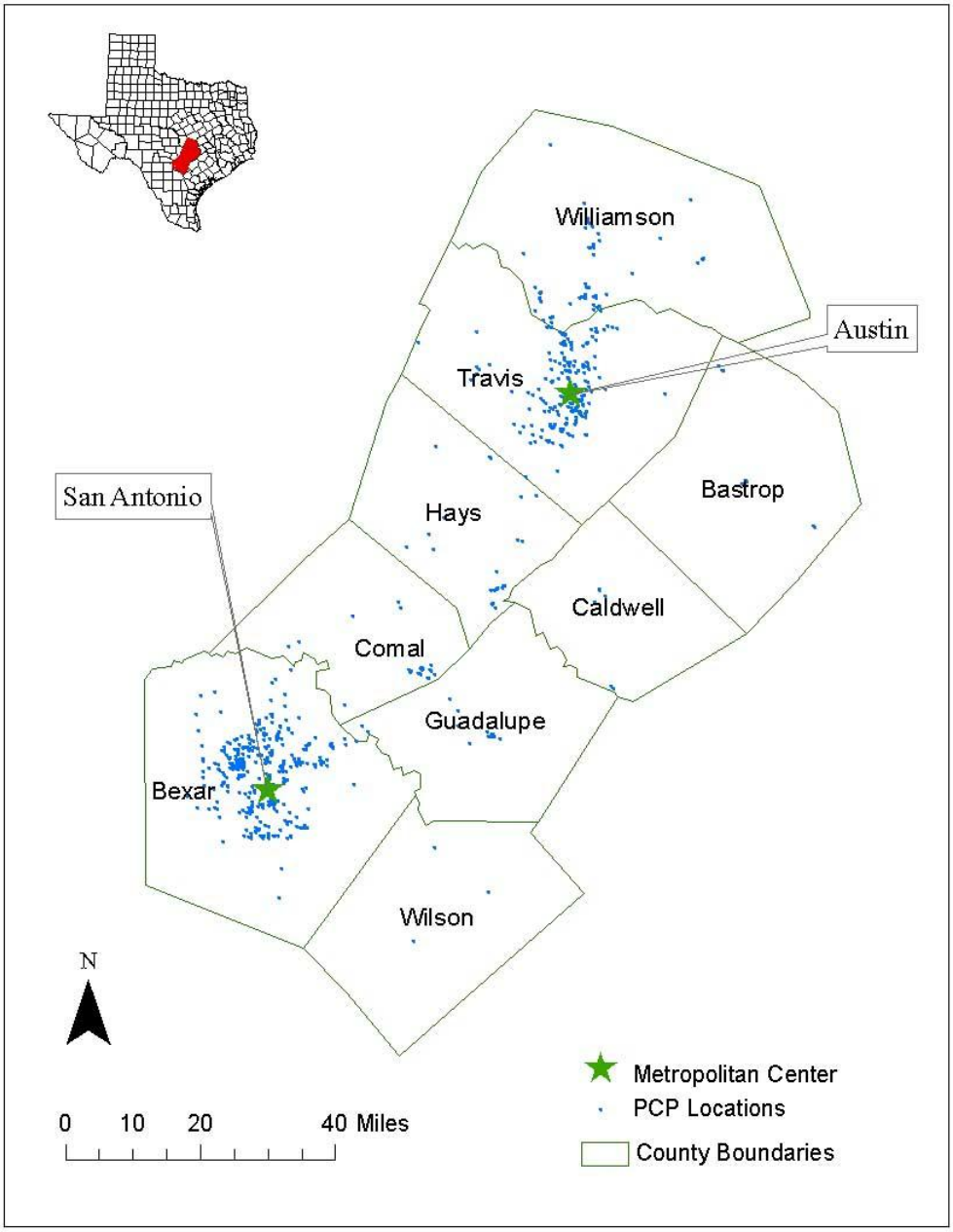

Figure 4.1 Study area of the sensitivity assessment

The assessment was implemented at the level of census tract. To account for the within-tract population variation, population-weighted centroids were calculated to represent the population locations of census tracts (Hwang and Rollow 2000). This

population-weighted centroid was derived based on the spatial distributions of census block population by

$$x_t = \sum_{i=1}^{n_t} p_i x_i / \sum_{i=1}^{n_t} p_i \quad 4.1$$

$$y_t = \sum_{i=1}^{n_t} p_i y_i / \sum_{i=1}^{n_t} p_i \quad 4.2$$

where x_t and y_t represent the weighted-centroid coordinates for census tract *t*, x_i and y_i represent the coordinates of the geometric centroid of the i^{th} census block within census tract *t*, p_i represents the population of the i^{th} census block, and n_t represents the total number of blocks within census tract *t*. The population data at the block level were collected from the Summary File 1 of the census 2000 datasets (US Census Bureau 2001a).

Shortest travel times between census tracts and medical sites were computed using the Network Analyst Extension of ArcGIS 9.3 (ESRI 2009). Briefly, given a set of origin points (i.e., census tract centroids) and a set of destination points (i.e., PCP sites), this function can create a matrix showing the travel time between all origin-destination pairs within a travel-time threshold. The computation considers the speed limit of roads as well as some general driving conventions. Data of the road network and the speed limit were derived from the US 2000 street map.

In the sensitivity assessment, the catchment size of the E2SFCA method was extended to 60 minutes because, according to a preliminary assessment, using this size could incorporate many isolated, rural census tracts and avoid "island" areas (McGrail and Humphreys 2009) in the computation. The extended area (i.e., 30 – 60 minutes) composed the fourth subzone of the catchment. The Gaussian function ($f(d) = e^{-d^2/\beta}$) was adopted as the distance impedance function because it has been proved superior to

other functions in simulating the distance impedance effect (Wang 2007). The mean travel time for each sub-zone (5, 15, 25, and 45 minutes for the four sub-zones, respectively) was used as d_{ij} in the computation of Gaussian weights. Seven sets of Gaussian weights with different impedance coefficients (Table 4.1) were used in the sensitivity assessment. As suggested by Kwan (1998), 0.01 is a critical point for Gaussian function approaching zero. For the travel cost (i.e., 45 minutes) of the outmost sub-zone (D_4), the β value of 440 corresponds to the Gaussian value of 0.01. Therefore, the minimum value of β was set to 440. The maximum of β was set to 1,040 because the Gaussian curve is relatively flat at this value. The other five sets of β range from 540 to 940 with an increment of 100. The increase of β represents the decrease of the extent of distance impedance effect.

Table 4.1 Gaussian weights with different distance impedances for catchment sub-zones

Distance Impedance Coefficient (β)	Sub-zone 1	Sub-zone 2	Sub-zone 3	Sub-zone 4
440	0.945	0.600	0.242	0.010
540	0.955	0.659	0.314	0.024
640	0.962	0.704	0.377	0.042
740	0.967	0.738	0.430	0.065
840	0.971	0.765	0.475	0.090
940	0.974	0.787	0.514	0.116
1040	0.976	0.805	0.548	0.143

4.2.2.3 Analyzing potential spatial access to CRC services in Texas

Potential spatial accesses to PCPs and oncologists for Texas census tracts were computed by the E2SFCA method. The population-weighted centroid was calculated for each census tract by formulas 4.1 and 4.2. For potential spatial access to PCPs, the catchment size was set to 60 minutes with the breaks at 10, 20, and 30 minutes for the

four sub-zones. For potential spatial access to oncologists, the catchment size was set to 180 minutes with the breaks of 30, 60, and 120 minutes. Larger catchment size was used for measuring spatial access to oncologists because, given the life threats of cancer, generally people would not mind traveling a long distance for cancer treatment services. The selection of the impedance coefficient (β) and the presenting form (spatial access index or spatial access ratio) of spatial access would be based on the sensitivity analysis results.

After the computation of potential spatial access to CRC services for census tracts, the geographic patterns of the access were examined. Also examined were the racial/ethnic, socioeconomic, and rural-urban characteristics of the spatial accesses. Specifically, the median potential spatial access with inter-quartile range (IRQ) was calculated by public health region, race/ethnicity, poverty rate, and rural-urban status. Since the analyses were implemented for the population of the whole state rather than a sample, statistical significances of the results were not tested. The whole analysis was realized using the Visual Basic for Applications (VBA) in ArcGIS 9.3 (ESRI 2009).

4.3 Results

4.3.1 Sensitivity assessment results

Table 4.2 presents the mean values and the normalized standard deviations of spatial access indices with varying impedance coefficients. For comparison, the standard deviations were normalized so that the mean value of spatial access indices for each β equals one.

Table 4.2 Mean and standard deviation of spatial access indices with different impedance parameters

Distance Impedance Coefficient (β)	Mean Spatial Access Index ($\times 10^{-3}$)	Normalized Standard Deviation of Spatial Access Index ($\times 10^{-3}$)
440	0. 761	0.400
540	1.523	0.391
640	2.286	0.384
740	3.050	0.376
840	3.816	0.369
940	4.581	0.362
1040	5.347	0.354

As shown in Table 4.2, larger impedance coefficients lead to higher spatial access indices and smaller normalized standard deviations. Since a larger β value (or weaker distance impedance) indicates a less extent of distance impedance, the increased spatial access index is reasonable. The decrease of normalized standard deviation indicates that larger βs have a smoothing effect (Fotheringham et al. 2000) and yield more homogeneous surfaces of spatial access.

To further assess the robustness of the two forms of spatial access presentation, two-sample *t* tests were performed between the first distance impedance coefficient ($\beta = 440$) and other impedance coefficients for both spatial access index and spatial access ratio. The null hypotheses were stated as H0A: $A_{\text{Spatial Access Index, } \beta 1} = A_{\text{Spatial Access Index, } \beta i}$, and H0B: $A_{\text{Spatial Access Ratio, } \beta 1} = A_{\text{Spatial Access Ratio, } \beta i}$ where βi represents the *i*-th ($i \in [2,7]$) impedance coefficient. As expected, the tests rejected H0A and preserved H0B (Table 4.3).

Table 4.3 Two-sample *t* test results of spatial access index and spatial access ratio

β pairs	Spatial Access Index		Spatial Access Ratio	
	t-score	*p*-value	t-score	*p*-value
1-2	-62.198	1.000	0.000	0.000
1-3	-63.432	1.000	0.000	0.000
1-4	-64.727	1.000	0.000	0.000
1-5	-66.073	1.000	0.000	0.000
1-6	-67.453	1.000	0.000	0.000
1-7	-69.250	1.000	0.000	0.000

The geographic patterns of spatial access indices and spatial access ratios computed by the E2SFCA method are shown in Figures 4.2 and 4.3, respectively. For spatial access index, census tracts were categorized into six groups with 0.00029, 0.00049, 0.00069, 0.00089, and 0.001 as the boundary values of intervals. For spatial access ratio, census tracts were categorized into six groups with boundary values (0.38, 0.64, 0.91, 1.17, and 1.31) that correspond to the spatial access index intervals of the first impedance coefficient ($\beta = 440$).

As shown in Figures 4.2 and 4.3, the geographic pattern of spatial access index varies greatly among the seven groups of Gaussian weights, with larger impedance coefficients yielding more homogeneous geographic distributions. On the other hand, the geographic distribution of spatial access ratio does not change much and remains relatively stable to the change of β. These results indicate that the relative measure (i.e., spatial access ratio) is less sensitive to the distance impedance coefficient than the absolute measure (i.e., spatial access index) in characterizing potential spatial access to medical services.

4.3.2 Geographic patterns of potential spatial access to CRC services in Texas

Based on the results of the sensitivity assessment, spatial access ratio was adopted to present potential spatial accesses to PCPs and oncologists. Since spatial access ratio is stable to the variation of distance impedance, the impedance coefficient was set to 440 for PCPs and 4,890 (i.e., the value which corresponds to the Gaussian value of 0.01 for the outmost subzone) for oncologists.

Figure 4.4 shows the geographic patterns of potential spatial access to the three CRC services in Texas. Generally, urban areas where PCPs and CRC screening facilities were highly clustered had the highest potential spatial access to CRC prevention services. Potential spatial access to oncologists was very low (< 0.40) across the whole state except for a limited number of metropolitan areas, among which the Houston metropolitan area had the best access. To further examine the regional differences of the potential spatial access, median spatial access ratios and interquartile ranges (IQRs) of the 11 public health regions (PHRs) (also shown in Figure 4.4) of Texas were compared (Table 4.4).

As shown in Table 4.4, residents of the westernmost tip (PHR 10) and the southernmost tip (PHR 11) of Texas have the lowest potential spatial access to PCPs. These two areas are also among the PHRs that have the least potential spatial access to oncologists. Residents of some northern parts (PHRs 2 and 4) and the easternmost part (PHR 5) have the longest travel time to CRC screening facilities. PHRs 2, 9, 10 and 11 are the most disadvantaged regions in potential spatial access to oncologists. PHRs 1, 3, 6, 7, and 8 have the best overall access to all three CRC services among all the 11 regions.

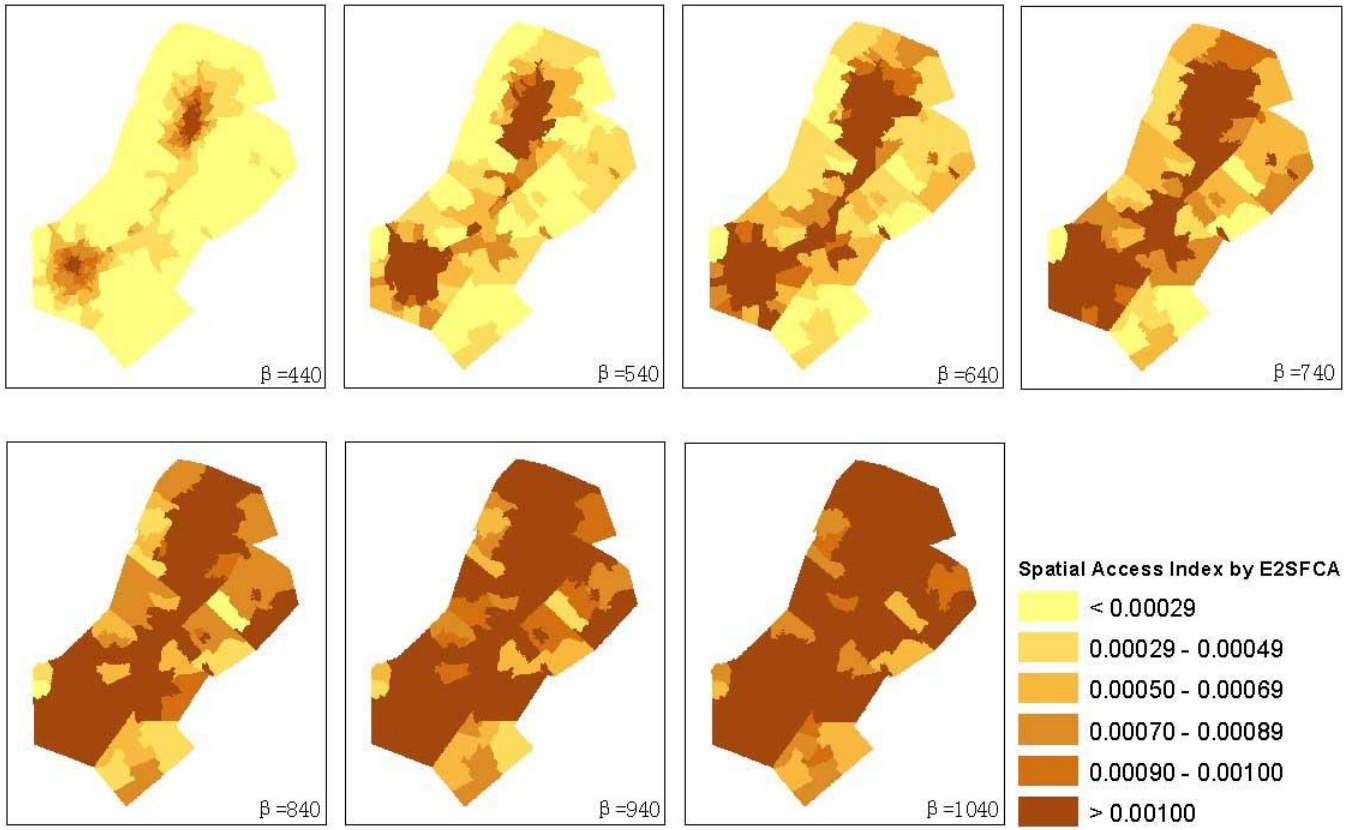

Figure 4.2 Geographic patterns of spatial access indices determined by the E2SFCA method (Note: β is the distance impedance coefficient of the impedance function.)

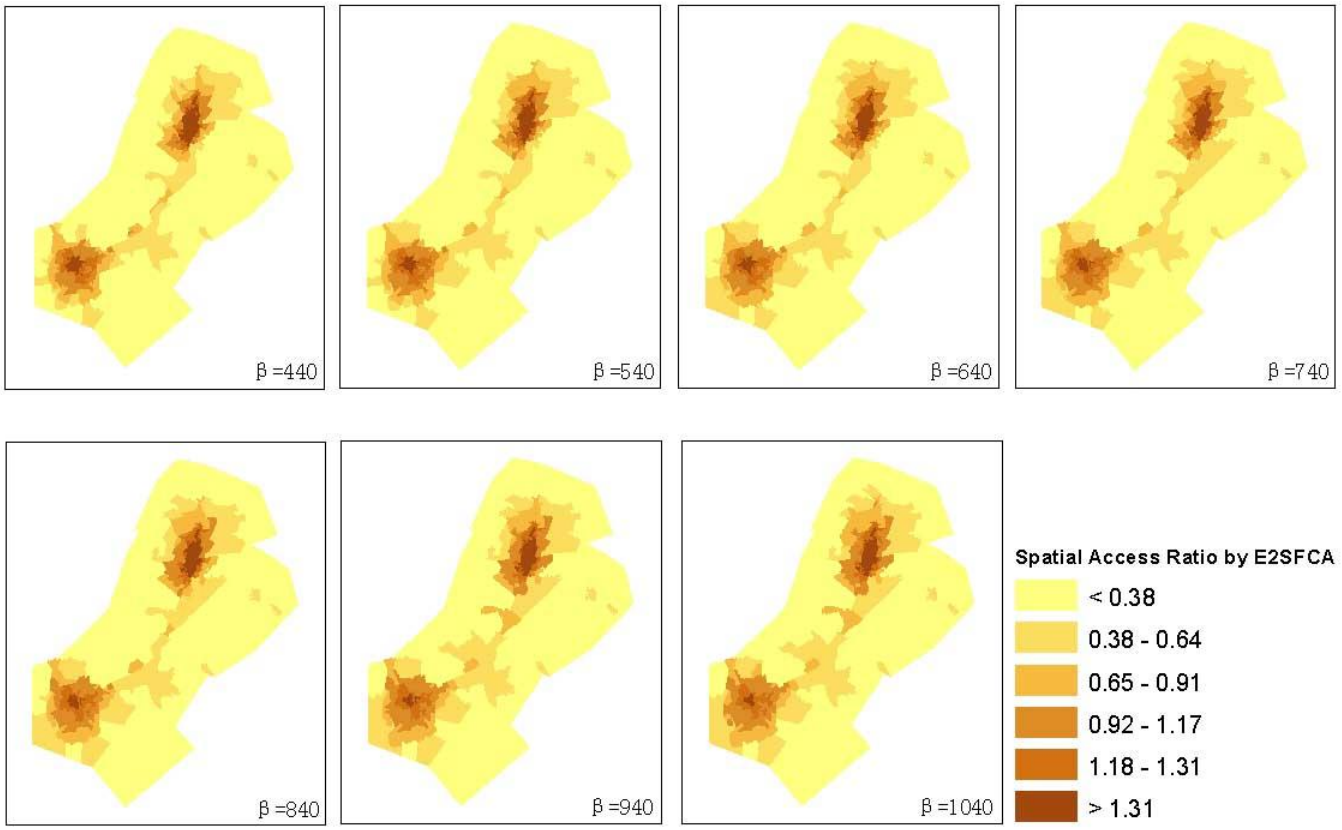

Figure 4.3 Geographic patterns of spatial access ratios determined by the E2SFCA method (Note: β is the distance impedance coefficient of the impedance function.)

