

**GEOSPATIAL APPROACHES FOR UNDERSTANDING THE ROLE OF
RESIDENTIAL MOBILITY AND AREA-LEVEL FACTORS IN COLON CANCER
SURVIVAL DISPARITIES.**

A Dissertation

Submitted to

the Temple University Graduate Board

In Partial Fulfillment

of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in Geography and Urban Studies

by

Daniel Wiese

May 2021

Examining Committee Members:

Dr. Kevin A. Henry, Advisory Chair, Temple University, Department of Geography and Urban Studies.

Dr. Shannon M. Lynch, Fox Chase Cancer Center, Temple University.

Dr. Victor H. Gutierrez-Velez, Temple University, Department of Geography and Urban Studies.

Dr. Krista Schroeder, External Examiner, Temple University, School of Public Health.

ABSTRACT

A primary reason geospatial approaches are important in cancer research is that health and disease are shaped not only by factors such as age, race/ethnicity, genes, and clinical care but also by the environment where individuals work and act. While the use of geospatial approaches in cancer research is growing, several limitations remain. For example, for most population-based studies, cancer patients' neighborhood environments are based on only a single location derived from the residence at the time of diagnosis. This dissertation aimed to address this limitation by using a unique dataset of colon cancer patients diagnosed in New Jersey that include residential histories obtained through a data linkage with LexisNexis, a commercial data collection company. By incorporating residential histories, I moved beyond a cross-sectional approach to examine how residential histories and socio-spatial mobility can change a patient's geographic context over time and influence survival.

To demonstrate the application of these data in this dissertation, I completed three case studies. In the first case study, I compared whether including residential histories changed the risk of death estimates by neighborhood poverty compared to the traditional approach when including only the location at the time of diagnosis. Results suggested that the risk of death estimates from neighborhood poverty were generally similar in strength and direction regardless of residential histories inclusion. This finding was likely a result of minimal socio-spatial mobility of colon cancer patients (i.e., patients generally moving to census tracts with similar poverty levels).

The second study aimed to compare the geographic risk of death estimates when using single location and residential histories in spatial models. Results overall showed that the

geographic patterns of the risk of death estimates were generally similar between the models. However, not accounting for residential mobility resulted in underestimated geographic risk of death in several areas. This finding was related to the fact that approximately 35% of the colon cancer patients changed the residency, and 12% of the initial study population left New Jersey after the diagnosis.

In the third case study, I examined whether landscape characteristics (e.g., built environment) were associated with the risk of death from colon cancer independent of individual-level factors, residential mobility, and neighborhood poverty. The results indicated that an increasing proportion of high-intensity developed-lands substantially increased the risk of death, while an increase in the aggregation and connectivity of vegetation-dominated low-intensity developed-lands reduced the risk of death. These findings suggested that places lacking greenspaces could have worse access to recreational sites that promote physical activity.

Overall, this dissertation expands our knowledge about the geographic disparities in colon cancer in New Jersey. It also provides specific examples of integrating residential histories and remote sensing-based products into cancer disparities research. Including residential histories opens up new avenues of inquiry to better understand the complex relationships between people and places, and the effect of residential mobility on cancer outcomes. Combining multiple socio-demographic and environmental domains to estimate the neighborhood effects on cancer outcomes will increase our potential to understand the underlying pathways.

ACKNOWLEDGMENTS

The path was long but would have been longer without the support and trust of several great people, whom I had a chance to meet during my educational development.

I am incredibly thankful to my advisors, Dr. Kevin Henry, Dr. Shannon Lynch, and Dr. Victor Gutierrez, for their enormous support and guidance during my time at Temple University. Additionally, a special thank you to Dr. Antoinette Stroup (Rutgers University) for constant advising, and Dr. Lee Hachadoorian and Dr. Charles Kaylor for helping me to develop the necessary analytic skills, and the Remote Sensing and Sustainability (rsensus) Lab (Temple University) for methodological support. I would also like to acknowledge my co-authors, Dr. Aniruddha Maiti, Dr. Slobodan Vucetic (Temple University), and Dr. Gerald Harris (Rutgers University), for their assistance on publishing these studies, as well as Dr. Bradley Gardener (Temple University) for inspiring to teach, and the entire Department of Geography and Urban Studies at Temple University for the established environment and support.

I also would like to thank my former mentors, Knut Neumann (Schule Königstrasse), Sabine Behrenwaldt (Emil-Krause Gymnasium), Dr. Peter Borchardt, Dr. Udo Schickhoff, and Dr. Jürgen Böhner (University of Hamburg), who substantially contributed to shaping my identity, and always encouraged me to proceed to higher education.

Finally, I would like to thank my wife Alina for her belief in me, and enlightenment, whenever it was needed. – You have been my absolute inspiration! And a special thank you to my entire family in Hamburg and Philadelphia and my friends across all countries for their support during these years.

This work was supported by the National Science Foundation (1560888). New Jersey State Cancer Registry data were collected using funding from NCI and the Surveillance, Epidemiology and End Results (SEER) Program (HHSN261201300021I), the CDC's National Program of Cancer Registries (NPCR) (5U58DP003931-02/NU58DP006279-02-00) as well as the State of New Jersey and the Rutgers Cancer Institute of New Jersey. Thank you to the Temple University, College of Liberal Arts, and the Department of Geography and Urban Studies for the provided funding and scholarship.