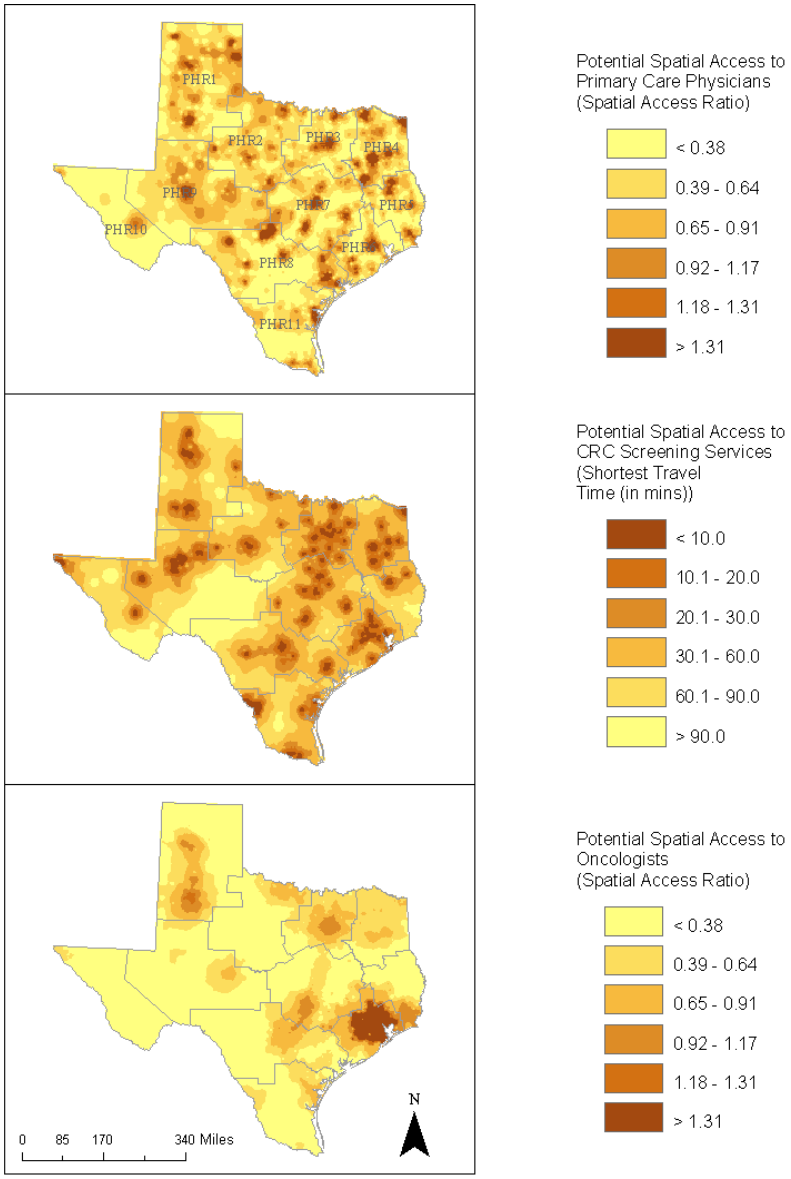

Figure 4.4 Geographic patterns of potential spatial access to colorectal cancer services in Texas (2000) (Note: PHR refers to public health regions.)

Table 4.4 Median potential spatial access (and the Inter-Quartile Range in parenthesis) to colorectal cancer services in Texas by race/ethnicity, socioeconomic status, rurality, and public health region (Note: a larger spatial access ratio represents higher spatial access whereas longer travel time means lower spatial access.)

	Potential spatial Access to Primary Care Physicians (Spatial Access Ratio)	Potential Spatial Access to CRC Screening Facilities (Minutes)	Potential Spatial Access to Oncologists (Spatial Access Ratio)
By race/ethnicity			
Non-Hispanic white	0.95 (0.51, 1.29)	14.2 (7.9, 31.4)	0.86 (0.55, 1.04)
Non-Hispanic black	1.19 (0.89, 1.42)	11.0 (7.1, 17.9)	0.98 (0.77, 2.49)
Hispanic	1.07 (0.72, 1.38)	10.0 (6.1, 20.5)	0.84 (0.42, 1.03)
Asian	1.17 (0.94, 1.41)	8.8 (5.9, 13.3)	1.03 (0.93, 2.49)
Native American	1.01 (0.60, 1.35)	11.9 (6.9, 26.9)	0.85 (0.48, 1.02)
By SES			
Q1 (High SES)	1.20 (0.78, 1.47)	9.3 (5.4, 23.3)	0.97 (0.81, 1.70)
Q2	1.05 (0.55, 1.38)	12.1 (6.8, 36.1)	0.82 (0.52, 1.04)
Q3	0.95 (0.51, 1.29)	14.4 (8.2, 32.2)	0.79 (0.40, 1.04)
Q4 (Low SES)	0.97 (0.65, 1.25)	12.1 (7.8, 18.2)	0.77 (0.37, 1.03)
By Rurality			
Metropolitan	1.11 (0.75, 1.38)	10.5 (6.5, 17.3)	1.11 (0.69, 1.48)
Micropolitan	0.72 (0.45, 1.01)	35.9 (21.1, 58.7)	0.42 (0.20, 0.69)
Small Rural	0.54 (0.32, 0.85)	50.4 (34.7, 65.9)	0.35 (0.18, 0.57)
Isolated Rural	0.29 (0.38, 0.54)	56.9 (37.8, 73.5)	0.30 (0.13, 0.51)
By Public Health Region (PHR)			
PHR1	1.17 (0.65, 1.47)	10.2 (6.0, 50.0)	0.98 (0.40, 1.16)
PHR2	0.96 (0.59, 1.30)	34.1 (19.1, 53.9)	0.19 (0.11, 0.78)
PHR3	1.09 (0.71, 1.34)	11.4 (7.8, 17.1)	0.95 (0.77, 0.97)
PHR4	0.90 (0.45, 1.46)	24.4 (9.4, 37.8)	0.55 (0.42, 0.66)
PHR5	0.92 (0.42, 1.32)	65.7 (50.1, 79.1)	0.74 (0.19, 1.05)
PHR6	1.07 (0.66, 1.43)	9.4 (6.3, 14.0)	2.53 (1.85, 2.59)
PHR7	1.01 (0.50, 1.41)	14.8 (7.1, 29.3)	0.74 (0.40, 1.03)
PHR8	1.10 (0.70, 1.36)	12.5 (6.8, 31.0)	0.89 (0.61, 0.89)
PHR9	1.03 (0.59, 1.15)	21.8 (5.3, 96.7)	0.39 (0.25, 0.40)
PHR10	0.73 (0.60, 0.84)	9.1 (5.7, 15.3)	0.48 (0.37, 0.52)
PHR11	0.88 (0.59, 1.14)	15.1 (7.2, 40.8)	0.37 (0.32, 0.48)

4.3.3 Socio-demographic characteristics of potential spatial access to CRC services in Texas

Table 4.4 also shows potential spatial access to CRC services in Texas by race/ethnicity, SES, and rural/urban status. For PCPs, Asians (median spatial access ratio of 1.17; interquartile range (IQR), 0.94-1.41) and non-Hispanic blacks (median spatial

access ratio of 1.19; IQR, 0.89-1.42) had the highest potential spatial access whereas non-Hispanic whites (median spatial access ratio of 0.95; IQR, 0.51-1.29) had the lowest spatial access. Hispanics (median spatial access ratio of 1.07; IQR, 0.72-1.38) and Native Americans (median spatial access ratio of 1.01; IQR, 0.60-1.35) were comparable to each other. For CRC screening services, Asians had the shortest travel times (median of 8.8 minutes; IQR, 5.9-13.3 minutes) to screening facilities whereas non-Hispanic whites had the longest travel time (median of 14.2 minutes; IQR, 7.9-31.4 minutes). The other three groups were in the middle level. For potential spatial access to oncologists, Asians (median spatial access ratio of 1.03; IQR, 0.93-2.49) and non-Hispanic blacks (median spatial access ratio of 0.98; IQR, 0.77-2.49) were still the most advantageous groups. The other three groups were in comparable levels, with the median spatial access ratio ranging from 0.84 to 0.86.

Potential spatial access to CRC services was also different for population groups with different socioeconomic status and lived in different areas. As shown in Table 4.4, the richest group (Q1) had the best potential spatial access to all three CRC services and the poorer people (Q3 and Q4) had relatively lower potential spatial access. Increased rurality corresponded to decreased potential spatial access to the services. Potential spatial accesses to PCPs and oncologists were more than three times higher for metropolitan dwellers than for isolated-rural residents. Median travel time to CRC screening facilities was 56.9 minutes (IQR: 37.8-73.5 minutes) for isolated-rural residents compared to 10.5 minutes (IQR: 6.5-17.3 minutes) for people who live in metropolitan areas. Compared to socioeconomic and racial/ethnic differences of potential spatial access to CRC services, differences by rural/urban status were more obvious.

4.4 Discussions and Conclusions

Potential spatial access to medical services is a critical factor for identifying medically underserved areas and populations. Knowledge about social groups' differences in potential spatial access to cancer services can also facilitate the interpretation of cancer disparities. This chapter proposed a relative spatial access assessment approach to compensate uncertainties associated with different values of impedance coefficient in the E2SFCA method. The approach was then used to investigate the potential spatial access to CRC-related services in Texas based on data in 2000.

While the results indicate that spatial access index, the absolute presentation of spatial access, could be greatly influenced by the extent of distance impedance; spatial access ratio, the relative spatial access assessment approach, is stable to the variation of the distance impedance. Therefore, spatial access ratio is a better alternative to spatial access index in expressing the results of the E2SFCA method in the absence of an appropriate value of impedance coefficient. Spatial access ratio can be extended to represent the results of other gravity-based spatial access models.

The analysis results about the difference of potential spatial access for people living in rural and urban areas suggest that rural residents of Texas had the lowest potential spatial access to all of the three types of CRC services. This result corroborates previous findings about the unfavorable geographical conditions of non-urban dwellers in accessing health care services (Chan et al. 2006; Onega et al. 2008; Wang et al. 2008). In addition, potential spatial access to CRC services also varied among different racial/ethnic and socioeconomic groups but the differences were relatively minor comparing to the rural/urban variation.

Differences in potential spatial access to CRC services by race/ethnicity and geographic regions might chiefly reflect the rural-urban characteristics of racial/ethnic groups and PHR regions. For example, Asians had the best spatial access to all three types of CRC services, with the 25th percentile's accesses to PCPs and oncologists being only slightly lower than 1 and the 75th percentile's median travel time to CRC screening facilities being slimly higher than 13 minutes. This may be explained by that fewer than 4% of Asians are non-urban dwellers in Texas (US Census Bureau 2001a). Non-Hispanic blacks' advantage in the potential spatial accesses is also closely related with their urban-living habits. In addition, PHRs with the best potential spatial accesses to CRC services also encompass the metropolitan centers (for example, Houston, Dallas, San Antonio, and Austin) of Texas. These facts suggest that rurality might be the major reason for the unequal distribution of potential spatial access to CRC services in Texas. Government interventions aiming at alleviating the inequality of potential spatial access should target at isolated rural and small town rural areas.

Meanwhile, there are several points need to be considered when interpreting the results of this chapter. First, although the travel time breaks for potential spatial access to PCPs were based on empirical studies (Lee 1991) and geographical characteristics of Texas, the determination of travel time breaks (i.e. 30, 60 and 120 minutes) for oncologists were relatively arbitrary. Currently, there is limited information about how cancer patients tradeoff between travel cost and oncologists or treatment sites. Three hours' traveling time, which is estimated as an half of a one day roundtrip traveling time (i.e., six hours), might be conservative in defining the catchment size for oncologists in this study. Patient surveys and cancer center attendance data are needed to determine the

traveling threshold(s) tolerated by patients and the appropriate time breaks for cancer service utilization in the future.

Second, the E2SFCA method, like other gravity-based models, does not differentiate individuals with and without vehicles. Individuals without personal automobiles, who are not uncommon in major metropolitan areas, would have to rely on public transportation to access medical services. In this case, their potential spatial access could be greatly compromised. This effect might be more evident for blacks, who have the lowest car-ownership rates among all racial/ethnic groups (US Census Bureau 2001b). In other words, black people's potential spatial access to CRC services might be overly estimated in this study. A potential solution to this problem is to introduce car-ownership information into the spatial access model. For example, one can incorporate the percentage of households with at least one vehicle, which is available from the census 2000 data, as a personal mobility weight in the E2SFCA method to account for this problem. In addition, gravity-based models assume ideal travel conditions (e.g., roads without any traffic congestions) for all individuals. This would lead to some biases in travel time estimation because, in reality, traffic conditions differ across different areas (For example, urban areas are more congested than low traffic micropolitan areas).

Third, this study did not differentiate the effectiveness and costs of CRC screening facilities, which may exert significant influence on the utilization of these facilities. For example, high-tech facilities such as colonoscopy have higher accuracy in diagnosing CRC but are expensive. Patients are also required to visit a hospital or clinic before being examined by these facilities. FOBT, on the other hand, is more economically affordable and can be conveniently offered by nurses or health professionals without an office visit

(Center for Disease Control and Prevention (CDC) 2010). The results would be more reasonable if these differences are taken into account when calculating potential spatial access to CRC screening facilities.

Despite the limitations, this chapter proposes an important improvement for gravity-based spatial access models. The spatial access ratio, albeit mathematically simple, is more stable to the impedance coefficient than spatial access index when used along with the E2SFCA method. This chapter also for the first time examined potential spatial access to cancer services at a relatively fine geographic scale in the United States. In addition, the state of Texas, which has a racially/ethnically diverse population residing in areas with heterogeneous rural/urban characteristics, represents an excellent study area for examining inequalities in potential spatial access to medical services.

CHAPTER 5

DISPARITIES OF COLORECTAL CANCER STAGE AT DIAGNOSIS IN TEXAS

5.1 Introduction

Cancer stage at diagnosis is a critical factor influencing cancer outcomes (Dimou et al. 2009; Govindarajan et al. 2003; McDavid et al. 2003). This influence is especially pronounced for colorectal cancer (CRC), for which the five-year survival rate is 90% for patients diagnosed at local stages but only 10% for those who present distant or more advanced stages at diagnosis (ACS 2005).

Poor access to CRC prevention services has long been suspected responsible for the high risk of late stage CRC diagnosis among some social groups. However, few US studies have investigated the impact of potential spatial access to CRC services in conjunction with non-spatial factors on CRC diagnosis stage. Recent studies have observed socioeconomic, racial/ethnic, and geographic disparities of CRC stage at diagnosis (Clegg et al. 2009; Dimou et al. 2009; Govindarajan et al. 2003; Henry et al. 2009; Schwartz et al. 2003; Singh et al. 2005; Singh et al. 2006) but few of them have investigated whether these disparities were associated with potential spatial access to prevention services or not.

This chapter systematically examines disparities of CRC stage at diagnosis in Texas by SES, race/ethnicity, and geographic location, and assesses the association

between these disparities and potential spatial access to CRC prevention services. The remaining of this chapter is composed of three sections. A detailed description of the data sources and methodology is given in section 5.2. Section 5.3 describes the results of disparities and the impacts of potential spatial access to CRC services on the disparities. Section 5.4 discusses the findings and concludes this chapter.

5.2 Data and Methodology

5.2.1 Data

The analyses of this chapter involved three datasets including CRC incidence data, socio-demographic data, and data about potential spatial access to CRC prevention services. CRC incidence data of Texas from 1995 to 2003 have been obtained from the Texas Cancer Registry and geocoded in Chapter 3. As this research involved individual level information such as age, race, Hispanic ethnicity, and stage at diagnosis, cases with missing or unclear information of these factors were excluded from the analyses (n=9,403). Paired *t*-tests did not reveal any significant differences by sex (p=0.89), age (p=0.99), race/ethnicity (p=0.42), area socioeconomic status (p=0.99), urban/rural designation (p=0.99), and potential spatial access to CRC prevention services (p=0.92) between the excluded cases and the retained cases.

Census tract level poverty rate was adopted to represent area SES in this study. Census tracts' potential spatial accesses to CRC screening facilities and PCPs have been computed by the methods presented in Chapter 4.

5.2.2 Methodology

Analyzing basic characteristics of CRC late stage diagnosis in Texas

Summary statistics were generated to describe the overall study population. Census tracts were categorized into quartiles according to poverty rate, potential spatial access to PCPs, and potential spatial access to CRC screening facilities, respectively. Rurality categories include metropolitan, micropolitan, small town rural, and isolated rural census tracts (Hart 2006). Racial/ethnic groups were limited to non-Hispanic whites, non-Hispanic blacks, Hispanics, Asians, and Native Americans. Associations between case characteristics and stage at diagnosis were investigated by Chi-square tests. Tests for trending were implemented for poverty rate, rurality, and potential spatial access to CRC prevention services by entering these indicators as ordinal variables in the regression analyses. Factors that showed no significant associations with late stage CRC diagnosis would be excluded from subsequent analyses.

Analyzing the influence of spatial access to CRC services on racial/ethnic and socioeconomic disparities of CRC stage at diagnosis

Age- and sex- adjusted odds ratios of late stage CRC diagnosis by race/ethnicity, SES, and potential spatial access to CRC prevention services were estimated using the generalized estimating equation (GEE) logistic regression. GEE regressions can model the relation between the average response of a population and relevant covariates without considering the correlation of covariates across the high levels (Roux 2002; Zeger et al. 1988). Since this study includes both census tract level covariates (i.e., poverty rate, potential spatial access to CRC prevention services) and individual level characteristics (i.e., age, race/ethnicity), GEE logistic regression is more suitable than the traditional logistic regression for the analyses.

Five GEE models were implemented in this study, with three of them for assessing the individual influence of race/ethnicity, SES, and potential spatial access to CRC prevention services and two for analyzing the joint effects of these factors. The first multi-factor model (Model I) included sex, age, poverty rate, and race/ethnicity as independent variables. The second multi-factor model (Model II) included all variables of model I plus potential spatial access to CRC services as independent variables.

Analyzing the influences of potential spatial access to CRC services on geographic disparities of CRC stage at diagnosis

The geographic variations of CRC stage at diagnosis was measured using the adaptive spatial filtering method (Rushton and Lolonis 1996; Talbot et al. 2000). When implementing this method, the background population was defined as all CRC incidence cases, and the case events were defined as cases diagnosed with late stage CRC. The grid size of was set to be fifteen miles. The threshold case number was set to be twenty two. To minimize the influence of the small-number problem, an indirect age-sex standardization method was used to compute the Standard Mortality Rate (SMR) (note: the term "Standard Mortality Rate" does not necessarily refer to mortality). Monte Carlo simulations (999 times) were implemented to generate the reference distribution. The adaptive spatial filtering method is available in the Disease Mapping and Analysis Program (DMAP) software, version IV (Cai 2007).

5.3 Results

Characteristics of CRC stage at diagnosis in Texas

Selected characteristics of CRC stage at diagnosis in Texas are shown in Table 5.1. The overall rate of late stage CRC diagnosis is 59.0% for the 60,298 study cases. Approximately 73.2% of the cases were non-Hispanic whites. Asian and Native

American patients only accounted for 1.3% of the total patients. The mean age at diagnosis was 68.9 for non-Hispanic whites, 64.8 for non-Hispanic blacks, 64.0 for Hispanics, 62.5 for Native Americans, and 61.7 for Asians. Almost 78% of the cases were living in metropolitan areas. For census tract level poverty rate, rurality, and potential spatial access to primary care, the fewest CRC cases were found in the most disadvantaged quartiles.

Table 5.1 Selected characteristics of colorectal cancer late stage diagnosis in Texas, 1995-2003

Variable	Cases	(%)	% Late Stage Diagnosis
Age (p<0.01)			
Group 1 (<50)	6,528	10.8	65.1
Group 2 (50-60)	9,644	16.0	61.0
Group 3 (61-70)	1,4401	23.9	58.4
Group 4 (70-80)	1,7607	29.2	57.4
Group 5 (>80)	1,2118	20.1	57.0
Sex			
Male	30,765	51.0	58.7
Female	29,533	49.0	59.2
Race/Ethnicity (p<0.001)			
Non-Hispanic White	44,136	73.2	57.5
Non-Hispanic Black	69,30	11.5	63.5
Hispanic	8,457	14.0	62.5
Asian	740	1.2	59.1
Native American	35	0.1	68.6
Area SES (p<0.001)			
Q1 (High SES)	15,370	25.5	56.5
Q2	16,845	27.9	59.0
Q3	15,458	25.6	59.2
Q4 (Low SES)	12,625	20.9	61.5
Rurality			
Metropolitan	46,767	77.5	59.1
Micropolitan	6,263	10.4	58.3
Small Town	3,862	6.4	58.8
Isolated Rural	3,412	5.7	57.9
Potential Spatial Access to PCPs (p<0.05)			
Q1 (High Access)	15,120	25.1	57.6
Q2	15,730	26.1	59.1
Q3	14,743	24.5	59.1
Q4 (Low Access)	14,705	24.4	60.1
Potential Spatial Access to CRC Screening Facilities			
Q1 (High Access)	15,781	26.2	57.9
Q2	14,925	24.8	59.6
Q3	14,786	24.5	59.3
Q4 (Low Access)	14,806	24.6	59.2

The univariate analyses revealed significant differences in late stage CRC diagnosis by age, race/ethnicity (limited to non-Hispanic whites, non-Hispanic blacks, and Hispanics here), area SES, and potential spatial access to primary care (Table 5.1). Groups with high proportions of late stage CRC cases were people younger than 50 years of age (65.1%), non-Hispanic blacks (63.5%), and Hispanics (62.5%). Native Americans had the highest rate of late stage CRC diagnosis (68.6%) among all racial/ethnic groups, but the differences were not significant (p=0.19). As census tract poverty rate increased, the rate of late stage CRC diagnoses increased (p<0.01). The enhancement of potential spatial access to PCPs lowered the risk of late stage CRC diagnosis (p<0.05). The analyses did not reveal any significant relationship between late stage CRC diagnosis and sex, rurality, or potential spatial access to CRC screening facilities.

Influence of potential spatial access to PCPs on racial/ethnic and socioeconomic disparities of CRC stage at diagnosis

The results of multivariate analyses are shown in Table 5.2. Asian and Native American patients were excluded from this step because of their small proportion of the study population. Age- and sex-adjusted regressions revealed significant disparities in CRC stage at diagnosis by SES, race/ethnicity, and potential spatial access to PCPs. Non-Hispanic black (OR: 1.25, 95% CI: 1.19 to 1.32) and Hispanic (OR: 1.20, 95% CI: 1.14 to 1.26) CRC patients had higher risk of late stage diagnosis than non-Hispanic white CRC patients. Worsening SES corresponded to increasing likelihood of presenting with a late stage CRC at the time of diagnosis ($p < 0.01$). Less potential spatial access to PCPs also led to higher likelihood of late stage CRC diagnosis ($p < 0.05$).

Table 5.2 Odds ratios of late stage colorectal cancer diagnosis by race/ethnicity, area socioeconomic status (SES), and potential spatial access to primary care

	Age- and sex- adjusted OR (95% CI)	OR of Model I (95% CI) [1]	OR of Model II (95% CI) [2]
Race/Ethnicity**			
Non-Hispanic White	1	1	1
Non-Hispanic Black	1.25 (1.19, 1.32)	1.21 (1.14, 1.27)	1.21 (1.15, 1.28)
Hispanic	1.20 (1.14, 1.26)	1.15 (1.09, 1.21)	1.15 (1.09, 1.22)
SES**			
Q1 (High SES)	1	1	1
Q2	1.12 (1.07, 1.17)	1.11 (1.06, 1.16)	1.11 (1.06, 1.16)
Q3	1.14 (1.09, 1.19)	1.11 (1.06, 1.16)	1.11 (1.06, 1.16)
Q4 (Low SES)	1.25 (1.19, 1.31)	1.15 (1.09, 1.21)	1.16 (1.10, 1.22)
Potential spatial access to PCPs*			
Q1 (High access)	1		1
Q2	1.06 (1.01, 1.11)		1.05 (1.00, 1.10)
Q3	1.06 (1.00, 1.10)		1.05 (1.00, 1.11)
Q4 (Low access)	1.09 (1.00, 1.12)		1.06 (1.01, 1.11)

*$p<0.005$; **$p<0.001$

[1]Model I includes sex, age, poverty rate, and race/ethnicity as independent variables.

[2]Model II includes all variables from model I plus potential spatial access to health services as independent variables.

The result of Model I indicates that, after the incorporation of SES, racial/ethnic disparities still remained significant despite a slight reduction in the odd ratios for non-Hispanic blacks and Hispanics. The interactions between SES and race/ethnicity also led to minor reductions of the risk of being diagnosed with late stage CRC for patients from low SES areas.

The addition of potential spatial access to PCPs in the analysis (Model II) did not exert significant influence on the racial/ethnic and socioeconomic disparities. Odds of late stage CRC diagnosis for disadvantaged racial/ethnic groups (i.e., Hispanics and non-Hispanic blacks) and low SES groups (Q3 and Q4) fluctuated slightly between Model I and Model II. For CRC patients with the lowest potential spatial access to PCPs (Q4), the risk of late stage diagnosis dropped slightly but still remained significant.

The geographic patterns of CRC late stage diagnosis

The overall spatial pattern of CRC late stage diagnosis is shown in Figure 5.1. The map reveals a heterogeneous spatial distribution of late stage CRC diagnosis after adjusting for age and sex. Generally, the northernmost panhandle, the western tip, and the southern part of Texas had elevated risk of late stage CRC diagnoses (relative risk (RR)> 1.19) relative to the state mean. Geographic areas with low risks of late stage CRC diagnosis (RR< 1) were primarily dispersed in the central-eastern part of Texas.

The adaptive spatial filtering method also identified six geographic areas as having significantly different risks of late stage CRC diagnosis compared to the statewide level (Figure 5.2). Among them, Areas 1, 2, 3, and 4 were high-risk areas with the relative risks of 1.49, 1.43, 1.50, and 1.43, respectively. Areas 5 and 6 were low risk areas with the relative risks of 0.57 and 1.42, respectively. The geographic distributions of these areas are consistent with the geographic pattern across the whole state. For example, the higher-risk areas were located at the northern panhandle (Area 1), the middle-western part (Area 2), the southernmost tip (Area 3), and the northern shoulder area (Area 4) of Texas. The low risk areas were located at the southwestern (Area 5) and the northeastern corners (Area 6) of the state. A comparison of these areas indicated that areas with lower-than-expected risks of late stage CRC diagnosis had higher proportions of non-Hispanic white residents, lower proportions of Hispanic residents, lower poverty rate, and higher spatial access to PCPs than areas with higher-than-expected risks (Table 5.3).

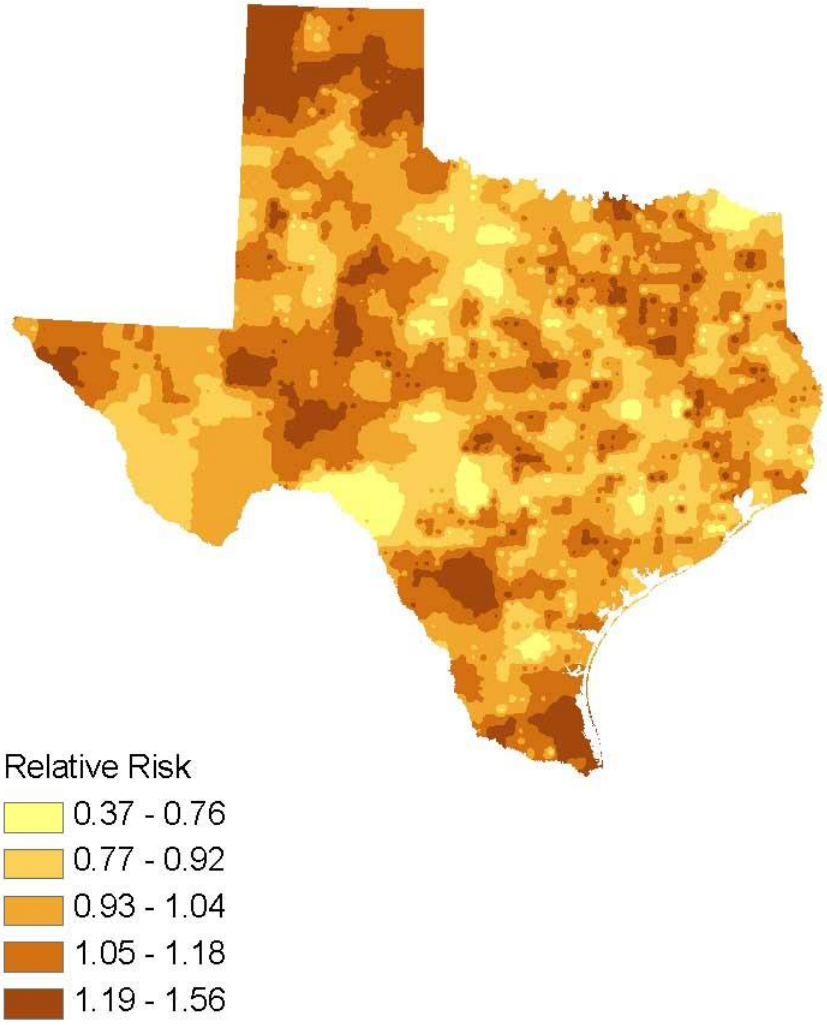

Figure 5.1 Geographic pattern of late stage colorectal cancer diagnosis in Texas (1995-2003)

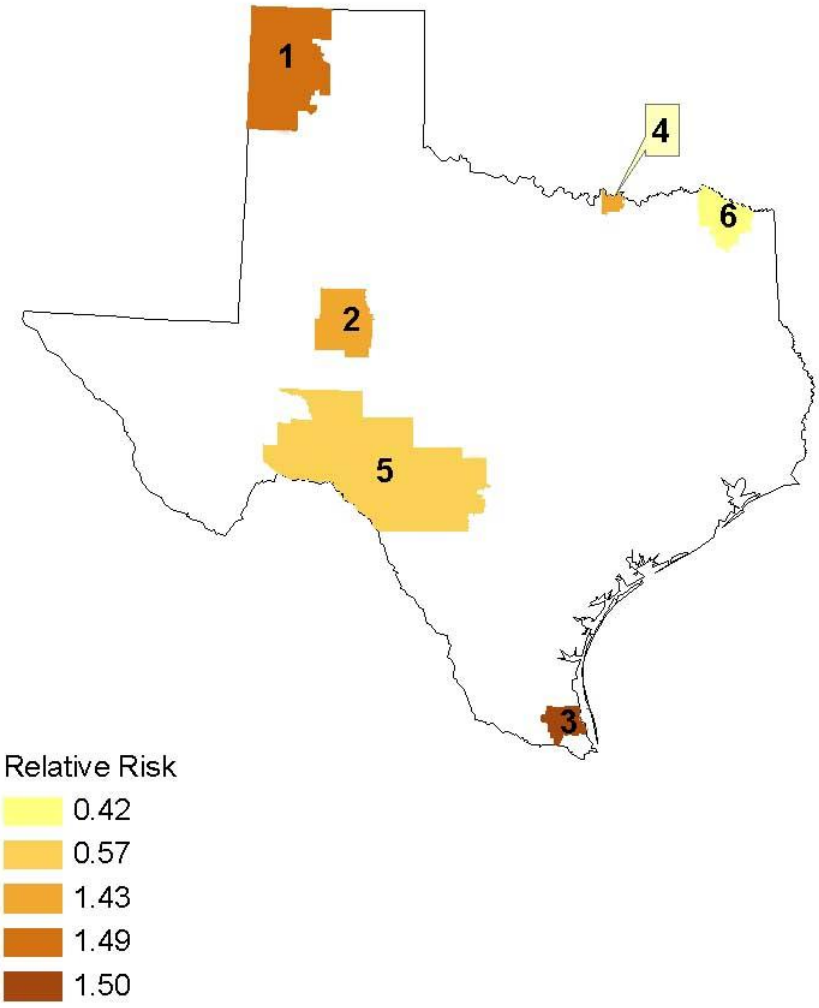

Figure 5.2 Spatial clusters of late stage colorectal cancer diagnosis in Texas (1995-2003)

Table 5.3 Characteristics of census tracts with significantly different risks of colorectal cancer late stage diagnosis in Texas

	Low risk areas	**High risk areas**	**Statewide**
Percentage of late stage cases	46.1	73.0	59.01
Total cases	605	642	60,298
Observed late stage cases	279	470	35,546
Expected late stage cases	553	327	35,546
Observed/Expected	0.50	1.44	1
p-value	<0.01	<0.01	
Percentage of non-Hispanic whites	58.9	46.0	49.28
Percentage of non-Hispanic blacks	5.5	1.4	10.87
Percentage of Hispanics	34.0	51.1	30.66
Poverty rate	18.5	21.3	14.9
Mean spatial access ratio to PCPs	0.65	0.54	0.98

5.4 Discussions and Conclusions

This chapter analyzed disparities of CRC stage at diagnosis from various factors, including race/ethnicity, SES, geographic location, and potential spatial access to CRC prevention services. The results revealed significant disparities of CRC stage at diagnosis from most of these factors. Non-Hispanic black and Hispanic CRC patients had significantly higher risks of being diagnosed at late stages than non-Hispanic white patients. Lower area SES corresponds to increased likelihood of late stage CRC diagnosis. Decreased potential spatial access to primary care also leads to elevated risk of late stage CRC diagnosis. There was an uneven geographic distribution of the risk of late stage CRC diagnosis across the whole state of Texas.

The results regarding the racial/ethnic and socioeconomic disparities of CRC stage at diagnosis in this chapter corroborate most of previous studies (Clegg et al. 2009; Schwartz et al. 2003; Wu et al. 2006), which revealed unfavorable risks of late stage CRC

diagnosis for racial/ethnic minorities and low socioeconomic groups. However, other researchers also reported different findings. In a study of CRC stage at diagnosis in Maine, Parsons and Askland (2007) found no significant associations between CRC stage and socioeconomic variables. A study in New Jersey suggested that race/ethnicity has no significant influence on CRC stage at diagnosis (Henry et al. 2009). These facts, along with the findings of this chapter, imply that racial/ethnic and socioeconomic disparities may be related with some local characteristics such as demographic structure, rurality, and some local health promotion programs (Parson and Askland 2007). A thorough investigation of these factors may reveal some socio-political (e.g. access to Medicare/Medicaid) or medical factors that could effectively prevent late stage CRC diagnosis among vulnerable population groups.

Prior studies on the interactive influence of race/ethnicity and SES on CRC indicate that SES may explain a large portion of racial/ethnic disparities in late stage CRC diagnosis (Dimou et al. 2009; Schwartz et al. 2003). However, this study indicated that non-Hispanic blacks' and Hispanics' unfavorable situation in CRC late stage diagnosis underwent little change before and after adjusting for area SES, suggesting that there might be some other factors contributing to the racial/ethnic disparities of CRC diagnosis in Texas. Cultural background, immigration status, and religious belief, in addition to SES, may provide potential reasons for the unexplained racial/ethnic disparities in this study. Surveys on how these factors (even within the same racial/ethnic group) influence individuals' life style as well as attitude on CRC screening are necessitated to account for the racial/ethnic disparities in future studies.

Inadequate access to CRC prevention services has been implied as the primary reason for the high risks of late stage CRC diagnosis among some racial/ethnic minorities and low SES groups (James et al. 2006; Matthews et al. 2005). However, this study suggests that potential spatial access to CRC prevention services, the basis for the service utilization, has limited influence on CRC diagnosis. Potential spatial access to primary care was related with the risk of late stage CRC diagnosis, but the relationship was much weaker than those with SES or race/ethnicity. The incorporation of potential spatial access to primary care into the regression did not attenuate the risks of late stage diagnosis for Hispanics, non-Hispanic blacks, and individuals from the poorest areas. These findings suggest that non-spatial factors such as the financial capability of potential patients might play a more important role in influencing the utilization of these services than spatial factors. Prior studies also indicated that language barriers and low literacy rates could effectively prevent one from navigating the healthcare system and communicating with health professionals (James et al. 2006; Matthews et al. 2005). Wang and Luo (2005) have proposed a systematical scheme for integrating spatial and non-spatial factors to characterize people's overall access to medical services. Their framework provides a potential way to extend the exploration of disparities beyond our study.

This study also found that potential spatial access to PCPs has a stronger association with the risk of late stage CRC diagnosis than do potential spatial access to CRC screening facilities. This finding, along with a previous study (Wang et al. 2008), highlights the importance of medical professionals who introduce patients to the cancer screening procedure. However, the geographic pattern indicates that residents of some

parts of Texas had limited potential spatial access to primary care, suggesting the necessity of promoting a more balanced supply of PCPs across the whole state.

The insignificant association between potential spatial access to CRC screening facilities and CRC stage at diagnosis might be explained by the fact that Texas residents had adequate potential spatial access to CRC screening services. For the majority of Texans, the shortest travel time to CRC screening facilities is less than 30 minutes, which has been suggested as a threshold for people's tolerance travel time for accessing medical service (Lee 1995). In other words, most residents of Texas can easily access screening facilities. In this case, differences in the shortest travel time to CRC screening facilities might not influence CRC stage at diagnosis much.

The geographic analyses revealed an uneven spatial distribution of CRC stage at diagnosis in Texas. Some areas had significantly different risks of late stage CRC diagnosis than others. Areas with elevated likelihood of late stage CRC diagnosis were characterized by low percentages of non-Hispanic whites, high proportions of Hispanic residents, and high poverty rate. Meanwhile, areas with a significant deficit of late stage CRC diagnosis had higher proportions of non-Hispanic whites, lower percentage of Hispanics, and lower poverty rates compared to the state level. This contradiction provides further evidence for the aforementioned comments on socioeconomic and racial/ethnic disparities.