*Eine Hauptsache der Armut in den Wissenschaften ist meist eingebildeter Reichtum.
Es ist nicht ihr Ziel, der unendlichen Weisheit eine Tür zu öffnen,
sondern eine Grenze zu setzen dem unendlichen Irrtum.*

*[One of the chief causes of poverty in science is usually imaginary wealth.
The aim of science is not to open a door to infinite wisdom,
but to set a limit to infinite error.]*

Bertold Brecht (Leben des Galilie, 1928)

TABLE OF CONTENTS

ABSTRACT..... i
ACKNOWLEDGMENTS iii
LIST OF TABLES ix
LIST OF FIGURES..... x
LIST OF ABBREVIATIONS xi

CHAPETR 1: INTRODUCTION

Colon Cancer Survival Disparities..... 1
The Neighborhood Effect on Colon Cancer Survival Disparities 2
 The Role of Neighborhood’s Socioeconomic Status 2
 The Role of Neighborhood’s Landscape Characteristics..... 5
Geographic Mobility 6
Socio-Spatial Mobility and Cancer Disparities..... 9
Gaps in Cancer Disparities Research 12
Approach..... 16
Methodology 19

CHAPTER 2: SOCIOECONOMIC DISPARITIES IN COLON CANCER SURVIVAL:
REVISITING NEIGHBORHOOD POVERTY USING RESIDENTIAL HISTORIES.

Abstract24
Introduction26
Methods28
 Study Population..... 28
 Residential History Data from LexisNexis 29

Neighborhood Poverty Data	30
Examining Residential Histories.....	30
Generating Area-Level Poverty Measures (Neighborhood SES)	31
Statistical Methods.....	32
Results.....	34
Study Population.....	34
Colon Cancer Deaths	37
Discussion.....	38
Conclusion	44

**CHAPTER 3: RESIDENTIAL MOBILITY AND GEOSPATIAL DISPARITIES
IN COLON CANCER SURVIVAL.**

Abstract.....	46
Introduction.....	48
Materials and Methods.....	50
Study Population.....	50
Residential History Data from LexisNexis	52
Neighborhood Socioeconomic Status Data	53
Statistical Methods.....	53
Results.....	56
Study Population.....	56
Geographic Clustering and Spatial Effects	57
Discussion.....	60
Conclusion	65

CHAPTER 4: MEASURING NEIGHBORHOOD LANDSCAPES: ASSOCIATIONS
BETWEEN NEIGHBORHOOD’S LANDSCAPE CHARACTERISTICS AND COLON
CANCER SURVIVAL.

Abstract.....	66
Introduction.....	68
Materials and Methods.....	72
Study Population.....	72
Residential Histories	73
Socioeconomic Variables.....	74
Environmental Variables	75
Statistical Analysis.....	80
Results.....	82
Study Population.....	82
Univariate Models.....	85
Multivariate Models.....	86
Discussion.....	88
Conclusion	95
CHAPTER 5: CONCLUSION	97
REFERENCES CITED.....	104

LIST OF TABLES

Table 1: Overview of Research Questions and Statistical Models	20
Table 2: Study population characteristics	34
Table 3: Distribution of patient characteristics by the number of changes in residential census tracts	36
Table 4: Comparison of HRs for different CT-poverty measures estimated from the Cox regression models.....	38
Table 5: Study population characteristics	56
Table 6: Model Comparison	58
Table 7: Overview of the land cover classes after the reclassification of the NLCD dataset.	76
Table 8: Overview of the area-based and individual variables	77
Table 9: Study population characteristics	83
Table 10: Hazard Rates of individual-level factors from the multivariate model.	86

LIST OF FIGURES

Figure 1: Conceptual model of the organization of influential sites factors and cancer survival. Several individual (inner white circle) and neighborhood factors (outer blue circle) are associated with the risk of death after the diagnosis with cancer. However, they might not remain constant over the entire follow-up period (blue timeline), because of changes in the neighborhood environment and sociodemographic structures as well as because of the patient’s socio-spatial residential mobility (red-dashed box).	11
Figure 2: Description of the different CT-poverty measurements used in the analysis...	32
Figure 3: Geographic variation in risk of death estimates after adjustment for sex, age, race/ethnicity, sub-stage at diagnosis, and CT-poverty (a) using information from place of diagnosis and (b) using residential histories during the follow-up. Figure c shows the difference map of risk of death estimates between models a and b ($c = a - b$). Red and black circled areas show the CTs with risk of death differences that were in the opposite direction in the time-varying model compared to the diagnosis model.....	60
Figure 4: Land cover classification and landscape characteristics of New Jersey.	79
Figure 5: Workflow diagram.....	81
Figure 6: Correlation matrix before and after reduction of significant variables. A: All significant variables based on the univariate models. B: Reduced set of significant variables used in final multivariate models.	82
Figure 7: Distribution of colon cancer patients by neighborhood/landscape characteristics.....	84
Figure 8: Hazard Ratios of neighborhood-level variables from the crude models, adjusted for each variable individually. Blue line indicates there is no difference in risk. Red bars indicate the 95% confidence interval of statistically significant variables.	86
Figure 9: Hazard Ratios of neighborhood-level variables from the multivariate model adjusted for all individual-level factors and each neighborhood/landscape variable individually. Blue line indicates a zero-coefficient or no difference in risk. Red bars indicate the 95% confidence interval of statistically significant variables.	88

LIST OF ABBREVIATIONS

ACS: American Community Survey. It is operated by the United States Census Bureau and conducted monthly every year. In contrast to the decennial Census, the data are collected based on the population samples from every state/territory.

CI: Confidence Interval.

CT: Census tract statistical subdivisions of a county and geographic units of analysis of approximately 4,000 residents as defined by the Census Bureau.

DLC: Date of Last Contact with the patient after the diagnosis.

DX: Diagnosis.

GD: Geographic Disparity is defined as percent of remaining unexplained spatial variance. High values assume large geographic disparities in the area.

HR: Hazard Ratio is a measure of the risk of death. A value of 1.0 indicates no difference in the estimates in comparison to the reference group.

NJSCR: New Jersey State Cancer Registry is a population-based cancer registry and is a part of the National Cancer Institute's Surveillance Epidemiology and End Results program.

SES: Socioeconomic Status.

CHAPTER 1: INTRODUCTION

Colon Cancer Survival Disparities

Colon cancer is one of the three major causes of cancer-specific deaths in the United States. Each year approximately 50,000 men and women die from this disease (American Cancer Society, 2020). Despite decreasing colon cancer deaths rates since the 1960s, colon cancer remains a significant burden, mostly due to racial and urban/rural disparities in colon cancer mortality (Singh & Jemal, 2017; Singh, Williams, Siahpush, & Mulhollen, 2011). Race/ethnicity is one of the key factors of colon cancer survival and mortality disparities in the United States. Compared to other race/ethnicity groups, Blacks tend to have poorer colon cancer survival after the diagnosis (White et al., 2017; White, Vernon, Franzini, & Du, 2010). Moreover, improvements in survival rates during the past decades for African-Americans are lower than for Whites (Murphy, Wallace, Sandler, & Baron, 2019). Rural populations also experience higher risk of death from colon cancer than urban populations, which are only partially explained by access to healthcare and screening (Carmichael, Cowan, McIntyre, & Velopulos, 2020).

While several factors have helped to improve colon and rectal cancer survival, including improvements in screening and early detection, treatment, and targeting modifiable lifestyle behaviors (Hagggar & Boushey, 2009; Labianca et al., 2010), the differential prevalence rates of these risk factors by race also likely contribute to survival disparities. There has been much attention focused on early detection and preventative screening because patient survival varies significantly by the stage of the disease at the time of diagnosis. According to the American Cancer Society, the difference between the 5-year

survival rates in all age groups may vary between 90% for patients diagnosed with local stage, 71% for regional stage and only 14% for distant stage patients (American Cancer Society, 2020). Black patients are more likely to be diagnosed at advanced or distant stage disease compared to other race/ethnic groups (Mandelblatt, Andrews, Kao, Wallace, & Kerner, 1996), and are also less likely to undergo colorectal cancer screening (Cole, Jackson, & Doescher, 2013; Jerant, Fenton, & Franks, 2008; Trivers, Shaw, Sabatino, Shapiro, & Coates, 2008), which has shown to help identify colon cancers at lower, more treatable stages (Rutter, Knudsen, Lin, & Bouskill, 2021). Additionally, while increasing age is associated with shorter survival, recently, there are also increasing death events among younger population under the age of 55 (American Cancer Society, 2020). The increasingly higher colon cancer rates at younger ages are believed to be driven by increases in colorectal cancer in primarily Black and Hispanic populations (Murphy et al., 2019) as well as attributed to increasing obesity rates (Mauri et al., 2019). However, the exact reasons remain unknown (Low et al., 2020). These persistent and differential rates of colon cancer mortality and colorectal cancer screening practice by urban/rural residence and race/ethnicity, particularly in younger age categories, demonstrates the role that disparities appear to play in colon cancer survival.

The Neighborhood Effect on Colon Cancer Survival Disparities

The Role of Neighborhood's Socioeconomic Status

Cancer survival disparities are multilevel and influenced by genetics, individual characteristics, behavior, and factors related to the physical and social environments where people work and live (Warnecke et al., 2008). In short, cancer survival depends on

several internal (body-related) and external (place-related) environmental factors that are interconnected.

Previous studies have found that colon cancer survival disparities are associated not only with age at diagnosis and race/ethnicity, but also health insurance (Cabo, Shu, Shu, Parikh, & Bailey, 2020; Pulte, Jansen, & Brenner, 2018), access to care (Wan, Zhan, Lu, & Tiefenbacher, 2012), socioeconomic status (SES) (Aarts, Lemmens, Louwman, Kunst, & Coebergh, 2010; Du et al., 2007; Henry et al., 2014; D. Zhang, Matthews, Powell-Wiley, & Xiao, 2019) as well as geography (Chien, Schootman, & Pruitt, 2013) or the place where you live. A large body of literature on health and cancer disparities provides compelling evidence that the contextual factors of the neighborhood and population independently influence health outcomes like cancer survival (A. V. Diez Roux, 2001; Ana V. Diez Roux & Mair, 2010; Henry et al., 2014; D. Wiese, Stroup, Crosbie, Lynch, & Henry, 2019; D. Wiese et al., 2020b). Often, the effect of these contextual (e.g. neighborhood properties) and compositional (e.g. characteristics of individuals) factors is referred to as *neighborhood effect* (A. V. Diez Roux, 2001). The concept of neighborhoods effecting health outcomes is related to the social capital (Sampson, 2003), first described by Bourdieu (Bourdieu, 1986). Social capital can be defined as a social network, which is a reflection of a neighborhood's collective efficacy or social cohesion that is based upon trust and ability to develop societal self-regulation (Sampson, 2003), and is a product of both individuals (a neighborhood's residents) and political structuralism that variously affects different population groups and causes unequal opportunities (Eriksson, 2011). In the context of cancer survival outcomes, generally, patients from low-SES neighborhoods had worse survival than patients that are living in

high-SES neighborhoods (Aarts et al., 2010; Chien et al., 2013; Henry, Niu, & Boscoe, 2009; Lian et al., 2011; Steinbrecher et al., 2012). Additionally, low SES may cause delays in the diagnosis (Henry et al., 2014) leading to a later stage, and thus a worse survival outcome.