This chapter includes several advancements relative to prior studies on disparities of CRC diagnosis. For the first time, the relative importance of potential spatial access to medical services on disparities of CRC stage at diagnosis in the United States has been assessed. Information about whether and how the spatial dimension of people's access to

medical services influences CRC incidence has important implications for health resource allocation and CRC disparity reduction in Texas. Second, this chapter adopted different spatial access measures for different CRC prevention services. The E2SFCA method and the shortest travel time method were employed to estimate potential spatial access to PCPs and CRC screening facilities, respectively. This differentiation is based on varying utilization rates of the two medical resources by individuals. Prior studies have used the shortest travel time to represent potential spatial access to all CRC prevention services, including primary care (Parson and Askland 2007). This method is problematic because it ignores the supply-to-demand ratios of PCPs – an important factor for primary care accessibility. In addition, the E2SFCA method is more intuitive and explicit in revealing potential spatial access to medical services than previous models (Luo and Qi 2009).

In conclusion, this chapter represents a comprehensive examination of disparities in CRC stage at diagnosis. It revealed significant disparities by race/ethnicity, area SES, and geographic location. It also confirmed a significantly negative association between potential spatial access to medical services and the risk of CRC late stage diagnosis. However, this study did not reveal any mediating effect of the spatial access factor on the racial/ethnic, socioeconomic, and geographic disparities. This finding suggests that policy makers and health planners who aim at minimizing disparities of CRC stage at diagnosis should focus attention more on socio-cultural factors rather than facility resources for optimization of CRC prevention services.

CHAPTER 6

DISPARITIES OF COLORECTAL CANCER SURVIVAL IN TEXAS

6.1 Introduction

Previous US studies have revealed substantial disparities of CRC survival in relation to SES, race/ethnicity, and geographic location, suggesting that racial/ethnic minorities, individuals with low SES, and individuals in some geographic regions had poorer survival than others (Clegg et al. 2002; Du et al. 2007; Henry et al. 2009; Huang et al. 2007; McDavid et al. 2003; Wudel et al. 2002). Factors that may explain CRC survival disparities include CRC stage at diagnosis, patients' access to treatment services, and some factors caused by low SES. Individuals diagnosed with a late stage CRC and those who did not get timely and appropriate treatment had a higher chance of dying from CRC than those diagnosed at early stages and those received good treatment services.

As discussed in Chapter 2, a major limitation of prior US studies on CRC survival disparity is that they seldom analyzed the joint impact of SES, race/ethnicity, geographic location. Focusing on only a portion of these factors may lead to partial or biased conclusions about CRC survival disparity. In addition, the role of potential spatial access to cancer treatment services on CRC survival has never been examined in the United States.

The primary goal of this paper is to investigate how potential spatial access to

cancer treatment services, along with non-spatial access factors, influences CRC survival at the census tract level. A factor analysis method was conducted to estimate non-spatial factors that influence the service utilization among patients in Texas. The study assesses the impacts of spatial and non-spatial factors on CRC survival in different rural/urban settings. The rest of this chapter is made up of three sections. The data and methods are described in section 6.2. The results of CRC survival disparities are given in section 6.3. Section 6.4 offers the discussion of the results.

6.2 Data and Methodology

6.2.1 Data

CRC Survival Data

The Texas CRC incidence data from January 1, 1995 through December 31, 2003 were obtained from the Texas DSHS. The data have been preprocessed in Chapters 3 and 5. The study population for the analyses in this chapter include 60,298 CRC cases with specific information about race (i.e., white, black, Native American, Asian, or others), Hispanic ethnicity, sex, age, stage at diagnosis (early stage or late stage), date of last contact, vital status (deceased or alive) at last contact, cause of death (if deceased), date of death (if deceased), and the census tract number. The TCR uses a passive follow-up approach to ascertain the vital status of cancer patients but also links its cancer registry information with the Texas vital statistics data, the National Death Index data, and the Social Security Death Index data. The patient follow-up period of the CRC data used in this study was from January 1, 1995 through March 1, 2010.

Cancer specific survival is adopted as the measure of survival in this study. Cancer specific survival (or cause specific mortality) is an approach for measuring the

survival from a specific underlying cause of death other than other causes of death (NCI 2010). Cancer specific survival has been proved useful for analyzing cancer disparities and has been adopted by NCI as a "policy based statistic" (NCI 2010).

This research specifies CRC as the cause of death. Both long-term and five-year CRC-specific survivals would be analyzed. For the long-term survival analysis, the survival time (in month) of a patient was measured from the date of CRC diagnosis and was censored at the date on which the case lost to the follow-up, the date the case die from other causes, or the last day of the follow-up (which was March 1, 2010), whichever occurred first. For the five-year survival analysis, the survival time (in month) of a patient was measured from the date of CRC diagnosis and was censored at the date on which the case lost to the follow-up, the date the case die from other causes, or the last day of the five-year period starting from the date of diagnosis, whichever occurred first.

Spatial Access Data

CRC treatment services were represented by oncologists in this study. The oncologist data of Texas in 2000 have been collected from Texas DSHS and have been geocoded in Chapter 3. The geocoding successfully located 204 oncologists (in 120 practicing sites). Potential spatial access to oncologists at the census tract level has been calculated in Chapter 4 using the E2SFCA method (Luo and Qi 2009) with a distance impedance coefficient (β) of 4,890. The relative spatial access (spatial access ratio) was used to present the potential spatial access for each census tract.

Non-spatial Access Data

Generally, a person's non-spatial access to medical services can be measured from a range of characteristics such as socioeconomic status, environment, education

level, English fluency, and transportation (Khan 1992; Wang et al. 2008). Factor analysis (FA) and principal component analysis (PCA) were conducted on several single socio-cultural indicators of census tracts to uncover important dimensions of non-spatial access in this chapter. Briefly, given a number of tract characteristics, PCA can generate the same number of indicators (components) with no colinearity to capture their variances; FA can further produce a smaller number of components to reflect the major aspects of the indicators. Based on previous studies (Wan et al. 2011; Wang et al. 2008), nine non-spatial access indicators (i.e., poverty rate, unemployment rate, median household income, median home value, percent of persons with complete high school education, percent of persons with complete college education, percent of linguistic isolated households, percent of households with more than one persons per room, and percent of households without vehicles) were used in the analysis. These indicators were derived from Summary File 3 of the census 2000 data (US Census Bureau 2001b).

According to the PCA, two components, which explain more than 74% of the original indicators, have eigen values greater than one. Therefore, these two components were retained for the subsequent factor analysis. Then, the varimax rotation method was employed to maximize the loadings of the two factors. The resulting factor loadings and the proportion of variance accounted by each factor are shown in Table 6.1. The two factors can be labeled socioeconomic factor and socio-environmental factor, respectively. The socio-environmental factor primarily reflects the cultural aspects such as linguistic isolation which may prevent some patients from effectively utilizing the healthcare services, while the socioeconomic factor primarily captures the economic aspects such as home value and household income which may influence the affordability of patients for

medical services. These two factors would be used as the non-spatial access factors in subsequent analyses.

Table 6.1 Factor loadings and the proportion of variances explained by the principal components in the factor analysis

	Factors	
Single Indicators	**Socio-environmental factor**	**Socioeconomic factor**
Poverty rate	**0.79**	-0.40
Unemployment rate	0.64	-0.29
Median household income	-0.37	**0.84**
Median home value	-0.13	**0.92**
Percent of persons with complete high school education	-0.75	0.49
Percent of persons with complete college education	-0.32	**0.85**
Percent of linguistic isolated households	**0.87**	-0.07
Percent of households with more than one persons per room	**0.85**	-0.19
Percent of households without vehicles	**0.69**	-0.29
Proportion of total variance explained by the factor	42.1%	32.1%

6.2.2 Methodology

The analyses of this chapter were implemented in three parts. The first part analyzed the basic characteristics of CRC survival in Texas. Kaplan-Meier estimators (Kaplan and Meier 1958) were employed to calculate CRC-specific survival rates and the corresponding 95% confidence intervals (CI) by sex, age, race/ethnicity, and access to medical services. In the analyses, Texas census tracts were categorized into quartiles according to the spatial and non-spatial access factors. Kaplan-Meier estimators were used for all stages and each of the four SEER stages, respectively.

The second part assessed the associations between CRC survival and health care access using Cox proportional hazard regressions. To differentiate the influence of urban/rural settings, the analysis was implemented for all areas, urban areas, and non-urban areas, respectively. The urban/non-urban status of census tracts was based on the Rural Urban Commuting Area (RUCA) classification (Hart 2006). More detailed rurality categories were not used in this chapter because of the small numbers of CRC cases in some categories. For each urban/rural setting, the analysis was composed of three steps. The first step estimated age-, sex- and stage-adjusted hazard ratios (HRs) of five-year CRC survival by race/ethnicity and each of the healthcare access indices (Model I). The second step incorporated age, sex, stage at diagnosis, race/ethnicity, and the non-spatial access factor into the model (Model II). The third step used all of the factors of Model II plus the spatial access factor into the regression (Model III). The change of the HRs was traced to assess the relative importance of each factor for CRC survival disparity among urban and rural patients. The Kaplan-Meier estimations and the Cox hazard proportional hazard regressions were accomplished with SPSS, version17.0 (SPSS 2008).

The third part of the analyses focused on the spatial pattern of CRC survival. The spatial scan statistic (exponential model) was employed to determine whether there were any significant geographic variations of CRC survival in Texas. Since this study's purpose was to detect areas with significantly different survival times, both Ha (Ha: $\theta_{in} < \theta_{out}$) and Hb (Hb: $\theta_{in} > \theta_{out}$) of the spatial scan statistic were tested. The statistical significance was tested by the Monte Carlo permutation method (Huang et al. 2007). The spatial scan statistic analyses were implemented using SaTScan, version 9.0 (Kulldorff 1997).

Four spatial scan statistic models were employed to identify areas with significantly longer or shorter CRC-specific survival rates, while sequentially adjusting for important covariates using an exponential model (Huang et al. 2007a). Specifically, the first model (Model *a*) used sex- and age-adjusted survival times. The second model (Model *b*) adjusted for sex, age, and stage at diagnosis. The covariates adjusted in the third model (Model *c*) involved all factors of the second model plus race/ethnicity. The fourth model (Model *d*) included all factors of the third model plus the non-spatial access factor. The covariate adjustments followed the procedures of Huang et al. (2007a) using MATLAB software, version 7.0 (Mathworks 2010).

Areas with significantly longer or shorter survival times after the covariate adjustments were mapped using ArcGIS 9.3 (ESRI 2009). Relevant indicators such as the number of cases, the observed/expected ratio of CRC mortality as determined by SaTScan, the racial/ethnic distribution, and the population-averaged non-spatial access factor were reported and compared between the high- and low-risk areas.

6.3 Results

6.3.1 Descriptive and regression results

Table 6.2 lists the five-year survival rates categorized by different factors. As can be seen in Table 6.2, the risk of five-year CRC-specific mortality is associated with stage at diagnosis, age, race/ethnicity, and access to healthcare services. Non-Hispanic whites had a significantly higher five-year survival rate than other racial/ethnic groups. Patients with different spatial and non-spatial access to healthcare services also differed significantly in the five-year survival rate. In addition, the rate of five-year survival

ranged from as high as 97.1% (i.e., patients younger than 50) for in-situ stage groups to as low as 5.1% (i.e., patients older than 80) for distant stage groups.

Table 6.2 Five-year survival rates of early stage and late stage colorectal cancer cases by sex, age, race/ethnicity, and access to healthcare services in Texas, 1995-2003

	Five-year survival rate (%) and 95% Confidential Intervals				
	All stages (n=56,734)	In-situ stage (n=3,151)	Localized stage (n=19,901)	Regional stage (n=23,555)	Distant stage (n=10,127)
Total population	58.9 (58.4, 59.4)	93.7 (92.6, 94.6)	83.3 (82.6, 83.9)	58.4 (57.3, 58.8)	9.2 (8.6, 9.9)
Sex					
Male	58.8 (58.2, 59.5)	94.0 (92.5, 95.2)	83.1 (82.2, 84.0)	58.4 (57.3, 59.4)	8.7 (7.8, 9.6)
Female	58.9 (58.3, 59.6)	93.4 (91.7, 94.8)	83.4 (82.5, 84.3)	57.7 (56.6, 58.8)	9.8 (8.8, 10.8)
p-value	0.14	0.44	0.96	<0.01	0.43
Age					
<50	50.4 (48.8, 52.0)	97.1 (92.5, 98.9)	84.1 (81.6, 86.4)	55.2 (52.6, 57.6)	12.2 (10.4, 15.2)
50-60	57.7 (56.5, 59.0)	96.5 (93.9, 98.0)	87.8 (86.1, 89.2)	60.1 (58.1, 62.0)	10.4 (8.9, 12.0)
61-70	61.3 (60.4, 62.3)	95.1 (92.9, 96.5)	85.8 (84.5, 87.0)	61.5 (60.0, 63.0)	9.6 (8.3, 11.0)
71-80	62.5 (61.6, 63.3)	94.9 (93.0, 96.3)	85.3 (84.2, 86.3)	59.8 (58.4, 61.1)	8.7 (7.5, 10.0)
>80	55.6 (54.5, 56.7)	86.1 (82.4, 89.0)	74.4 (72.8, 75.9)	51.4 (49.7, 53.1)	5.1 (3.8, 6.8)
p-value	<0.001	<0.001	<0.001	<0.001	<0.001
Race/Ethnicity					
Non-Hispanic white	61.1 (60.6, 61.7)	94.3 (93.1, 95.3)	84.6 (83.8, 85.2)	59.5 (58.6, 60.3)	10.0 (9.3, 10.9)
Non-Hispanic black	50.9 (49.4, 52.3)	91.5 (87.4, 94.2)	76.5 (74.1, 78.7)	53.0 (50.7, 55.2)	7.3 (5.9, 9.0)
Hispanic	53.2 (51.8, 54.6)	92.3 (88.1, 95.1)	79.9 (77.7, 81.9)	55.0 (52.9, 57.0)	6.9 (5.4, 8.6)
Asian	55.4 (50.2, 60.4)	90.0 (47.3, 98.5)	88.8 (81.4, 93.4)	53.3 (45.0, 60.8)	9.5 (4.3, 17.1)
Native American	43.9 (21.1, 64.6)	No Cases	85.7 (33.4, 97.9)	23.6 (1.3, 62.0)	14.3 (7.1, 46.5)
p-value	<0.001	0.29	<0.001	<0.001	<0.001
Potential spatial access to oncologists					
Q1 (High access)	59.3 (58.3, 60.2)	94.9 (92.5, 96.6)	84.2 (82.9, 85.5)	61.2 (59.6, 62.7)	9.4 (8.1, 10.8)
Q2	59.7 (58.8, 60.5)	93.0 (90.6, 94.9)	84.0 (82.8, 85.1)	57.2 (55.7, 58.8)	8.4 (7.2, 9.7)
Q3	58.0 (57.0, 59.0)	93.8 (91.6, 95.5)	83.0 (81.7, 84.3)	56.5 (55.0, 57.9)	10.3 (9.1, 11.7)
Q4 (Low access)	58.5 (57.5, 59.4)	93.3 (91.0, 95.0)	81.6 (80.2, 82.9)	57.7 (56.2, 59.2)	8.7 (7.4, 10.1)
p-value	<0.001	<0.001	<0.001	<0.001	<0.001
Socioeconomic factor					
Q1 (High status)	59.5 (58.5, 60.4)	94.9 (93.0, 96.3)	82.9 (81.7, 84.1)	57.3 (55.8, 58.7)	10.2 (8.9, 11.6)
Q2	59.4 (58.5, 60.3)	92.7 (90.4, 94.5)	84.4 (83.2, 85.5)	58.5 (57.0, 59.9)	9.5 (8.3, 10.7)
Q3	59.4 (58.4, 60.4)	93.0 (90.5, 94.8)	84.0 (82.7, 85.3)	59.2 (57.7, 60.7)	8.4 (7.2, 9.8)
Q4 (Low status)	56.8 (55.7, 57.8)	94.3 (91.4, 96.2)	81.1 (79.6, 82.6)	57.1 (55.4, 58.7)	8.6 (7.2, 10.1)
p-value	<0.001	<0.001	<0.001	<0.01	<0.001
Socio-environmental factor					
Q1 (High status)	60.2 (59.2, 61.1)	95.4 (93.1, 96.9)	85.6 (84.3, 86.7)	58.9 (57.3, 60.4)	10.8 (9.4, 12.2)
Q2	59.2 (58.2, 60.2)	95.3 (93.1, 96.8)	83.7 (82.4, 85.0)	59.4 (57.9, 60.9)	9.3 (8.0, 10.7)
Q3	58.1 (57.1, 59.0)	92.5 (90.0, 94.4)	82.0 (80.6, 83.2)	57.3 (55.8, 58.7)	8.8 (7.6, 10.1)
Q4 (Low status)	58.2 (57.2, 59.1)	92.2 (89.8, 94.1)	81.9 (80.6, 83.2)	56.7 (55.2, 58.2)	8.0 (6.8, 9.4)
p-value	<0.001	<0.01	<0.001	<0.01	<0.001

Table 6.3 shows the HRs and the 95% confident intervals (CIs) of the Cox proportional regressions. Asians and Native Americans were excluded from the regression analyses because of their relatively smaller numbers. The socio-environmental factor was not involved in the regression because, according to a preliminary analysis, they are not significantly associated with CRC survival at neither setting. Age-, sex and stage-adjusted regression indicates that non-Hispanic black (HR: 1.33, 95% CI: 1.28 to 1.39) and Hispanic (HR: 1.16, 95% CI: 1.11 to 1.21) CRC patients had significantly higher risks of five-year CRC-specific mortality than non-Hispanic white patients. Patients with the lowest socioeconomic access to healthcare had slightly higher (HR: 1.09, 95% CI: 1.04 to 1.14) risk of five-year CRC-specific mortality than those with the highest socioeconomic access. The incorporation of both the socioeconomic factor and race/ethnicity (Model II) in the regression led to slight reductions of the HRs for disadvantaged groups but did not attenuate the significance of disparities by the two factors. The regressions revealed no significant differences for five-year CRC-specific survival based on potential spatial access to oncologists across the state.