In addition to the studies that examine neighborhood effects on health outcomes, biosocial theory suggest that neighborhoods can also impact mental and psychosocial conditions of the residents. Wiese et al (2018) explain that changes in biological and physiological body systems (e.g. immune system, metagenome, epigenetic markers) are among factors that influence the individual reaction to disease regardless of genetics, and that space, place and time may influence those changes (D. Wiese, Rodriguez Escobar, Hsu, Kulathinal, & Hayes-Conroy, 2018). Additionally, Lynch and Rebbeck (2013) introduce a Multilevel Biologic and Social Integrative framework, which outlines the relation between macro-environment, individual, and biologic factors on cancer etiology (S. M. Lynch & Rebbeck, 2013). Specifically, this framework connects levels of carcinogenesis in the context of interventions (e.g., type of treatment), and translation/implementation (e.g., health care policy). They argue that several disease biomarkers essential for measuring the effect of exposure may have caused spontaneous mutation but developed a response to an initial macro-environment or individual level exposure as a reaction to physiological stressor and behavior. A cross-sectional study of 1,488 individuals revealed associations between blood leukocyte telomere shortage (related to cancer onset and other chronic diseases) and neighborhood factors, suggesting that unfavorable neighborhood sociodemographic circumstances may influence the biologic mechanisms (Shannon M. Lynch, Mitra, Ravichandran, et al., 2017).

The Role of Neighborhood's Landscape Characteristics

In contrast to the large number of studies focused on measuring the relationship between neighborhood SES and cancer outcomes, a focus on the role of the built environment in cancer outcomes like survival is less common. The built environment includes physical landscape characteristics of neighborhoods that are human-made or modified (Rapoport, 2011) such as buildings, roads, housing conditions, parks, sidewalks or greenspace that provide the setting for human activity (Renalds, Smith, & Hale, 2010). Understanding the role of the built environment in cancer outcomes is especially important given that the majority of the United States population lives in urban and urbanized areas.

There are various approaches and variables that can be used to examine the built environment or landscape characteristics. For example, different land cover types, area size, and patterns of mixed land use can be used to define the characteristics of a neighborhood's built environment. To date, these types of data have not been extensively utilized. Gomez and colleagues (Gomez et al., 2015) conclude that because some studies have found an association between the built environment and health behaviors including physical activity, more research using built environmental data is necessary to better understand cancer etiology and outcomes. A recent review on built environmental variables noted that only few studies examined cancer incidence, while no studies have examined the association between landscape characteristics and cancer survival (Kondo, Fluehr, McKeon, & Branas, 2018).

The evaluation of the neighborhood's landscape characteristics is challenging because of the different nature of the data (i.e. raster files), which cannot be joint to the patient's

neighborhood based on the census tract code. However, it can be accomplished using landscape ecology techniques. Landscape ecology is an environmental/geographic science approach of landscape characterization, evaluation, and design (Turner, Gardner, O'Neill, & O'Neill, 2001). The methodology includes quantification of landscape characteristics and features, and finds application in urban planning/aesthetics, biodiversity, species richness, and conservation as well as disease control studies (Evelyn Uuemaa, Mander, & Marja, 2013). Applying these techniques can help transform land cover data into more meaningful environmental *landscape metrics* essential for neighborhood landscape characteristics and even aesthetic. Combining environmental and socioeconomic data may increase the accuracy in quantifying neighborhood living quality and SES as well as increase capacities of neighborhood studies when accounting for complexity of influential factors on cancer outcomes.

Understanding the relations between neighborhood SES and landscape characteristics factors as well as their associations with the cancer outcomes may help to understand the nature of cancer disparities in the population and provide the impetus to implement policies to decrease the risk of death. Additionally, identifying neighborhood environments, which are associated with the cancer risk increase, may help in improving the identification of vulnerable populations (Gomez et al., 2015).

Geographic Mobility

As discussed, multiple studies report associations between cancer outcomes and neighborhood SES and environmental conditions. However, the influential neighborhood factors may not necessarily remain constant over time because of changes in the

environment due to physio-geographic or socio-demographic processes, policy, and, nevertheless, geographic and social mobility.

Geographic or spatial mobility assumes any residential relocation. Traditionally, a relocation over a short distance is referred to as residential mobility, while long distance moves are referred as migration (Gilliland, 1998). Earlier frameworks of geographic mobility suggest that positive migration flow would go into the newly expanded urban peripheries as a result of industrialization and transportation development (Long, 1988). Later on, geographic mobility was explained by the human capital models, and individual estimations of costs and benefits associated with residential moves (Gillespie, 2016). Modern-day geographic mobility may be influenced by several factors such as family, education, occupation, language, ancestry, ethnicity and even policy. Coulter and Scott (Coulter & Scott, 2015) distinguish between *targeted* and *diffuse* reasons for residential relocation, and conclude that life-course events may influence moving motivations and behavior. Additionally, mobility theories assume associations between distance and motivations: employment/profession-driven relocations can go over longer distances (Böheim & Taylor, 2007), while short-distance moves may be associated with neighborhood preferences (van Ham, 2012). However, the choice of changing residence is not always voluntary, and the choice of where to move is often decided based on available financial resources (W. A. V. Clark & Morrison, 2012; Hulchanski, 2010). Three forms of residential mobility are identified: 1) *preference dominated* (individual decision because of living environment dissatisfaction), 2) *imposed* (driven by a life-course event such as family expansion), and 3) *forced* (involuntary relocation) (Gillespie, 2016). Lack of resources and absence of institutional support can lead to enforced change

of residency (Sell, 1983), evictions (Skobba & Goetz, 2013) and/or deprivation *traps* on both individual and geographic levels (Meen, 2009). It results in missing chances escaping the *trap*, and consequently the concentration of poverty in certain neighborhoods, which can become socio-economically segregated (Glennster, Lupton, Noden, & Power, 1999; Power & Wilson, 2000; Townsend, 2002).

In contrast to geographic mobility, social mobility assumes changes in an individual's or group's socioeconomic position over the life course independent of physical relocation. However, geographic mobility may be associated with the movement between different residential neighborhoods with various levels of socioeconomic deprivation. – The so-called *socio-spatial mobility* can be influenced by industrialization, immigration, urbanization (Strauss, 2009), and employment (Purcell, 2020). Traditional social mobility theories argue that the direction of SES may vary through time, influencing not only individual people, but also families, neighborhoods, communities, and regions (Strauss, 2009). However, social mobility may also be regulated through policies. Padgett (1990) (Padgett, 1990), for example, summarizes that according to Weber's and Marx's theories, social mobility is controlled through centralized management (e.g., government) in order to preserve the social class structure. In contrast, considering a point of view where social mobility is self-selection and a sorting process based on individuals' education-level, skills, and preferences (Padgett, 1990) would suggest that everyone has opportunities to improve their socio-economic position. In this case, the definition of social class and related mobility patterns is usually associated with opportunities for selecting occupations and employers.

Thernstorm (1968) argues that historically the United States' population has experienced upward social mobility over several generations (Thernstrom, 1968). However, it usually remains relatively low on the individual level, with children ending up in the same or similar socioeconomic position as their parents (Thernstrom, 1968). In contrast, Straus (1971) identified several aspects of the individual and collective downward mobility in the United States. She argues that enforced labor, gender, race/ethnicity, and natural and social disasters are primarily reason for a decrease in social mobility (Strauss, 2009) because ethnic minorities and women historically experienced fewer opportunities in occupation and social mobility than men (Jacobs, 1990).

Socio-Spatial Mobility and Cancer Disparities

There has been an ongoing interest in examining socio-spatial mobility patterns in a population, among others, due to their association with health outcomes. Geographic mobility appears to be one of many stressors that negatively affect individual health. Previous studies suggest that relocation before the cancer diagnosis or shortly after is an essential psychosocial stressor (Lix et al., 2006; McGrath & Rawson, 2013). Geographic mobility negatively affects mental and physical health conditions (Lin, Huang, Bai, & Kuo, 2012) and may result in a delayed diagnosis and earlier mortality (Muralidhar, Nguyen, & Tucker-Seeley, 2016). Additionally, residential relocation may cause a loss of access to treatment or be a new barrier (Baugh & Verghese, 2013) and can alter social networks that can have psychosocial impacts (Cornwell & Waite, 2009). Previous studies on colon cancer survival suggest that change of residency and neighborhood SES trajectories might be as important as other individual-level factors such as age and

race/ethnicity in explaining existing disparities in the risk of death (Shvetsov et al., 2020; D. Zhang et al., 2019). At the population-level, residential relocation can also impact neighborhoods and places by changing the sociodemographic structure and other contextual factors (Gillespie, 2016).

Mobility experience theory suggests that changing residence is a combination of social and environmental experiences. It is based on previous migration processes, including motivational reasons for moving, the decision about where to move, distance moved, and adaptation to the conditions at their new home (Bolan, 1997). On the one hand, the positive or negative experiences related to moving can influence health and well-being (Brett, 1982; Stokols & Shumaker, 1982; Stokols, Shumaker, & Martinez, 1983). On the other hand, Dunn et al. (Dunn, Winning, Zaika, & Subramanian, 2014) argue that changing residence is also influenced by the current health status and is underlying a reciprocal process. In their study, they found that the willingness to move was influenced by poor health and desire to change the environment. Thus, understanding the mobility experience theory (e.g. role of the socio-spatial mobility) may help to understand better the etiological mechanism behind diseases and outcomes (Richter, Günther, & Herke, 2018).

Geographic and socioeconomic movements are interconnected and may be influential in cancer outcomes and thus require more attention. Figure 1 introduces a conceptual model of the organization of influential sites factors and cancer survival. After the diagnosis, patient's survival time and the risk of death is influenced by several individual (inner white circle) and neighborhood factors (outer blue circle). These factors include

physiological and genetic characteristics such as age and stage at the diagnosis, sex/gender and race/ethnicity, marital status as well as general physical and mental health. – All are strong predictors of cancer survival outcomes. However, the role of neighborhood factors such as SES, access to health care, social cohesion and even nature, walkability, and neighborhood aesthetics have also been shown to be important in several health outcomes (Akpinar, Barbosa-Leiker, & Brooks, 2016; Bratman et al., 2019; Chen, Stephens, & Jones, 2019; Hunter & Brown, 2012; Mears & Brindley, 2019; Mears, Brindley, Jorgensen, & Maheswaran, 2020; Wendelboe-Nelson, Kelly, Kennedy, & Cherrie, 2019). However, a key question is *how do these factors impact patient survival when we account for both residential and social mobility?* – This question is important because the majority of studies focused on neighborhood effects and cancer outcomes utilize only a single measure based on the patient’s location at time of diagnosis.

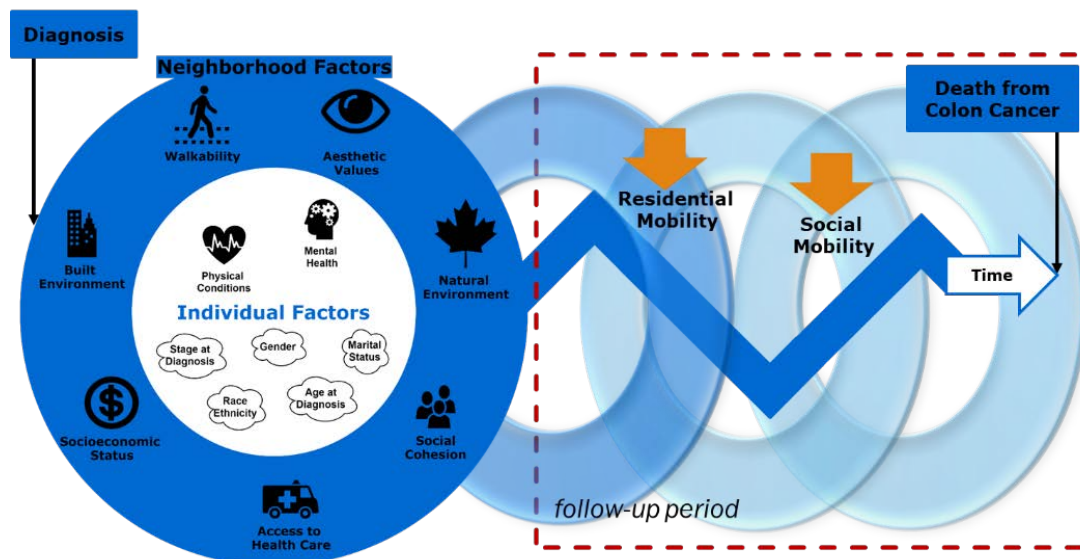


Figure 1: Conceptual model of the organization of influential sites factors and cancer survival. Several individual (inner white circle) and neighborhood factors (outer blue circle) are associated with the risk of death after the diagnosis with cancer. However, they might not remain constant over the entire follow-up period (blue timeline), because of changes in the neighborhood environment and sociodemographic structures as well as because of the patient’s socio-spatial residential mobility (red-dashed box).