Table 6.3 Hazard ratios (HRs) of colorectal cancer five-year survival by race/ethnicity and access to medical services

	HR of Model I[a] (95% CI)	**HR of Model II (95% CI)**	**HR of Model III (95% CI)**
All areas (n=55,941)			
Race/ Ethnicity			
Non-Hispanic white	1	1	1
Non-Hispanic black	1.33* (1.28, 1.39)	1.32* (1.27, 1.38)	1.32* (1.27, 1.38)
Hispanic	1.16** (1.11, 1.21)	1.16** (1.11, 1.21)	1.16** (1.11, 1.21)
Socioeconomic factor			
Q1 (High access)	1	1	1
Q2	1.03 (0.99, 1.08)	1.02 (0.97, 1.06)	1.02 (0.97, 1.06)
Q3	1.07** (1.03, 1.11)	1.05*** (1.00, 1.10)	1.05*** (1.00, 1.09)
Q4 (Low access)	1.09** (1.04, 1.14)	1.06** (1.02, 1.13)	1.06** (1.02, 1.11)
Spatial access to oncologists			
Q1 (High access)	1		1
Q2	1.02 (0.97, 1.06)		1.03 (0.98, 1.07)
Q3	1.00 (0.96, 1.04)		1.02 (0.97, 1.06)

Table 6.3-Continued

Q4 (Low access)	1.00 (0.96, 1.04)		1.01 (0.97, 1.06)
Urban areas (n=43,242)			
Race/ Ethnicity			
Non-Hispanic white	1	1	1
Non-Hispanic black	1.34* (1.27, 1.40)	1.32* (1.27, 1.39)	1.32* (1.27, 1.39)
Hispanic	1.17** (1.12, 1.23)	1.16** (1.11, 1.21)	1.16** (1.11, 1.21)
Socioeconomic factor			
Q1 (High access)	1	1	1
Q2	1.04 (0.99, 1.08)	1.02 (0.98, 1.07)	1.02 (0.98, 1.07)
Q3	1.07** (1.03, 1.12)	1.05*** (1.01, 1.10)	1.05*** (1.01, 1.10)
Q4 (Low access)	1.09** (1.04, 1.14)	1.06*** (1.01, 1.10)	1.06*** (1.01, 1.11)
Spatial access to oncologists			
Q1 (High access)	1		1
Q2	1.02 (0.97, 1.06)		1.03 (0.98, 1.07)
Q3	1.00 (0.96, 1.04)		1.02 (0.97, 1.06)
Q4 (Low access)	1.00 (0.95, 1.04)		1.00 (0.95, 1.05)
Non-urban areas (n=12,699)			
Race/ Ethnicity			
Non-Hispanic white	1	1	1
Non-Hispanic black	1.33* (1.20, 1.48)	1.33* (1.20, 1.49)	1.34* (1.20, 1.49)
Hispanic	1.13** (1.03, 1.25)	1.14** (1.02, 1.26)	1.13** (1.02, 1.26)
Socioeconomic factor			
Q1 (High access)	1	1	1
Q2	1.02 (0.98, 1.08)	1.01 (0.95, 1.08)	1.01 (0.96, 1.08)
Q3	1.04*** (1.00, 1.09)	1.03*** (1.00, 1.08)	1.03*** (1.00, 1.08)
Q4 (Low access)	1.07*** (1.01, 1.10)	1.06*** (1.01, 1.11)	1.06*** (1.00, 1.12)
Spatial access to oncologists			
Q1 (High access)	1		1
Q2	0.95 (0.89, 1.13)		0.95 (0.90, 1.13)
Q3	0.98 (0.83, 1.16)		0.98 (0.83, 1.16)
Q4 (Low access)	1.10*** (1.01, 1.39)		1.09*** (1.00, 1.36)

[a] Model I adjusted for age, sex, and stage at diagnosis. Model II adjusted for all factors of Model I, race/ethnicity, and the non-spatial access factor. Model III adjusted for all factors of Model II plus potential spatial access to oncologists.

* $p<0.001$; ** $p<0.01$; *** $p<0.05$

A separate analysis of data in urban areas gives results similar to those at the state level, as race/ethnicity and the non-spatial access factor are associated with significant disparities in CRC five-year survival. However, the analysis of non-urban areas revealed influences of the spatial access factor, as patients with the lowest potential spatial access to oncologists (HR: 1.10, 95% CI: 1.01 to 1.39) had significantly higher risk of five-year CRC-specific mortality than those with the highest potential spatial access. This relationship remained stable (HR: 1.09, 95% CI: 1.00 to 1.36) after adjusting for

race/ethnicity and the non-spatial access factor. In addition, the non-spatial access factor had less influence on CRC survival in non-urban areas than in urban areas.

6.3.2 Geographic analysis results

Several regions of Texas appeared to experience significant differences in CRC survival (Figure 6.1). Relevant characteristics of these regions are listed in Table 6.4.

Based on age- and sex-adjusted survival times (Figure 6.1.a), one area with longer-than-expected survival times (Area 1) and two areas with shorter-than-expected survival times (Areas 2 and 3) were detected. Area 1 is located between the cities of Austin and Houston and is characterized by a higher percentage of non-Hispanic whites (54.8%), a relatively advantaged socioeconomic status (-0.53) (note: the non-spatial access factor for multiple census tracts as a whole was calculated as the population-weighted average of these census tracts), and a low rate of late-stage CRC diagnosis (57.4%). The O/E ratio of CRC-specific mortality for this area is 0.89 ($p < 0.001$). Area 2 is located at the center of Dallas County and is characterized by a high proportion of non-Hispanic blacks (64.4%), a relatively disadvantaged socioeconomic status (0.56), and a high rate of late-stage CRC diagnosis (69.0%). CRC patients of this area had an excessive risk of CRC-specific mortality (O/E = 1.53, $p < 0.001$). Area 3 is in Bexar County of central Texas and is characterized by a high proportion of Hispanic residents, a relatively disadvantaged socioeconomic status (0.45), and a high rate of late-stage CRC diagnosis (62.5%). The O/E ratio of CRC mortality is 1.52 ($p < 0.01$) for this area.

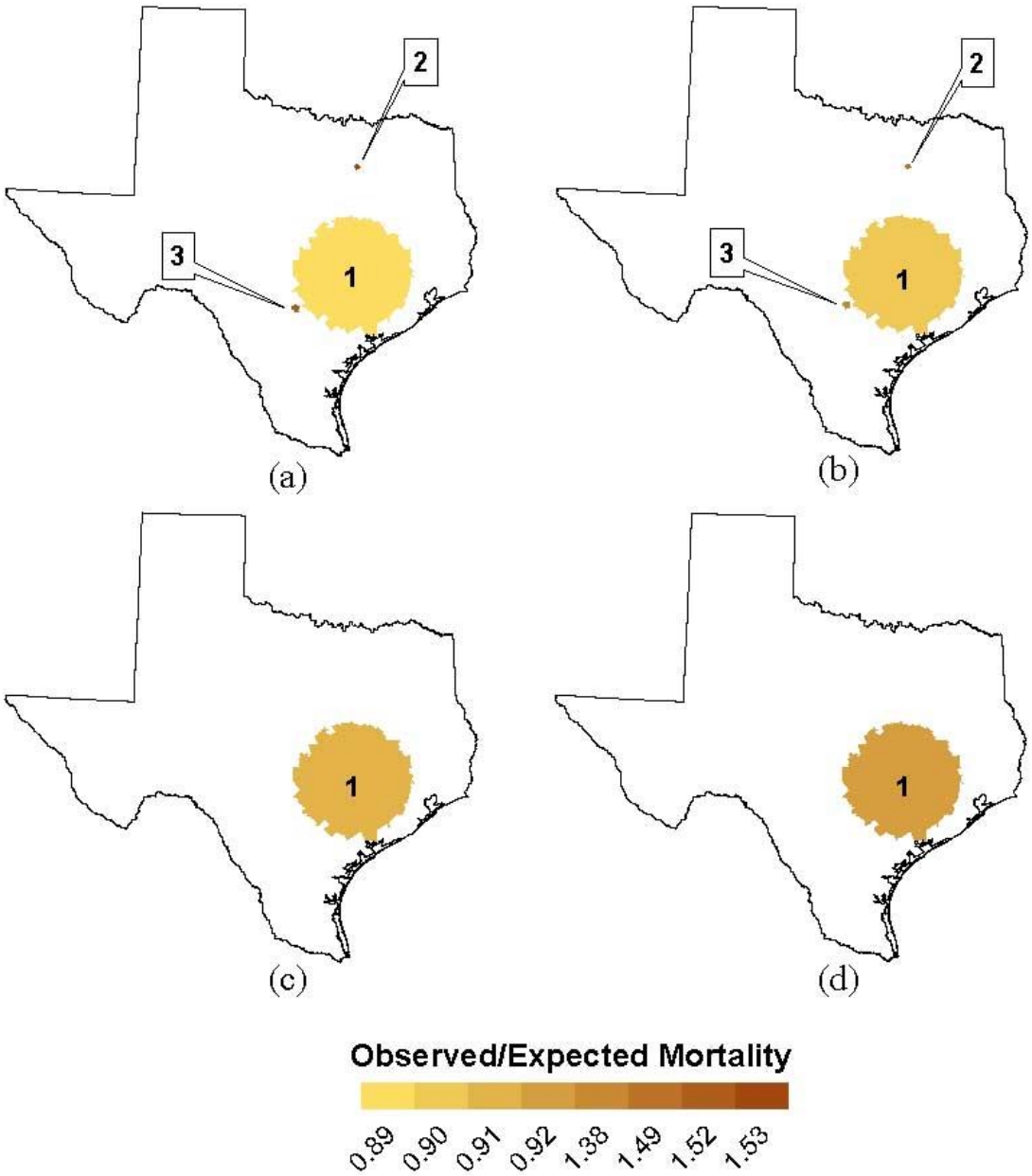

Figure 6.1 Spatial clusters of colorectal cancer survival in Texas (1995-2003) (a). Survival time adjusted for sex and age. (b). Survival time adjusted for sex, age, and stage at diagnosis. (c). Survival time adjusted for sex, age, stage at diagnosis, and race/ethnicity. (d). Survival time adjusted for sex, age, stage at diagnosis, race/ethnicity, and the socioeconomic factor.

Table 6.4 Case characteristics for areas with significantly different risks of colorectal-cancer-specific survival

Survival time adjusted for:	Cluster area	Observed/Expected ratio of death	Percent of non-Hispanic whites	Percent of non-Hispanic blacks	Percent of Hispanics	Socioeconomic factor	Late stage cases (%)
a. Age and sex							
	1	0.89	54.8	12.5	24.2	-0.53	57.4
	2	1.53	4.4	64.4	30.0	0.56	69.0
	3	1.52	10.1	3.9	84.9	0.45	62.5
b. Age, sex, and stage at diagnosis							
	1	0.90	54.8	12.5	24.2	-0.53	57.4
	2	1.49	4.4	64.4	30.0	0.56	69.0
	3	1.38	10.1	3.9	84.9	0.45	62.5
c. Age, sex, stage at diagnosis, and race/ethnicity							
	1	0.91	54.8	12.5	24.2	-0.53	57.4
d. Age, sex, stage at diagnosis, race/ethnicity, and non-spatial access to healthcare services							
	1	0.92	54.8	12.5	24.2	-0.53	57.4

Additional adjustments for diagnosis stage, race/ethnicity, and the socioeconomic factor led to the attenuation of CRC mortality risks in some clusters and the elimination of some clusters. For example, the O/E ratio of Area 1 gradually approaches one (0.89->0.90->0.91->0.92) with the adjustments, indicating that the lower risk of CRC mortality of this area is partly related to a lower rate of late-stage CRC diagnosis, better non-spatial access, and a higher percentage of non-Hispanic whites. Areas 2 and 3 vanished after adjusting for race/ethnicity, suggesting a significantly positive correlation between the percentage of Hispanic and non-Hispanic black residents and CRC-specific mortality risks in these areas. The adjustments did not produce any new apparent clusters. Since there was no significant association between the spatial access factor and CRC-survival across the entire state, adjusting for the spatial access factor was unnecessary.

The geographic analysis described above follows a covariate-adjustment sequence of "age-and-sex->stage at diagnosis->race/ethnicity->socioeconomic factor." Since the sequence of covariate-adjustment might influence the results, an alternative sequence (i.e.,

"age-and-sex->stage at diagnosis-> socioeconomic factor ->race/ethnicity") was also evaluated in the analysis. The evaluation did not suggest any obvious differences between the results corresponding to the two sequences about the locations of spatial clusters and the change of HRs with the covariate adjustment. Consequently, the results of the alternative sequence are not shown here.

6.4 Discussions and Conclusions

This chapter analyzed the influence of multiple factors on CRC survival in Texas. The results indicated that, after adjusting for important covariates, CRC survival was significantly associated with the non-spatial access factor across the state of Texas and with the spatial access factor in non-urban areas of Texas. However, the influences of both spatial and non-spatial access factor were much weaker than that of race/ethnicity.

The geographic analyses suggest that CRC survival in Texas varies by place of residence. The spatial scan statistic method based on covariate-adjusted survival times detected several areas with significantly different CRC survival likelihood. Further adjustments for race/ethnicity and the non-spatial access factor eliminated the areas with significantly shorter CRC survival periods. However, the area with the longest CRC survival still remained unchanged.

CRC survival disparities by race/ethnicity, the non-spatial access factor, and geographic location revealed in this paper are similar to previous findings of other studies undertaken in the United States (Clegg et al. 2002; Du et al. 2007; Henry et al. 2009; Huang et al. 2007b; McDavid et al. 2003; Wudel et al. 2002). For example, Hispanics, blacks, and individuals with limited non-spatial access to healthcare have been systematically experiencing lower CRC survival rates than whites and those from affluent

areas (Clegg et al. 2002; Du et al. 2007). Studies in Los Angeles County, California (Huang et al. 2007b), and New Jersey (Henry et al. 2009) revealed that several areas had significantly shorter survival times after adjusting for important covariates. These areas, like those identified in this study, were characterized by high proportions of black and Hispanic residents and poor non-spatial access to healthcare services.

The different associations between spatial access to oncologists and CRC survival in urban and non-urban settings may reflect the uneven distribution of oncologists of the two settings. As most oncologists were practicing in metropolitan areas, the spatial access of these areas might have reached a 'threshold value' above which the influence of spatial access drastically diminishes. In other words, there were enough oncologists (and corresponding cancer treatment services) within these areas that use of the service was not influenced by the supply-to-demand ratio. However, rural areas were characterized by poor spatial access perhaps below the 'threshold value' and use of the service was heavily influenced by both the supply-to-demand ratio and travel distance. Future studies based on patient survey and/or oncologist-visiting data could be used to verify this finding.

The initial spatial pattern of age-, sex, and stage-adjusted CRC survival might reflect the geographic distributions of social groups who have greater or less chance of surviving from CRC regardless of age, sex, and stage at diagnosis. Blacks, Hispanics and other patients from areas with poor non-spatial access may possess greater risks of CRC-specific mortality than whites and those who live in affluent areas (Clegg et al. 2002; Du et al. 2007). In this study, the area with significantly longer CRC survival was found to be neighborhoods (i.e., the Austin-Houston connecting area) characterized by higher than average percentages of non-Hispanic white, wealthy residents, whereas areas with worse

CRC survival were predominantly Hispanic or non-Hispanic black neighborhoods with low non-spatial access indices (i.e., Areas 2 and 3 in Figure 3). This result simply reinforces the aforementioned discussion of the influence of demographic and non-spatial access factors on CRC survival.

It is unclear why CRC patients in the relatively rural region separating Austin and Houston had longer-than-expected survival times after factoring in race/ethnicity and the non-spatial access factor. One possible reason may be this area's closer proximity to cancer treatment services in both urban areas. As shown in the geocoded distribution, 114 out of the 204 oncologists in Texas were concentrated in or near this area. The University of Texas MD Anderson Cancer Center, a world-class center for cancer treatment, is located at the edge of this area. The advanced treatment facilities and techniques of the MD Anderson Cancer Center might greatly influence the CRC survival of area patients. Other factors that might promote CRC survival in this area are health insurance coverage and doctor referral patterns. More investigation is needed to assess how these factors separately or collectively facilitate CRC survival in this area.

The results of this chapter have important practical implications for government intervention strategies. The insignificant association between spatial access to healthcare services and CRC survival in urban areas implies that it is not necessary to increase oncology resources in these areas. Health programs in these areas should be encouraged toward improving on the non-spatial factors that influence survival, aspects such as health insurance coverage and language barriers. On the other hand, the significant association between spatial access to oncologists and CRC survival in rural areas suggests that promotion of spatially accessible cancer treatment services in areas distant

from metropolitan centers might have more significant impacts on improvement of CRC survival in rural areas. In addition, the areas with significantly shorter CRC survival times, as determined by the geographic analysis, might be good places to initiate programs aiming to enhance CRC prevention and treatment services as well as CRC awareness education.

Some limitations of this study deserve mention when describing and interpreting the results. First, the results could have been influenced by the exclusion of cases and data as a result of incomplete information. Although this study found no significant differences in individual and neighborhood characteristics between the excluded cases (n=16,004) and the study cases, the former accounted for about 22% of the total population. This could dampen confidence in the results of this study. Second, the role of health insurance coverage was not investigated due to the lack of availability of data. As suggested by previous studies, the type of health insurance (e.g., commercial insurance, Medicaid, and Medicare) and the deductibles they carry may influence the quality of the treatment delivered and, consequently, the survival time of the patient (Du et al. 2007; McDavid et al. 2003). Previous studies that considered individual level health insurance information were primarily based on the national samples (e.g., the SEER data), Medicaid or Medicare enrollment data, or hospital treatment records. Similar information is not available in the Texas cancer registry datasets. Third, this chapter did not examine the all-cause survival of CRC, which means other competing risks of mortality were ignored in the analysis. This may further limit the usefulness of the results. In addition, SaTScan uses circles to define cluster shapes. Real clusters may not be circular in size,

and the circular shapes of the clusters presented in Figure 6.1 should not give the false impression that the clusters are circular.

Despite these limitations, this research is the first US study to examine the independent roles of spatial and non-spatial accesses to medical services on CRC survival. Results regarding disparities in terms of race/ethnicity, geographic location, and access to healthcare services can help us better understand how these factors would determine CRC survival. These results can benefit future cancer-disparity elimination programs

CHAPTER 7

CONCLUSION

7.1 Results and Discussions

The major objective of this research is to investigate disparities of CRC survival in Texas from race/ethnicity, SES, geographic location, and potential spatial access to medical services. It was hypothesized that race/ethnicity, area SES, geographic location, and potential spatial access to medical services were significantly related to CRC diagnosis and CRC-specific survival. It was also hypothesized that race/ethnicity and SES were the major factors influencing CRC survival disparities in Texas. To verify these hypotheses, this research employs a variety of spatial and non-spatial methods to investigate the disparities.

Secondly, this research analyzes potential spatial access to medical services that might be associated with CRC stage at diagnosis and CRC-specific survival. This purpose leads to two questions: 1) how can we accurately estimate and appropriately present potential spatial access to medical services, and 2) did racial/ethnic, socioeconomic, and geographic groups in Texas have different potential spatial access to CRC-related services? To answer those two questions, a relative spatial access assessment approach was proposed to estimate potential spatial accesses to primary care and cancer treatment services in Texas.

The sensitivity assessment indicated that the proposed approach of spatial access presentation could effectively minimize the uncertainty problem of gravity-based spatial access models. The analyses of potential spatial access to CRC prevention and treatment services revealed that rural residents of Texas had disadvantaged potential spatial access to CRC services compared to residents of other areas. Potential spatial access to CRC services also differed among racial/ethnic and socioeconomic groups but the differences were less obvious than the rural/urban differences.

The regression analyses identified significant disparities of CRC stage at diagnosis and CRC-specific survival in Texas from the aspects of race/ethnicity, SES, and geographic location. Native American, non-Hispanic black, and Hispanic CRC patients experienced higher risks of late stage CRC diagnosis and CRC-specific mortality than non-Hispanic white and Asian patients. Lower SES corresponded to higher risks of unfavorable CRC outcomes. Multivariate analyses indicated that SES and race/ethnicity are the key independent factors influencing disparities of CRC stage at diagnosis and CRC-specific survival. Decreased potential spatial access to PCPs was found to be associated with elevated risk of late stage diagnosis for CRC patients, although the association was much weaker than those between CRC diagnosis and SES or race/ethnicity. Potential spatial access to oncologists was found to be associated with CRC-specific survival in non-urban areas only.