Note: Figure developed for this dissertation using draw.io and MS PowerPoint

Gaps in Cancer Disparities Research

Social epidemiologists have been increasing interest in integrating life course approaches when studying cancer risk and outcomes like survival. To incorporate exposures to social and physical environments over time, researchers are interested in finding ways to move beyond the standard approach of only including neighborhood environmental variables (e.g., census tract poverty) at the time of diagnosis. They are looking towards integrating neighborhood environmental variables over time by incorporating residential histories. Researchers are well aware that populations are mobile and that neighborhood conditions change over time. And most importantly, that the timing of social and physical environmental exposures during different life stages and their accumulation can impact risk of disease and outcomes like survival. Despite a keen interest in integrating residential histories into social-epidemiologic investigations, most neighborhood population-based studies still utilize cross-sectional study designs by only using the cases' residential location at the time of diagnosis (Singh & Jemal, 2017; D. Zhang et al., 2019). The main reasons why there are so many cross-sectional studies are: 1) cancer registries, the primary source of population-based cancer data, do not yet routinely collect residential histories, and 2) researchers have not invested in collecting these data to the extent that has been done in environmental epidemiology.

Studies that examine the effects of geographic exposures or neighborhood attributes (e.g., poverty) and use only the residence at the time of diagnosis to measure the exposure, face a fundamental methodological problem related to the spatial and temporal uncertainty about the timing and duration of exposure (Kwan, 2012). Using location at the time of diagnosis limits the researcher's ability to assess how different exposures over time based

on past residences might result in different outcomes compared to a single location. Routinely integrating residential histories into cancer registry data to account for the timing of neighborhood or residential exposures would allow us to ask new social epidemiological questions and conduct environmental epidemiological studies more frequently. Integrating residential histories into cancer data will help us to move beyond the limitation of the cross-sectional study and provide new avenues for measuring the spatial-temporal variation of contextual or neighborhood-based (e.g., poverty, segregation) or environmental influences (e.g., contaminated water), identifying when individuals are potentially affected by them (e.g., critical periods), and evaluating the accumulated exposures with respect to residential mobility. For example, *does an upward trajectory of neighborhood SES result in better survival outcomes than individuals with a declining trajectory of SES over time?*

In geographic studies, the residential information can also be used to identify geographic regions or clusters of elevated incidence rates or odds ratios (G. M. Jacquez et al., 2005), which may act as an evidence of some past environmental exposure in certain places. Since the primary source of location data in geographic studies of cancer remains residential address at the time of diagnosis, the full extent of not using residential histories in cancer disparities research remains mostly unknown and requires further investigations. In conclusion, residential histories provide new research opportunities to study cancer survival because they are pertinent to life course theory and focus on exposures during critical periods and accumulative exposures (Espejo-Herrera et al., 2016; Hystad, Demers, Johnson, Carpiano, & Brauer, 2013; Geoffrey M. Jacquez et al., 2014; Ling et al., 2019; Roswall et al., 2016; Ruder & Bertke, 2017; Schullehner,

Hansen, Thygesen, Pedersen, & Sigsgaard, 2018; S. P. Tsai et al., 2004; D. Zhang et al., 2019).

The National Cancer Institute (NCI) has directed efforts to develop and improve maps and geospatial epidemiological methods across the cancer continuum (cancer development, detection, diagnosis, treatment, mortality, and survivorship). In 2016, the NCI's Division of Cancer Control and Population Sciences sponsored a conference to bring together researchers conduct geospatial research across the cancer control continuum to develop new methods and tools and address cancer disparities (M. Schootman et al., 2017). Furthermore, through NCI's Surveillance, Epidemiology, and End Results Program (SEER; <https://seer.cancer.gov>), and CDC's national cancer control prevention programs funding is allocated for population-based cancer registries to routinely geocode the residential location at the time of diagnosis. These data are regularly used to conduct cancer health disparities research. Despite NCI and CDC's efforts, an increasing number of studies that utilize Geographic Information System (GIS), significant gaps and limitations in current geospatial methods and approaches remain.

As part of the NCI's 2016 conference on Geospatial Approaches to Cancer Control and Population Sciences, Schootman et al. (2017) summarized key methodologic issues and gaps in current approaches. They stated that future research should 1) *incorporate attributes of both secondary data and self-reported* perceptions about neighborhoods, going beyond the use of administrative boundaries as neighborhoods; 2) *integrate residential history* information so researchers can examine exposures at critical periods, cumulative exposures, and exposure pathways; 3) use conceptual and theoretical models

that *integrate various types of data to measure environmental and community contexts* (such as work, residential, and activity settings) as well as biological and social factors; and 4) *develop or modify existing analysis methods*, so they are appropriate for the geospatial and multilevel nature of the data (M. Schootman et al., 2017).