The geographic analyses detected inhomogeneous spatial patterns of CRC stage at diagnosis and CRC survival in Texas. Areas with significantly different risks of late stage CRC diagnosis and CRC survival were identified. Generally, areas with unfavorable CRC outcomes had higher proportions of Hispanics and blacks, lower proportions of

non-Hispanic whites, and lower SES than other areas. For CRC-specific survival, the adjustments for SES and race/ethnicity did not eliminate the area with longer-than-expected survival times, suggesting some other reasons for better CRC-specific survival in this area.

Results of this study may be useful for reducing CRC survival disparities in Texas in several ways. First, these results could serve as a complementary guideline for government intervention programs. Alleviating the high burden of cancer survival for disadvantaged social groups has been identified as a long-term priority by Texas health departments (Texas Cancer Council 2006). Some preliminary strategies and action steps have been proposed to serve this priority (Texas Cancer Council 2006). However, these strategies and plans lack a basis from which health professionals can identify whom, where, and how the intervention programs should be implemented. The comprehensive information about racial/ethnic, socioeconomic, and geographic characteristics of CRC survival obtained in this research could fill this gap.

For example, since race/ethnicity and SES were the independently important factors for CRC stage at diagnosis and CRC five-year survival, the most vulnerable population should be identified as Hispanics and non-Hispanic blacks from low SES areas. In addition, neighborhoods with higher proportions of late stage CRC cases and higher risks of CRC-specific mortality provide appropriate candidate sites for future intervention. Health professionals could strengthen CRC prevention education, promote CRC screening, and ensure appropriate treatment services in these areas. Existing programs such as the Patient Navigation could also be implemented in these areas to help vulnerable population groups to access CRC screening facilities and treatment services

more conveniently (Freund et al. 2008; Jandorf et al. 2005). Medical-cost alleviation programs could target areas with significantly shorter CRC survival times to ensure that patients in those areas could obtain timely and appropriate treatment services.

Second, this research revealed some omission of previous CRC prevention works. Current CRC prevention guidelines recommend that individuals age 50 and older should regularly take CRC screening examinations (ACS 2010). However, this study found that, although Texans under the age of 50 had a smaller CRC incidence rate than the older ones, they had a higher probability of presenting with a late stage CRC, if diagnosed. A further examination suggested that ages at diagnosis for most of these cases were between 40 and 50. The high risk of late stage CRC diagnosis of this age group also made them less likely to attain the five-year survival than other age groups. This finding emphasizes the necessity to enhance CRC screening among younger groups in Texas. Age 40 might be an appropriate starting age for regular CRC screening.

In addition, analysis results about potential spatial access to CRC services could benefit the health resource allocation work of Texas. The research indicated that most Texans had adequate spatial access to CRC-related services. This information suggests that planners could avoid excessive investments in allocating human and hardware resources but rather focus more on addressing barriers introduced by non-spatial factors such as limited health insurance coverage and cultural differences.

7.2 Contributions and Limitations

The primary contribution of this dissertation is that it extended people's understanding about the disparities of CRC survival in Texas. Texas has different socio-demographic and geographic characteristics compared to other states. For example, the

proportion of Hispanics in Texas was 32% in 2000, which is much higher than most of other states (US Census Bureau 2001a). Texas also had the second largest number of foreign-born Hispanics, who were experiencing higher burdens of unfavorable cancer outcomes than other Hispanics (US Census Bureau 2001a; Ward et al. 2004). In addition, there is a 1,000-mile US-Mexico border in Texas which winds through an extremely rural region with predominantly Hispanic residents. These characteristics imply that Texas might have different characteristics of CRC survival disparity than other states or the whole nation. It is necessary to conduct separate analyses on CRC survival disparities for Texas. However, such studies have been rare. This study for the first time comprehensively examined racial/ethnic, socioeconomic, and geographic disparities of CRC diagnosis stage and CRC-specific survival at a fine geographic scale. The results did reveal a distinct scenario of CRC survival disparity of Texas.

The second major contribution of this research is that it showed how GIS and spatial analysis techniques could facilitate and complement traditional statistical methods in health disparity studies. Traditional statistical methods, which were primarily from the field of social epidemiology, could attain great clarity and precision in revealing how socio-demographic factors separately and collectively influence health outcomes. This information is important for detecting the disparity but insufficient for health disparity elimination. Identifying areas with significantly higher burden of health is a first important step for any meaningful intervention. However, social epidemiological methods could not provide such information. The spatial adaptive filtering method and the spatial scan statistics used in this research could effectively remedy this shortcoming. By comparing the health rates inside and outside a scanning window, these methods

could determine which areas had excessive or low risks of the health outcome and whether the differences were significant or not. By sequentially adjusting for important prognostic factors, one could also determine whether the different health risks of these areas are due to sex, age, SES, race/ethnicity, or other factors. The identification of these areas not only collaborates the results of traditional statistical areas but also provides information about appropriate sites for subsequent disparity-elimination work.

In addition, this research comprehensively reviewed current spatial access models and proposed an approach to better express the results of gravity-based spatial access models. It used a series of Gaussian impedance coefficients to assess the model's sensitivity to distance impedance and proposed a relative spatial access approach to present spatial access results. The evaluations demonstrated that spatial access ratio is more stable to the variance of distance impedance than the commonly-used spatial access index. This approach of spatial access estimation, when used along with other socio-demographic indicators, has a great potential in identifying areas and population groups with limited access to medical services.

This study has a number of limitations. First, as discussed in Chapters 5 and 6, the data exclusion could lower the reliability of the results. About 22% of the original CRC cases were excluded due to the absence of important covariates. Although no significant differences in racial/ethnic, socioeconomic, and geographic distributions were observed between the excluded cases and the retained cases, the results of this study still could be slightly influenced. Second, this study did not incorporate health insurance information in the analyses. Compared to poverty status, health insurance coverage might have a more straightforward impact on CRC outcomes (Du et al. 2007). In addition, health insurance

status, when considered along with potential spatial access to CRC medical services, could more accurately delineate people's access to CRC services. However, the lack of individual level or fine scale health insurance data prevented this study from investigating how health insurance influences CRC stage at diagnosis and CRC-specific survival.

Third, this study was restrained by the limitations of current spatial analysis methods. For example, the geographic analysis of late stage CRC diagnosis in Chapter 4 was only adjusted for sex and age. Using indirect standardization to adjust for more factors such as SES, race/ethnicity, and potential spatial access to CRC services might lead to many small-number population groups and, consequently, unreliable estimations. However, a mature method for adjusting multiple covariates for such analyses has not been available. This limitation prevents one from further exploring the true reasons for the geographic disparity of CRC stage at diagnosis. Also unavailable was a method for characterizing the overall spatial pattern of cancer survival. This study only detected the spatial clusters of CRC survival, that is, areas with significantly different survival times. The overall spatial distribution of CRC survival for Texas would be good background information for interpreting these clusters. Some researchers have used cancer mortality-to-incidence ratios (MIR) to illustrate the spatial patterns of cancer survival (Hebert et al. 2009). However, MIR fails to consider cancer censoring information, an indispensable factor for cancer survival analysis. In addition, although this research has pointed out possible influences of area-level socioeconomic indicators and spatial dependence on the results, it did not provide effective solutions to these two problems.

7.3 Future Research

Since this study has identified population groups and geographic areas with significantly divergent risks of late stage CRC diagnosis and CRC-specific survival after adjusting for important prognostic factors, the next focus could be on comparing the differences of medical care in terms of access, quality, and utilization among these population groups and geographic areas. For persons older than 65, such information could be obtained by linking the cancer registry data with the Medicare claim records (Schootman et al. 2009). Other determinants of CRC stage at diagnosis and CRC survival such as diet and physical inactivity also could be examined. Surveying cancer patients and cancer survivors with different socioeconomic and racial/ethnic backgrounds as well as individuals from areas with significantly different CRC risks may reveal some unreported preventive and protective efforts that can positively influence CRC. These effects could be replicated for other populations or areas to reduce CRC disparities.

Another extension of this research should consider the temporal trend of CRC survival disparities. The time span of this research is nine years (1995 to 2003). During this time period, there might be some minor fluctuations of the racial/ethnic, socioeconomic, or geographic characteristics of CRC survival. A thorough investigation of these trends would not only allow one to better understand how CRC disparities develop over time but also provide more precise information for CRC prevention. Some studies have investigated the spatial-temporal disparities of breast cancer and prostate cancer in Texas (Hsu et al. 2004; Hsu et al. 2007). Their methods could also be used for CRC survival disparities.

REFERENCES

Aamodt, G., S. O. Samuelsen, and A. Skrondal. 2006. A simulation study of three methods for detecting disease clusters. *International Journal of Health Geographics*, 5, 15.

Action Plan on Colorectal Cancer for the State of Texas. 2000. The University of Texas M.D. Anderson Cancer Center, Houston, TX.

Aday, L., and R. Andersen. 1974. A framework for the study of access to medical care. *Health Services Research*, 9(3), 208-220.

Albano, J. D., E. Ward, A. Jemal, R. Anderson, V. E. Cokkinides, T. Murray, J. Henley, J. Liff, and M. J. Thun. 2007. Cancer mortality in the united states by education level and race. *Journal of the National Cancer Institute*, 99(18), 1384-1394.

Albert, D. P., and F. B .Butar. 2005. Estimating the de-designation of single-county HPSAs in the United States by counting naturopathic physicians as medical doctors. *Applied Geography*, 25(3), 271–285.

Ali, M., P. Goovaerts, N. Nazia, M. Z. Haq, M. Yunus, and M. Emch. 2006. Application of Poisson kriging to the mapping of cholera and dysentery incidence in an endemic area of Bangladesh. *International Journal of Health Geographics*, 5, 45.

Altekruse, S. F., L. Huang, J. E. Cucinelli, T. S. McNeel, K. M. Wells, and M. N. Oliver. 2010. Spatial patterns of localized-stage prostate cancer incidence among white and black men in the southeastern United States, 1999-2001. *Cancer Epidemiology Biomarkers & Prevention*, 19(6), 1460-1467.

American Cancer Society. 2005. Cancer Facts & Figures 2005. Atlanta: American Cancer Society.

American Cancer Society. 2008. Cancer Facts & Figures 2008. Atlanta: American Cancer Society.

American Cancer Society. 2009. Cancer Facts & Figures 2009. Atlanta: American Cancer Society.

American Cancer Society. 2010. Cancer Facts & Figures 2010. Atlanta: American Cancer Society.

American Cancer Society. Colorectal Cancer Early Detection. http://www.cancer.org/Cancer/ColonandRectumCancer/MoreInformation/ColonandRectumCancerEarlyDetection/colorectal-cancer-early-detection-a-c-s-recommendations Last accessed October 10, 2010

American Cancer Society, High Plains Division, Inc. 2008. Texas Facts & Figures 2008. Austin, Texas: American Cancer Society, High Plains Division.

Anand, S. 2002. The concern for equity in health. *Journal of Epidemiology and Community Health,* 56(7), 485-487.

Ananthakrishnan, A. N., K. G. Schellhase, R. A. Sparapani, P. W. Laud, and J. M. Neuner. 2007. Disparities in colon cancer screening in Medicare population. *Archives of Internal Medicine,* 167(3), 258-264.

Andrulis, D. P. 1998. Access to care is the centerpiece in the elimination of socioeconomic disparities in health. *Annals of Internal Medicine,* 129(5), 412-416.

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

_____. 1988. Spatial Econometrics: Methods and Models. Dordercht: Kluwer Academic Publishers.

Anselin, L., and A. Bera, 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. In Ullah, A. and Giles, D. E., editors, *Handbook of Applied Economic Statistics*, 237–289.

Anselin, L., S. Ibnu, and K.Youngihn. 2006. GeoDa: An Introduction to Spatial Data Analysis. *Geographical Analysis,* 38(1), 5-22.

Apelberg, B. J., T. J. Buckley, and R. H. White. 2005. Socioeconomic and racial disparities in cancer risk from air toxics in Maryland. *Environmental Health Perspectives,* 113(6), 693-699.

Auvinen, A. and S. Karjalainen. 1997. Possible explanations for social class differences in cancer patient survival. In Kogevinas M, Pearce N, Susser M, Boffetta P (eds): Social inequalities and cancer, IARC Scientific Publications No. 138. Lyon: IARC 377-397.

Aylin, P., R. Maheswaran, J. Wakefield, S. Cockings, L. Jarup, R. Arnold, G. Wheeler, and P. Elliott. 1999. A national facility for small area disease mapping and rapid initial assessment of apparent disease clusters around a point source: The UK small area health statistics unit. *Journal of Public Health Medicine,* 21(3), 289-298.

Bach, P. B., D. Schrag, O. W. Brawley, A. Galaznik, S. Yakren, and C. B. Begg. 2002. Survival of blacks and whites after a cancer diagnosis. *JAMA-Journal of the American Medical Association,* 287(16), 2106-2113.

Baldwin, L. M., S. A. Dobie, K. Billingsley, Y. Cai, G. E. Wright, J. A. Dominitz, W. Barlow, J. L. Warren, and S. H. Taplin. 2005. Explaining black-white differences in receipt of recommended colon cancer treatment. *Journal of the National Cancer Institute,* 97(16), 1211-1220.

Banerjee, S., B. P. Carlin, and A. E. Gelfand. 2004. Hierarchical modeling and analysis for spatial data. Boca Raton, FL: Chapman & Hall/CRC.

Bazargan, M., S. H. Bazargan, M. Farooq, and R. S. Baker. 2004. Correlates of cervical cancer screening among underserved Hispanic and African-American women. *Preventive Medicine,* 39(3), 465-473.

Besag, J. and J. Newell. 1991. The detection of clusters in rare diseases. *Journal of the Royal Statistical Society Series a-Statistics in Society,* 154(1), 143-155.

Bigby, J. A. and M. D. Holmes. 2005. Disparities across the breast cancer continuum. *Cancer Causes & Control,* 16(1), 35-44.

Bigby, J., L. K. Ko, N. Johnson, M. M. A. David, B. Ferrer, and Reach. 2003. A community approach to addressing excess breast and cervical cancer mortality among women of african descent in boston. *Public Health Reports,* 118(4), 338-347.

Black, R. J., L. Sharp, and J. D. Urquhart. 1996. Analysing the spatial distribution of disease using a method of constructing geographical areas of approximately equal population size. In Methods for investigating localized clustering of disease, ed. F. E. Alexander and P. Boyle, 28–39. Lyon, France: IARC Scientific Publications.

Boscoe, F. P. and C. Sherman. 2006. On socioeconomic gradients in cancer registry data quality. *Journal of Epidemiology and Community Health,* 60(6), 551-551.

Borruso, G. 2005. Network density estimation: analysis of point patterns over a network. In Gervasi et al. (Eds.): *Computational Science and Its Applications* - ICCSA 2005, International Conference, Singapore, May 9-12, 2005, Proceedings, Part III.

Bradley, C. J., C. W. Given, and C. Roberts. 2001. Disparities in Cancer Diagnosis and Survival. *Cancer,* 91(1), 178-188.

_____. 2002. Race, Socioeconomic Status, and Breast Cancer Treatment and Survival. *Journal of the National Cancer Institute,* 94(7), 490-496.

Brawley, O. W., and H. P. Freeman. 1999. Race and outcome: is this the end of the beginning for minority health research? *Journal of the National Cancer Institute*, 91(22), 1433-1440.

Brawley, O. W. 2002. Disaggregating the Effects of Race and Poverty on Breast Cancer Outcomes. *Journal of the National Cancer Institute,* 94(7), 471-473.

Braveman, P. 2006. Health Disparities and Health Equity: Concepts and Measurement. *Annual Review of Public Health*, 27, 167-194.

Brown, Z. 1999. The unequal burden of cancer: an assessment of NIH research and programs for ethnic minorities and the medically underserved. Haynes MA, Smith BL, editors. Washington (DC): National Academy Press. P166.

Brual, J., S. Gravely-Witte, N. Suskin, D. E. Stewart, A. Macpherson, and S. L. Grace. 2010. Drive time to cardiac rehabilitation: at what point does it affect utilization? *International Journal of Health Geographics*, 9, 27.

Byers, T., J. Mouchawar., J. Marks., B. Cady., N. Lins., G. M. Swanson., et al. 1999. The American Cancer Society challenge goals. How far can cancer rates decline in the US by the year 2015? *Cancer*, 86(4), 715-727.

Cai, Q. 2007. DMAP IV User Manual.

Center for Disease Control and Prevention. 2010. http://www.cdc.gov/cancer/colorectal/basic_info/screening last accessed on May 18 2010

Cerfolio, R. J., A. S. Bryant, E. Scott, M. Sharma, F. Robert, S. A. Spencer, and R. I. Garver. 2006. Women with pathologic stage i, ii, and iii non-small cell lung cancer have better survival than men. *Chest,* 130(6),1796-1802.

Cervigni, F., Y.Suzuki, T. Ishii, and A. Hata. 2008. Spatial accessibility to pediatric services. *Journal of Community Health,* 33(6), 444–448.

Chan, L., L. G. Hart, and D. C. Goodman. 2006. Geographic access to health care for rural Medicare beneficiaries. *Journal of Rural Health,* 22, 140-146.

Chandra, A. and J. Skinner. 2003. Geography and racial health disparities. Working papers of National Bureau of Economic Research, 9153.

Chu, K. C., C. A. Lamar, and H. P. Freeman. 2003. Racial disparities in breast carcinoma survival rates - separating factors that affect diagnosis from factors that affect treatment. *Cancer,* 97(11), 2853-2860.

Chu, K. C., B. A. Miller, and S. A. Springfield. 2007. Measures of racial/ethnic health disparities in cancer mortality rates and the influence of socioeconomic status. *Journal of the National Medical Association*, 99(10), 1092-1100.

Ciccone, G., C. Prastaro, C. Ivaldi, R. Giacometti, P. Vineis, et al. 2000. Access to hospital care, clinical stage and survival from colorectal cancer according to socioeconomic status. *Annals of Oncology*, 11(9), 1201-1204.

Clayton, D. and J. Kaldor. 1987. Empirical bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, 43(3), 671-681.

Clegg, L. X., F. P. Li, B. G. Hankey, K. Chu, and B. K. Edwards. 2002. Cancer survival among US whites and minorities - a SEER (surveillance, epidemiology, and end results) program population-based study. *Archives of Internal Medicine*, 162(17), 1985-1993.

Cooper, R. A. 1994. Seeking a balanced physician work-force for the 21st century. *Journal of the American Medical Association*, 272(9), 680- 687.

Cockerham, W. C. 1998. Medical sociology. 7th ed. Upper Saddle River, NJ: Prentice Hall.

Cooper, L. A., M. N. Hill, and N. R. Powe. 2002. Designing and evaluating interventions to eliminate racial and ethnic disparities in health care. *Journal of General Internal Medicine*, 17(6), 477-486.

Correa, P. 2003. Helicobacter pylori infection and gastric cancer. *Cancer Epidemiology Biomarkers Preview*, 12, 238S-241S.

Coughlin, S. S., T. D. Thompson, H. I. Hall, P. Logan, and R. J. Uhler. 2002. Breast and Cervical Carcinoma Screening Practices among Women in Rural and Non-rural Areas of the United States, 1998-1999. *Cancer,* 94(11), 2801-2812.

Cummings, D. M., L. M. Whetstone, J. A. Earp, and L. Mayne. 2002. Disparities in mammography screening in rural areas: Analysis of county differences in North Carolina. *Journal of Rural Health,* 18(1), 77-83.

Cutchin, M. P. 2007. The need for the "new health geography" in epidemiologic studies of environmental and health. *Health & Place,* 13(3), 725-742.

Cuzick, J. and R. Edwards. 1990. Spatial clustering for inhomogeneous populations. *Journal of the Royal Statistical Society Series B-Methodological,* 52(1), 73-104.

Dejardin, O., A. M. Bouvier, J. Faivre, S. Boutreux, G. De Pouvourville, and G. Launoy. 2008. Access to care, socioeconomic deprivation and colon cancer survival. *Alimentary Pharmacology & Therapeutics,* 27(10), 940-949.

Dejardin, O., A. M. Bouvier, C. Herbert, M. Velten, A. Buemi, P. Delafosse, N. Maarouf, S. Boutreux, and G. Launoy. 2005. Social and geographic disparities in access to reference care site for patients with colorectal cancer in France. *British Journal of Cancer,* 92(10), 1842-1845.

Dejardin, O., L. Remontet, A. M. Bouvier, A. Danzon, B. Tretarre, P. Delafosse, F. Molinie, N. Maarouf, M. Velten, E. A. Sauleau, N. Bourdon-Raverdy, P. Grosclaude, S. Boutreux, G. De Pouvourville, and G. Launoy. 2006. Socioeconomic and geographic determinants of survival of patients with digestive cancer in France. *British Journal of Cancer,* 95(7), 944-949.

Dignam, J. J. 2000. Differences in breast cancer prognosis among African-American and Caucasian women. *CA-A Cancer Journal for Clinicians,* 50(1), 50-64.

Du, X. L., S. Fang, S. W. Vernon, H. El-Serag, Y. T. Shih, J. Davila, and M. L. Rasmus. 2007. Racial disparities and socioeconomic status in association with survival in a large population-based cohort of elderly patients with colon cancer. *Cancer,* 110 (3), 660-669.

Eley, J. W., H. A. Hill, V. W. Chen, D. F. Austin, M. N. Wesley, et al. 1994. Racial Differences in Survival from Breast-Cancer - Results of the National-Cancer-Institute Black/White Cancer Survival Study. *JAMA-Journal of the American Medical Association,* 272(12), 947-954.

Elhefni, H. 2006. Black-white disparity in colorectal cancer: Is the gap narrowing? *American Journal of Epidemiology,* 163(11), S139-S139.

El-Serag, H. B. 2002. Hepatocellular carcinoma - an epidemiologic view. S72-S78. Original edition, *Journal of Clinical Gastroenterology*.

Environmental Systems Research Institute. ESRI, ArcGIS, V. 9.3, Redlands, CA.

Espey, D. K., X. C. Wu, J. Swan, C. Wiggins, M. A. Jim, et al. 2007. Annual report to the nation on the status of cancer, 1975-2004, featuring cancer in American Indians and Alaska natives. *Cancer,* 110(10), 2119-2152.

Field, T. S., S. M. Buist, C. Doubeni, S. Enger, H. Fouayzi, et al. 2005. Disparities and survival among breast cancer patients. *Journal of the National Cancer Institute Monographs,* 2005(35), 88-95.

Fiscella, K., P. Franks, M. P. Doescher, and B. G. Saver. 2002. Disparities in health care by race, ethnicity, and language among the insured - findings from a national sample. *Medical Care,* 40(1), 52-59.

Fotheringham, A. S., and D. W. S. Wong. 1991. The modifiable areal unit problem in multivariate statistical analysis. *Environmental and Planning A*, 23(7), 1025–44.

Fotheringham, A. S., and F. B. Zhan. 1996. A comparison of three exploratory methods for cluster detection in spatial point patterns. *Geographical Analysis*, 28(3), 200–218.

Fotheringham, A. S., C. Brunsdon, and M. Charlton. 2000. Quantitative Geography: Perspectives on Spatial Data Analysis (Sage, London).

Freeman, H. P. 1998. The meaning of race in science - considerations for cancer research - concerns of special populations in the national cancer program. *Cancer*, 82 (1), 219-225.

Freund, K. M., T. A. Battaglia, E. Calhoun, D. J. Dudley, K. Fiscella, E. Paskett, P. C. Raich, R. G. Roetzheim, and Grp Patient Navigation Res Program. 2008. National cancer institute patient navigation research program methods, protocol, and measures. *Cancer*, 113(12), 3391-3399.

Gilewski, T., S. Adluri, G. Ragupathi, S. L. Zhang, T. J. Yao, K. Panageas, M. Moynahan, A. Houghton, L. Norton, and P. O. Livingston. 2000. Vaccination of high-risk breast cancer patients with mucin-1 (muc1) keyhole limpet hemocyanin conjugate plus qs-21. *Clinical Cancer Research*, 6(5),1693-1701.

Gilligan, T. 2005. Social disparities and prostate cancer: Mapping the gaps in our knowledge. *Cancer Causes & Control*, 16(1), 45-53.

Goel, M. S., C. C. Wee, E. P. McCarthy, R. B. Davis, Q. Ngo-Metzger, and R. S. Phillips. 2003. Racial and ethnic disparities in cancer screening - the importance of foreign birth as a barrier to care. *Journal of General Internal Medicine*, 18(12), 1028-1035.

Goldman, D. A. and J. D. Brender. 2000. Are standardized mortality ratios valid for public health data analysis? *Statistics in Medicine*, 19(8), 1081-1088.

Gomez, S. L., C. D. O'Malley, A. Stroup, S. J. Shema, and W. A. Satariano. 2007. Longitudinal, population-based study of racial/ethnic differences in colorectal cancer survival: Impact of neighborhood socioeconomic status, treatment and comorbidity. *BMC Cancer*, 7, 193.

Goovaerts, P. 2005. Geostatistical analysis of disease data: estimation of cancer mortality risk from empirical frequencies using Poisson Kriging. *International Journal of Health Geographics*, 4, 31.

Goovaerts, P. 2006. Geostatistical analysis of disease data: visualization and propagation of spatial uncertainty in cancer mortality risk using Poisson Kriging and p-field simulation. *International Journal of Health Geographics*, 5, 7.

Govindarajan, R., R. V. Shah, L. G. Erkman, and L. F. Hutchins. 2003. Racial differences in the outcome of patients with colorectal carcinoma. *Cancer,* 97(2), 493-498.

Goy, J. M. Rosenberg, and W. King. 2007. Health Risk Behaviors: Examining Social Inequalities in Bladder and Colorectal Cancers. *Annals of Epidemiology,* 18(2), 156-162.

Grann, V., A. B. Troxel, N. Zojwalla, D. Hershman, S. A. Glied, and J. S. Jacobson. 2005. Regional and racial disparities in breast cancer-specific mortality. *Social Science & Medicine,* 62(2), 337-347.

Guagliardo, M. F. 2004. Spatial accessibility of primary care: concept, methods and challenges. *International Journal of Health Geographics*, 3, 3.

Haas, J. S., C. C. Earle, J. E. Orav, P. Brawarsky, M. Keohane, B. A. Neville, and D. R. Williams. 2008. Racial Segregation and Disparities in Breast Cancer Care and Mortality. *Cancer*, 113(8), 2166-2172.

Halpern, M. T., A. L. Pavluck, C. Y. Ko, and E. M. Ward. 2009. Factors associated with colon cancer stage at diagnosis. *Digestive Diseases and Sciences* 54(12), 2680-2693.

Harper, S. and J. Lynch. 2006. Methods for measuring cancer disparities: a review using data relevant to Healthy People 2010 cancer-related objectives. Washington, DC: National Cancer Institute, 2006. (http://seer.cancer.gov/publications/disparities/).

Harper, S., J. Lynch, S. C. Meersman, N. Breen, W. W. Davis, and M. E. Reichman. 2008. An overview of methods for monitoring social disparities in cancer with an example using trends in lung cancer incidence by area-socioeconomic position and race-ethnicity, 1992-2004. *American Journal of Epidemiology*, 167(8), 889-899.

Hart, G. 2006. Rural-urban commuting areas codes, version 2.0. http://depts.washington.edu/uwruca/ and http://www.doh.wa.gov/data/Guidelines/RuralUrban1.htm Last accessed18 May 2010

Hassan, M. O., Z. Arthurs, V. Y. Sohn, and S. R. Steele. 2009. Race does not impact colorectal cancer treatment or outcomes with equal access. *American Journal of Surgery*, 197(4), 485-490.

Hastie, T. and R. Tibshirani. 1987. Generalized additive-models - some applications. *Journal of the American Statistical Association*, 82(398), 371-386.

Hawley, S. T., L. Foxhall, S. W. Vernon, B. Levin, and J. E. Young. 2001. Colorectal cancer screening by primary care physicians in Texas: A rural-urban comparison. *Journal of Cancer Education*, 16(4), 199-204.

Haynes, M. A., B. D. Smedley, editors. 1999. The unequal burden of cancer: an assessment of NIH research and programs for ethnic minorities and the medically underserved. Institute of Medicine, Committee on Cancer Research Among Minorities and the Medically Undeserved: Washington DC: Academy Press.

Hebert, J. R., G. D. Virginie, M. H. Deborah, et al. 2009. Mapping cancer mortality-to-incidence ratios to illustrate racial and sex disparities in a high-risk population. *Cancer*, 115(11), 2539-2552.

Henry, K. A., X. L. Niu, and F. P. Boscoe. 2009a. Geographic disparities in colorectal cancer survival. *International Journal of Health Geographics*, 8, 13.

Henry, K. A., R. Sherman, and L. M. Roche. 2009b. Colorectal cancer stage at diagnosis and area socioeconomic characteristics in New Jersey. *Health & Place*, 15(2), 505-513.

Higgs, G. 2004. A literature review of the use of GIS-based measures of access to health care services. *Health Services & Outcome Research Methodology*, 5(2), 119-139.

Horner, M. J., L. A. G. Ries, M. Krapcho, N. Neyman, R. Aminou, N. Howlader, S. F. Altekruse, E. J. Feuer, L. Huang, A. Mariotto, B.A. Miller, D. R. Lewis, M. P. Eisner, D. G. Stinchcomb, and B. K. Edwards (eds). 2009. SEER Cancer Statistics Review, 1975-2006. National Cancer Institute. Bethesda, MD.

Hsu, C. E, H. Jacobson, and M. F. Soto. 2004. Evaluating the disparity of female breast cancer mortality among racial/ethnic groups—a spatiotemporal analysis. *International Journal of Health Geographics*, 3, 4.

Hsu, C. E., F. S. Mas, J. M. Hickey, J. A. Miller, and D. J. Lai. 2006. Surveillance of the colorectal cancer disparities among demographic subgroups: A spatial analysis. *Southern Medical Journal*, 99(9), 949-956.

Hsu, C. E., F. S. Mas, J. A. Miller, and E. T. Nkhoma. 2007. A spatial-temporal approach to surveillance of prostate cancer disparities in population subgroups. *Journal of the National Medical Association*, 99(1), 72-80.

Huang, B., M. Dignan, D. Han, and O. Johnson. 2009. Does distance matter? Distance to Mammography facilities and stage at diagnosis of breast cancer in Kentucky. *Journal of Rural Health*, 25(4), 366-371.

Huang, L., L. W. Pickle, D. Stinchcomb, and E. J. Feuer. 2007. Detection of spatial clusters - application to cancer survival as a continuous outcome. *Epidemiology*, 18(1), 73-87.

Hwang, H. L., and J. Rollow. 2000. Data processing procedures and methodology for estimating trip distances for the 1995 American Travel Survey (ATS). ORNL/TM-2000/141, Oak Ridge National Laboratory, Oak Ridge, TN.

International Science and Engineering Fair. 2006. ISEF rules: human subjects. Science Service. Available from http://www.societyforscience.org/Page.aspx?pid=317#Rules. Accessed May 30, 2010

Jacquez, G. M., and L. A. Waller. 1999. The effect of uncertain locations on disease cluster statistics. In Quantifying spatial uncertainty in natural resources: Theory and applications for GIS and remote sensing, ed. H. T. Mowrer and R.G. Congalton, 53–64. Chelsea, MI: Sleeping Bear Press.

Jacquez, G. M., and D. A. Greiling. 2003. Local clustering in breast, lung and colorectal cancer in Long Island, New York. *International Journal of Health Geographic*, 2, 1.

Jemal, A., M. Kulldorff, S. S. Devesa, R. B. Hayes, and J. F. Fraumeni. 2002. A geographic analysis of prostate cancer mortality in the US 1970-89. *International Journal of Cancer*, 101(2), 169-174.

James, T. M., K. A. Greiner, E. F. Ellerbeck, C. Y. Feng, and J. S. Ahluwalia. 2006. Disparities in colorectal cancer screening: A guideline-based analysis of adherence. *Ethnicity & Disease*, 16(1), 228-233.

Jandorf, L., Y. Gutierrez, J. Lopez, J. Christie, and S. H. Itzkowitz. 2005. Use of a patient navigator to increase colorectal cancer screening in an urban neighborhood health clinic. *Journal of Urban Health-Bulletin of the New York Academy of Medicine*, 82(2), 216-224.

Jones, A. P., R. Haynes, V. Sauerzapf, S. M. Crawford, H. Zhao, and D. Forman. 2008. Travel times to health care and survival from cancers in northern England. *European Journal of Cancer*, 44(2), 269-274.

Joseph, A. E., and D. R. Phillips. 1984. Accessibility and utilization: Geographical perspectives on health care delivery. Harper & Row, London, UK.

Joseph, A. E., and P. R. Bantock. 1982. Measuring potential physical accessibility to general practitioners in rural areas: A method and case study. *Social Science & Medicine*, 16(1), 85–90.

Kafadar, K. 1996. Smoothing geographical data, particularly rates of disease. *Statistics in Medicine*, 15(23), 2539-2560.

Kanna, B., T. K. Narang, T. Atwal, D. Paul, and S. Azeez. 2009. Ethnic disparity in mortality after diagnosis of colorectal cancer among inner city minority new yorkers. *Cancer*, 115(23), 5550-5555.

Kaplan, E. L. and P. Meier. 1958. Nonparametric estimation from incomplete observations. *Journal of American Statistical Association*, 53(282), 457-481.

Kirsner, R. S., F. C. Ma, L. Fleming, D. G. Federman, E. Trapido, R. Duncan, and J. D. Wilkinson. 2006. The effect of Medicare health care delivery systems on survival for patients with breast and colorectal cancer. *Cancer Epidemiology Biomarkers & Prevention*, 15(4), 769-773.

Klassen, A. C., F. Culticlo, N. Kulldorff, A. J. Alberg, E. A. Platz, and S. T. Neloms. 2006. Missing stage and grade in Maryland prostate cancer surveillance data, 1992-1997. *American Journal of Preventive Medicine*, 30(2), S77-S87.

Krieger, N. 2005. Defining and investigating social disparities in cancer: Critical issues? *Cancer Causes & Control*, 16(1), 5-14.

Krieger, N., J. T. Chen, P. D. Waterman, D. H. Rehkopf, and S. V. Subramanian. 2003. Race/ethnicity, gender, and monitoring socioeconomic gradients in health: A comparison of area-based socioeconomic measures - the public health disparities geocoding project. *American Journal of Public Health*, 93(10), 1655-1671.

Krieger, N., J. T. Chen, P. D. Waterman, D. H. Rehkopf, R. H. Yin, and B. A. Coull. 2006. Race/ethnicity and changing US socioeconomic gradients in breast cancer incidence: California and Massachusetts, 1978-2002 (United States). *Cancer Causes & Control*, 17(2), 217-226.

Krieger, N., C. Quesenberry, T. Peng, P. Horn-Ross, S. Stewart, S. Brown, K. Swallen, T. Guillermo, D. Suh, L. Alvarez-Martinez, and F. Ward. 1999. Social class, race/ethnicity, and incidence of breast, cervix, colon, lung, and prostate cancer among Asian, black, Hispanic, and white residents of the San Francisco bay area, 1988-92 (United States). *Cancer Causes & Control*, 10(6), 525-537.

Krieger, N., D. R. Williams, and N. E. Moss. 1997. Measuring social class in US public health research: Concepts, methodologies, and guidelines. *Annual Review of Public Health*, 18, 341-378.

Kulldorff, M. 1997. A spatial scan statistic. *Communications in Statistics-Theory and Methods*, 26(6), 1481-1496.

Kulldorff, M. and N. Nagarwalla. 1995. Spatial disease clusters - detection and inference. *Statistics in Medicine*, 14(8), 799-810.

Kulldorff, M., and Information Management Services, Inc. 2009. SaTScanTM v8.0: Software for the spatial and space-time scan statistics. http://www.satscan.org/.

Kwan, M. P. 1998. Space-time and integral measures of individual accessibility: A comparative analysis using a point-based framework. *Geographical Analysis* 30, 191-216.

Lai, S. M., K. B. Zhang, R. J. Uhler, J. N. Harrison, G. G. Clutter, and M. A. Williams. 2006. Geographic variation in the incidence of colorectal cancer in the United States, 1998-2001. *Cancer*, 107(5), 1172-1180.

Langford, M., and G.Higgs. 2006. Measuring potential access to primary healthcare services: the influence of alternative spatial representations of population. *The Professional Geographer*, 58(3), 294–306.

Lawson, A. B. 2001. Statistical methods in spatial epidemiology. New York: Wiley.

Lee, P. R. 1995. Health system reform and the generalist physician. *Academic Medicine* 70: S10-S13.

Lee-Feldstein, A., P. J. Feldstein, T. Buchmueller, and G. Katterhagen. 2000. The relationship of HMOS, health insurance, and delivery systems to breast cancer outcomes. *Medical Care*, 38(7), 705-718.

Lin, G. 2003. A spatial logit association model for cluster detection. *Geographical Analysis*, 35(4), 329-340.

Lisovicz, N., R. E. Johnson, J. Higginbotham, J. A. Downey, C. M. Hardy, M. N. Fouad, A. W. Hinton, and E. E. Partridge. 2006. The deep south network for cancer control - building a community infrastructure to reduce cancer health disparities. *Cancer*, 107(8), 1971-1979.

Liu, L. H., D. Deapen, and L. Bernstein. 1998. Socioeconomic status and cancers of the female breast and reproductive organs: A comparison across racial/ethnic populations in Los Angeles county, California (United States). *Cancer Causes & Control*, 9(4), 369-380.

Loerzel, V. W., and A. Busby. 2005. Interventions that address cancer health disparities in women. *Family & Community Health*, 28(1), 79-89.

Luo, J. C., C. H. Zhou, Y. Leung, J. S. Zhang, and Y. F. Huang. 2002. Scale-space theory based regionalization for spatial cells. *Acta Geogr. Sin*, 57, 167-173 (in Chinese).

Luo, W. 2004. Using a GIS-based floating catchment method to assess areas with shortage of physicians. *Health & Place*, 10(1), 1–11.

Luo, W., and F.Wang. 2003a. Spatial accessibility to primary care and physician shortage area designation: a case study in Illinois with GIS approaches. In: Skinner, R., Khan, O. (Eds.), Geographic Information Systems and Health Applications. Idea Group Publishing, Hershey, PA, 260–278.