The primary goal of my dissertation is to address the current gaps in geospatial research related to the incorporation of various data formats, integration of residential histories and expansion of spatial modeling techniques in cancer survival disparities while answering three important questions:

- Does **including residential histories change the risk of death estimates** by neighborhood poverty **compared to the traditional approach** when including only the location at the time of diagnosis?
- Does the inclusion of **residential histories in spatial models change the geographic risk estimate patterns** compared to the models, where only the location from the time at diagnosis was used?
- Is there a significant association between **landscape characteristics (e.g. built environment)** and **the risk of death** independent of individual-level factors and neighborhood-poverty?

In chapters 1 and 2, I utilize residential histories of colon cancer patients to account for socio-spatial mobility and estimate whether the risk of death associated with neighborhood-poverty does significantly vary between time-varying and traditional single-value/single-location models. Additionally, the integration of residential histories

allows to evaluate the extent of geographic bias related to changes in residency and neighborhood-poverty during the follow-up period.

In chapter 3, I investigate the associations between neighborhood landscape characteristics (e.g. composition and configuration of roads, buildings and greenspaces) and the risk of death, while including remote sensing-derived products and residential mobility.

Approach

I use approaches from medical geography as a framework for my dissertation research.

Medical geography is a sub-discipline of geography that applies geographic/spatial methodologies to investigate medical and health outcomes. Its roots are in pre-20th-century tropical medicine research, which shifted toward an inclusive field with various specializations, diverse research questions, and tools. Early works in medical geography focused on disease diffusion and ecology, especially the description of tropical diseases (e.g., malaria, yellow fever). May (1950) is considered to be the first scientist to develop a conceptual framework of disease ecology or landscape epidemiology (May, 1950).

Following that, cartographic production and knowledge of cultures were the key elements that geographers brought into medical sciences and understanding of diseases.

While early works in medical geography were concentrated on descriptions of infectious diseases from a biomedical perspective (Audy, 1954), increased interest by geographers in health inequalities (Emch, Root, & Carrel, 2017) led to integration of nontraditional frameworks into ecological models of human disease (McLeroy, Bibeau, Steckler, & Glanz, 1988). Now, medical and health geography could be defined as a combination of theory- and method-driven research mostly including but not limited to healthcare

planning, disease and landscape ecology, cluster analysis, critical health geography, geography of health and place, and political ecology (Emch et al., 2017). While medical and health geography shares many theoretical frameworks and approaches, health geography also focuses on emotional and personal experiences. In contrast, medical geography is focused on analyzing population-related health outcomes in relation to geospatial factors and mapping distributions of disease.

Additional characteristics differentiate medical geography from other related disciplines such as social epidemiology. For example, multilevel modeling is commonly used in epidemiological studies on cancer disparities. Most neighborhood studies in epidemiology, however, are aspatial in nature. Also, while social epidemiology was able to account for previously unexplained effects related to the social environment and cohesion (Berkman, Kawachi, & Glymour, 2014), medical geography researchers have shown that geographic location, spatial characteristics, and relations are important factors explaining health behavior and related outcomes. Several researchers have contributed to understanding the importance of healthcare access, showing that location and transportation modes directly influence population-based health outcomes (Wan, Zhan, Lu, et al., 2012; Wan, Zhan, Zou, & Chow, 2012; Wan, Zou, & Sternberg, 2012; F. Wang, 2012; F. Wang & Luo, 2005; L. Wang, Wilson, Stewart, & Hollenbeak, 2011). Other medical geographers presented that effects of SES and related disparities follow geographic patterns and can be explained by accounting for spatial relations (Chien et al., 2013; Henry et al., 2009; Henry, Stroup, Warner, & Kepka, 2016; Roche, Niu, Stroup, & Henry, 2017; Sahar et al., 2019; M. Schootman et al., 2017). Therefore, neighborhood SES-related disparities can be eliminated by concentrating on specific areas, while

ignoring geographic location can result in biases in epidemiological studies (Oliver, Matthews, Siadaty, Hauck, & Pickle, 2005) or be inconclusive in healthcare-quality research (Rushton, 2003). Nevertheless, application of correct mapping tools and geospatial modeling techniques affects the presentation of disease (Beyer, Tiwari, & Rushton, 2012; Rushton, 2003).

Medical geography's unique position as a discipline enables researchers to study diseases and health disparities from various perspectives (Pyle, 1976). The combination of social and environmental sciences' approaches and theories also resulted in a broad spectrum of available quantitative methods. Therefore, the value of the GIS must be presented as one of the major factors that enable the study of population, habitat, and behavior of the human disease ecology. Progress in GIS, particularly, allowed the implementation of theoretical assumptions into applicable case studies. While earlier research in medical geography used cartography exclusively for descriptive purposes, GIS tools opened opportunities to account for spatial relations, variations, and modeling in infectious and chronic disease studies (Sui, 2007). Since the early 1990s the GIS has become the system in which data of various origins and formats are combined (van Beurden & de Lepper, 1995), stored, manipulated, and spatially displayed, helping identify spatial patterns and potential causes for disease clustering (Douven & Scholten, 1995). Its importance is still growing, and it is dominating medical and health geography research (Rosenberg, 2015). The added value of the GIS to health analysis has a broad range, from cluster identification to provision of data that would be almost impossible to obtain from field studies, to overlay of information (Glass, 2000). GIS has significantly contributed to public health research in terms of data manipulation, analysis, and visualization (Rushton,

2003), spatial noise reduction (A. Lawson et al., 1999; A. B. Lawson, 2001; A. B. Lawson, Williams, & Williams, 2001), application of spatial smoothing (Kafadar, 1996, 1997, 1999, 2014), and communication (Sahar et al., 2019).

In contrast to several *aspatial* socio-epidemiological studies that identified disparities by neighborhood SES measures (e.g. high and low SES), GIS and GIScience enables researchers to study relations between the location and health outcomes and disparities (Sahar et al., 2019); including population residential mobility by identifying associations between time of exposure to risk factors and developing a disease (Francis P. Boscoe, 2011). In particular, cancer control and prevention has seen an increasing amount of geospatial approaches that consider the geographic patterns in cancer incidence and mortality in the United States (Sahar et al., 2019; M. Schootman et al., 2017); that is where cancer remains a burden across all populations and regions.