Luo,W., and F. Wang. 2003b. Measures of spatial accessibility to health care in a GIS environment: Synthesis and a case study in the Chicago region. *Environment and Planning B: Planning and Design*, 30(6), 865–884.

Luo, W., and Y. Qi. 2009. An enhanced two-step floating catchment area (E2SFCA) method for measuring spatial accessibility to primary care physicians. *Health & Place*, 15(4), 1100-1107.

Ma, G. X., Y. Tan, J. I. Toubbeh, R. L. Edwards, S. E. Shive, P. Siu, P. Wang, and C. Y. Fang. 2006. Asian tobacco education and cancer awareness research special population network - a model for reducing Asian American cancer health disparities. *Cancer*, 107(8), 1995-2005.

Mackenbach, J. P., and A. E. Kunst. 1997. Measuring the magnitude of socioeconomic inequalities in health: an overview of available measures illustrated with two examples from Europe. *Social Science & Medicine*, 44(6), 757–771.

Mayberry, R. M., R. J. Coates, H. A. Hill, L. A. Click, V. W. Chen, D. F. Austin, C. K. Redmond, C. M. Fenogliopreiser, C. P. Hunter, M. A. Haynes, H. B. Muss, M. N. Wesley, R. S. Greenberg, and B. K. Edwards. 1995. Determinants of black-white differences in colon-cancer survival. *Journal of the National Cancer Institute*, 87(22), 1686-1693.

Marshall, R. J. 1991. Mapping disease and mortality-rates using empirical bayes estimators. *Applied Statistics-Journal of the Royal Statistical Society Series* C, 40(2), 283-294.

Mathworks, Inc. MATLAB,v.7, Natick, MA.

Matthews, B. A., R. C. Anderson, and A. B. Nattinger. 2005. Colorectal cancer screening behavior and health insurance status (United States). *Cancer Causes & Control*, 16(6), 735-742.

McDavid, K., T. C. Tucker, A. Sloggett, and M. P. Coleman. 2003. Cancer survival in Kentucky and health insurance coverage. *Archives of Internal Medicine*, 163(18), 2135-2144.

McGrail, M. R., and J. S. Humphreys. 2009. Measuring spatial accessibility to primary care in rural areas: improving the effectiveness of the two-step floating catchment area method. *Applied Geography*, 29(4), 533-541.

Meliker, J. R., G. M. Jacquez, P. Goovaerts, G. Copeland, and M. Yassine. 2009. Spatial cluster analysis of early stage breast cancer: A method for public health practice using cancer registry data. *Cancer Causes & Control*, 20(7), 1061-1069.

Mu, L., and F. Wang. 2008. A scale-space clustering method: Mitigating the effect of scale in the analysis of zone-based data. *Annals of the Association of American Geographers*, 98(1), 85–101.

Muirhead, C. R., and B. K. Butland. 1996. Testing for over-dispersion using an adapted form of the Potthoff–Whittinghill method. In Methods for investigating localized clustering of disease, ed. F. E. Alexander and P. Boyle, 40–52. Lyon, France: IARC Scientific Publications.

National Cancer Institute (NCI). 2008. Cancer health disparities. http://www.cancer.gov/cancertopics/types/disparities Accessed on 05/04/2009

National Cancer Institute (NCI). http://crchd.cancer.gov/disparities/defined.html Accessed on 11/30/2009

National Cancer Institute (NCI). Statistical Research & Applications Branch, Measures of Cancer Survival. http://srab.cancer.gov/survival/measures.html Accessed 05/30/2010.

NIH National Center for Minority Health and Heath Disparities. http://crchd.cancer.gov/disparities/defined.html Accessed on 11/30/2009

Nattinger, A. B., R. T. Kneusel, R. G. Hoffmann, and M. A. Gilligan. 2001. Relationship of distance from a radiotherapy facility and initial breast cancer treatment. *Journal of the National Cancer Institute*, 93(17), 1344-1346.

Onega, T., E. J. Duell, X. Shi, D. M. Wang, E. Demidenko, and D. Goodman. 2008. Geographic access to cancer care in the U.S. *Cancer*, 112(4), 909-918.

Onega, T., E. J. Duell, X. Shi, E. Demidenko, D. Gottlieb, and D. C. Goodman. 2009. Influence of NCI cancer center attendance on mortality in lung, breast, colorectal, and prostate cancer patients. *Medical Care Research and Review*, 66(5), 542-560.

Openshaw, S. 1984. Ecological fallacies and the analysis of areal census data. *Environmental and Planning A*, 16(1), 17–31.

Openshaw, S., M. Charlton, A. W. Craft, and J. M. Birch. 1988. Investigation of leukemia clusters by the use of a geographical analysis machine. *Lancet*, 1(8580), 272–273.

Openshaw, S., M. Charlton, C. Wymer, and A. Craft. 1987. Developing a Geographical Analysis Machine for the automated analysis of point data sets. *International Journal of Geographical Information Systems*, 1(4), 335–58.

O'Sullivan, D., and D. J. Unwin. 2003. Geographic information analysis. Hoboken, NJ: Wiley.

Parsons, M. A., and K. D. Askland. 2007. Cancer of the colorectum in Maine, 1995-1998: Determinants of stage at diagnosis in a rural state. *Journal of Rural Health*, 23(1), 25-32.

Pickens, A., and M. B. Orringer. 2003. Geographical distribution and racial disparity in esophageal cancer. In Symposium on Understanding Disparities in Cardiovascular and Thoracic Surgical Outcomes in African Americans, 76:S1367-S1369. San Diego, California.

Pickle, L. W., and A. A. White. 1995. Effects of the choice of age-adjustment method on maps of death rates. *Statistics in Medicine*, 14(4), 615-627.

Read, T. E., and I. J. Kodner. 1999. Colorectal cancer: Risk factors and recommendations for early detection. *American Family Physician*, 59(11), 3083-3092.

Radke, J., and L. Mu. 2000. Spatial decomposition, modelling and mapping service regions to predict access to social programs. *Geographic Information Sciences*, 6(2), 105–112.

Robert, S. A., I. Strombom, A. Trentham-Dietz, J. M. Hampton, J. A. McElroy, P. A. Newcomb, and P. L. Remington. 2004. Socioeconomic risk factors for breast cancer - distinguishing individual- and community-level effects. *Epidemiology*, 15(4), 442-450.

Ricketts, T. C., L. J. Goldsmith, G. M. Holmes, R. Randorph, R. Lee, and D. H. Taylor. 2007. Designating places and Populations as medically underserved: A proposal for a new approach. *Journal of Health Care for the Poor and Underserved*, 18(3), 567-589.

Risser, D. R., M. A. Williams, B. Mokry, P. Betts, and E. A. Miller. 2009. Cancer Incidence & Mortality in Texas, 2001–2005. Austin, Texas: Texas Department of State Health Services.

Roche, L. M., R. Skinner, and R. B. Weinstein. 2002. Use of a Geographic Information System To Identify and Characterize Areas with High Proportions of Distant Stage Breast Cancer. *Journal of Public Health Management and Practice*, 8(2), 26-32.

Roetzheim, R. G., E. C. Gonzalez, J. M. Ferrante, N. Pal, D. J. Van Durme, and J. P. Krischer. 2000. Effects of health insurance and race on breast carcinoma treatments and outcomes. *Cancer*, 89(11), 2202-2213.

Ross, C. E., and J. Mirowsky. 2000. Does medical insurance contribute to socioeconomic differentials in health? *Milbank Quarterly*, 78(2), 291.

Roux, A. V. D. 2000. Multilevel analysis in public health research. *Annual Review of Public Health*, 21(2000), 171-192.

Roux, A. V. D., S. S. Merkin, P. Hannan, D. R. Jacobs, and C. I. Kiefe. 2003. Area characteristics, individual-level socioeconomic indicators, and smoking in young adults - the coronary artery disease risk development in young adults study. *American Journal of Epidemiology*, 157(4), 315-326.

Rushton, G., and P. Lolonis. 1996. Exploratory spatial analysis of birth defect rates in an urban population. *Statistics in Medicine*, 15(7-9), 717–726.

Schleinitz, M. D., D. DePalo, J. Blume, and M. Stein. 2005. Can differences in breast cancer utilities explain disparities in breast cancer care? In 28th Annual Meeting of the Society-of-General-Internal-Medicine, 21:1253-1260. New Orleans, LA.

Schootman, M., D. B. Jeffe, M. Lian, W. E. Gillanders, and R. Aft. 2009. The role of poverty rate and racial distribution in the geographic clustering of breast cancer survival among older women: A geographic and multilevel analysis. *American Journal of Epidemiology*, 169(5), 554-561.

Shavers, V. L., C. F. Lynch, and L. F. Burmeister. 2002. Racial differences in factors that influence the willingness to participate in medical research studies. *Annals of Epidemiology*, 12(4), 248-256.

Shokar, N. K., C. A. Carlson, and S. C. Weller. 2007. Prevalence of colorectal cancer testing and screening in a multiethnic primary care population. *Journal of Community Health*, 32(5), 311-323.

_____. 2008. Factors associated with racial/ethnic differences in colorectal cancer screening. *Journal of the American Board of Family Medicine*, 21(5), 414-426.

Schuurman, N., M. Berube, and V. A. Crooks. 2010. Measuring potential spatial access to primary health care physicians using a modified gravity model. *Canadian Geographer*, 54, 29-45.

Schuurman, N., N. Bell, J. R. Dunn, and L. Oliver. 2007. Deprivation indices, population health and geography: An evaluation of the spatial effectiveness of indices at multiple scales. *Journal of Urban Health-Bulletin of the New York Academy of Medicine*, 84(4), 591-603.

Schwartz, K. L., H. Crossley-May, F. D. Vigneau, K. Brown, and M. Banerjee. 2003. Race, socioeconomic status and stage at diagnosis for five common malignancies. *Cancer Causes & Control*, 14(8), 761-766.

Settimi, L., P. Comba, S. Bosia, C. Ciapini, E. Desideri, et al. 2001. Cancer risk among male farmers: a multi-site case-control study. *International Journal of Occupational Medicine and Environmental Health*, 14(4), 187-192.

Shavers, V. L., and M. L. Brown. 2002. Racial and ethnic disparities in the receipt of cancer treatment. *Journal of the National Cancer Institute*, 94(5), 334-357.

Shi, X. 2009. A Geocomputational Process for Characterizing the Spatial Pattern of Lung Cancer. *Annals of the Association of American Geographers*, 99(3), 521-533.

Silverman, B. W. 1986. Density estimation for statistics and data analysis. New York: Chapman & Hall.

Siminoff, L. A., and L. Ross. 2005. Access and Equity to Cancer Care in the USA: A Review and Assessment. *Postgraduate Medical Journal*, 81(961), 674-679.

Singh, G. K., B. A. Miller, B. F. Hankey, and B. K. Edwards. 2003. Area socioeconomic variations in US cancer incidence, mortality, stage, treatment, and survival 1975-1999. Bethesda, MD: National Cancer Institute. NCI Cancer Surveillance Monograph Series, no. 4 : NIH Publication No. 03-5417.

Singh, G. K., B. A. Miller, B. F. Hankey, E. J. Feuer, and L. W. Pickle. 2002. Changing area socioeconomic patterns in US cancer mortality, 1950-1998: Part I - all cancers among men. *Journal of the National Cancer Institute*, 94(12), 904-915.

Skibber, J. M., B. D. Minsky, and H. M. Paulo. 2001. In: Cancer: Principles and Practice of Oncology, vol. 1, sixth ed. Lippincott Williams & Wilkins, New York.

Sonneveld, D. J. A., M. Schaapveld, D. T. Sleijfer, G. J. T. Meerman, W. T. A. van der Graaf, R. H. Sijmons, H. S. Koops, and H. J. Hoekstra. 1999. Geographic clustering of testicular cancer incidence in the northern part of the Netherlands. *British Journal of Cancer*, 81(7), 1262-1267.

SPSS for Windows, Rel. 16.0. (2008) Chicago: SPSS Inc.

Steenland, K., J. Henley, and M. Thun. 2002. All-cause and cause-specific death rates by educational status for two million people in two American Cancer Society cohorts, 1959-1996. *American Journal of Epidemiology*, 156(1), 11-21.

Szwarcwald, C. L. 2002. On the World Health Organization's measurement of health inequalities. *Journal of Epidemiology and Community Health*, 56, 177-182.

Talbot, T. O., M. Kulldorff, S. P. Forand, and V. B. Haley. 2000. Evaluation of spatial filters to create smoothed maps of health data. *Statistics in Medicine*, 19(17-18), 2399-2408.

Texas Cancer Registry, Statistics, http://www.texascancer.info/statistics.html Last accessed 10/25/2010

Texas Cancer Plan 2005, 4th edition. Texas Cancer Council. 2006.

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

Turnbull, B. W., E. J. Iwano, W. S. Burnett, H. L. Howe, and L. C. Clark. 1989. Monitoring for clusters of disease - application to leukemia incidence in upstate New York. In National Conf on Clustering of Health Events, 132:S136-S143. Atlanta, GA.

Ugnat, A. M., L. Xie, J. Morris, R. Semenciw, and Y. Mao. 2004. Survival of women with breast cancer in Ottawa, Canada: variation with age, stage, histology, grade and treatment. *British Journal of Cancer*, 90(6), 1138-1143.

US Census Bureau, 2000. URL: http://www.census.gov/main/www/cen2000.html Last accessed September 20, 2010

US Census Bureau, 2001a, Census 2000 Summary File 1 (SF1) Texas. US Census Bureau, Washington, D.C.

US Census Bureau, 2001b, Census 2000 Summary File 3 (SF3) Texas. US Census Bureau, Washington, D.C.

US Census Bureau, 2001c, Census 2000 TIGER/Line Files Texas, US Census Bureau, Washington, D.C.

US Department of Health and Human Services. 2000. Healthy people 2010: understanding and improving health. 2nd ed. Washington, D.C.: US Government Printing Office.

US General Accounting Office (GAO). 1995. Health Care Shortage Areas: Designation Not a Useful Tool for Directing Resources to the Underserved. GAO/HEHS-95-2000, General Accounting Office, Washington, D.C.

Wagstaff, A., P. Paci, and E. VanDoorslaer. 1991. On the measurement of inequalities in health. *Social Science and Medicine*, 33(5), 545–557.

Wan, N., F. B. Zhan, and Z. Cai. 2011. Socioeconomic disparity in prostate cancer mortality and the impact of geographic scale. *Southern Medical Journal* (In press).

Wang, L. 2007. Immigration, ethnicity, and accessibility to culturally diverse family physicians. *Health & Place*, 13(3), 656–671.

Wang, F.H., S. McLafferty, V. Escamilla, and L. Luo. 2008. Late-stage breast cancer diagnosis and health care access in Illinois. *Professional Geographer*, 60(1), 54-69.

Wang, F. H., and V. M. O'Brien. 2005. Constructing geographic areas for analysis of homicide in small populations: Testing herding-culture-of-honor proposition. In Geographic information systems and crime analysis, ed. F. Wang, 84–101. Hershey, PA: Idea Group.

Wang, F. H., and W. W. Manor. 2002. Where the jobs are: employment access and crime patterns in Cleveland. *Annals of the Association of American Geographers*, 92(3), 435-450.

Ward, E., A. Jemal, V. Cokkinides, G. K. Singh, C. Cardinez, A. Ghafoor, and M. Thun. 2004. Cancer disparities by race/ethnicity and socioeconomic status. *CA-a Cancer Journal for Clinicians*, 54(2), 78-93.

Wardle, J., A. Miles, and W. Atkin. 2005. Gender differences in utilization of colorectal cancer screening. *Journal of Medical Screening*, 12(1), 20-27.

Washington State Department of Health. 2009. http://depts.washington.edu/uwruca/ Last accessed on May 08, 2010

Weibull, J. W. 1976. An axiomatic approach to the measurement of accessibility. *Regional Science and Urban Economics*, 6(4), 357–379.

Wheeler, D. C. 2007. A comparison of spatial clustering and cluster detection techniques for childhood leukemia incidence in Ohio, 1996-2003. *International Journal of Health Geographics*, 6, 13.

Whitehead, M. 1992. The concepts and principles of equity and health. *International Journal of Health Services*, 22(3), 429-445.

Wing, P., and C. Reynolds. 1988. The availability of physician services: A geographic analysis. *Health Services Research*, 23(5), 649–667.

Woods, L. M., B. C. Rachet, and M. P. Coleman. 2006. Origins of socio-economic inequalities in cancer survival: a review. *Annals of Oncology*, 17(1), 5-19.

Wu, X. C., V. Cokkinides, V. W. Chen, M. Nadel, Y. Ren, J. Martin, and G. L. Ellison. 2006a. Associations of subsite-specific colorectal cancer incidence rates and stage of disease at diagnosis with county-level poverty, by race and sex. *Cancer*, 107(5), 1121-1127.

Wudel, L. J., C. C. William, S. Yu, et al. 2002. Disparate outcomes in patients with colorectal cancer. *Archives of Surgery*, 137(5), 550-556.

Yang, D., R. Goerge, and R. Mullner. 2006. Comparing GIS-based methods of measuring spatial accessibility to health services. *Journal of Medical Systems*, 30(1), 23–32.

Yood, M. U., C. C. Johnson, A. Blount, J. Abrams, E. Wolman, B. D. McCarthy, U. Raju, D. S. Nathanson, M. Worsham, and S. R. Wolman. 1999. Race and differences in breast cancer survival in a managed care population. *Journal of the National Cancer Institute*, 91(17), 1487-1491.

Yost, K., C. Perkins, R. Cohen, C. Morris, and W. Wright. 2001. Socioeconomic status and breast cancer incidence in California for different race/ethnic groups. *Cancer Causes & Control*, 12(8), 703-711.

Young, J. L., S. D. Roffers, L. A. G. Ries, and A. G. Fritz. 2000. SEER summary staging manual-2000: codes and coding instructions. National Cancer Institute, NIH, Bethesda, MD.

Yu, X. Q. 2009. Socioeconomic disparities in breast cancer survival: Relation to stage at diagnosis, treatment and race. *BMC Cancer*, 9, 364.

Zeger, S. L., K. Y. Liang, and P. S. Albert. 1988. Models for longitudinal data - a generalized estimating equation approach. *Biometrics*, 44(4), 1049-1060.

Zhou, X. 2000. Geographic concentrations of lung cancer mortality in Texas and their relationships to environmental and socioeconomic conditions. ETD Collection for Texas State University. Paper AAI3025226.

Zorn, C. J. W. 2001. Generalized estimating equation models for correlated data: A review with applications. *American Journal of Political Science*, 45(2), 470-490.

VITA

Neng Wan was born in Tianmen, Hubei, China on February 29, 1980, to Zhongqing Wan and Mingqing Pan. After graduating from Tianmen High School, he attended Wuhan University, Wuhan, where he received a B.S. degree in Geodesy in 2003, and a M.S. degree in Geographic Information Science in 2006. His main research interests during the master period were image point detection and image-based 3-dimensional object reconstruction. He started his doctoral studies in the Department of Geography at Texas State University-San Marcos in January, 2007. His current research interests include health disparity, spatial cluster analysis, health service accessibility analysis, and human-environment interactions.

Permanent Address: 312 Fujiapo 1 RD. 25

Wuhan, Hubei 430070

CHINA - MAINLAND

This dissertation was typed by Neng Wan.