Using medical geography, GIS, and multilevel modeling approaches for my dissertation, I highlight and focus on the importance of geographic location when analyzing colon cancer survival disparities. My work contributes to a variety of geospatial epidemiological approaches for study cancer outcomes by expanding applied tools and methods while incorporating residential histories as well as utilizing remote sensing-derived products.

Methodology

Multilevel models are commonly used in epidemiological studies to reflect the complexity of interactions between individual characteristics and contextual factors and their influence on cancer outcomes (S. M. Lynch & Rebbeck, 2013). Multilevel models allow the inclusion of data of varying scales (e.g. individual and neighborhood factors)

into regression analysis. They also have a clear ability to identify importance of individual and neighborhood effects, whereby the characteristics of a group are considered as influential factors for personal health outcomes (Zahnd & McLafferty, 2017). While multilevel modeling research originates in social epidemiology, increased progress in GIS and implementation of spatial tools in multilevel analysis became an essential part of geographic cancer disparities research. Among other, geospatial multilevel models allow estimation of the geographic risk of death ratios and definition of statistically significant clusters of elevated risk of death.

Several statistical models are applied (Table 1). To estimate the effects of neighborhood-poverty on colon cancer survival using residential histories vs. place of the diagnosis (Question 1), I first developed two non-spatial proportional Cox hazard regressions.

Then, to evaluate the spatial differences and to estimate the potential geographic bias in the risk estimates (Question 2), I am applying a Bayesian extension of conventional Cox regression survival models. Lastly, time-varying multilevel models are utilized for the evaluation of the landscape characteristics (Question 3).

Table 1: Overview of Research Questions and Statistical Models

	Research Question	Modeling Outcome	Model Type
Chapter 1	Are the risk estimates for neighborhood poverty different in models when using single value from models with residential histories?	Hazard Ratios (HR) – Risk of death estimates for every independent variable. HR range -1 (Low Risk) to +1 (High Risk)	Aspatial analysis. Cox proportional hazard regression survival models (Cox Regression) including a) time-independent (single-value), b) time-varying neighborhood poverty

Table 1 continued

Chapter 2	Does the inclusion of residential histories in spatial models changes the geographic risk estimate patterns compared to the models, where only location from the time at diagnosis was used?	Geographic risk estimates – Spatially smoothed risk of death estimates for every location (census tract). Range -1 (Low Risk) to +1 (High Risk)	Spatial analysis. Bayesian extension of conventional Cox regression survival models including a) time-independent covariates (single-value), b) time-varying covariates
Chapter 3	Is there a significant association between physical/natural landscape characteristics and the risk of death after controlling for the individual-level factors and neighborhood-poverty?	Hazard Ratios (HR) – Risk of death estimates for every independent variable. HR range -1 (Low Risk) to +1 (High Risk)	Aspatial analysis. Cox proportional hazard regression survival models (Cox Regression) including time-varying land cover data

In conventional survival analysis, T_i typically denotes the observed survival time for patient i , and C_i denotes the right censoring time. Therefore, a patient's observed time is defined as $t_i = \min(T_i, C_i)$, and δ_i is used to denote whether the patient is censored or not (e.g., 1=death, 0=censored). $x_1 \dots x_n$ are the covariate vectors (i.e., explanatory variables) for the patients.

The traditional Cox proportional hazard model is defined as:

$$\text{Model 1: } \log[h(t_i, \delta_i, x_i)] = \log[h_0(t_i)] + x_i\beta$$

where $\log[h(t_i, \delta_i, x_i)]$ is the log-baseline function. Model 1 finds application in chapters 1 and 3. The utilized Bayesian model in chapter 2 is an extension of Model 1. It was described by Kneib and Fahrmeir (Kneib & Fahrmeir, 2007), and includes s_i , which represents the geographic location of the patients census tract at the time of diagnosis.

The spatial function $f_{spat}(s_i)$ is used to estimate the spatial effect.

The final spatial model can be established as:

$$\text{Model 2: } \log[h(t_i, \delta_i, x_i, s_i)] = \log[h_0(t_i)] + x_i\beta + f_{spat}^1(s_i)$$

The spatial function allows estimating the geographic variation in the risk of death estimates after controlling for various covariates (individual and neighborhood variables) (Nikolaus Umlauf, Adler, Kneib, Lang, & Zeileis, 2012). The geographic risk estimates measure the instantaneous event rate or the probability that an individual would experience an event (death from colon cancer) at a particular time point after diagnosis in comparison to the overall average. The spatial function is based on stationary Gaussian random fields (bivariate penalized splines). P-spline smoothed spatial effects were incorporated into the model based on an adjacency matrix of geographic neighbors (weights based on rook's case) by neighborhoods (e.g., Census tracts) (Brezger & Lang, 2006; Lang & Brezger, 2004). In case of high spatial heterogeneity or small numbers, no smoothing is applied (Belitz et al., 2015). The geographic risk estimates for each census tract is the smoothed rate for the census tract based on those living in that area at time of diagnosis. In the time-varying settings, the geographic risk estimates are estimated as a cumulative value weighed by the time of residency of each patient in each census tract where they lived.

In contrast to traditional models, Bayesian approaches use Markov chain Monte Carlo simulation in generalized additive and semiparametric mixed models. These models allow an inclusion of different types of covariates (e.g. multilevel or time-varying longitudinal data), metrical covariates with non-linear effects, unstructured random effects accounting for overdispersion caused by unobserved heterogeneity or for correlation in longitudinal or spatial data, spatial effects and different forms and degrees of smoothness (Fahrmeir & Lang, 2001). A detailed description of the model parameters and settings as well as the utilized software are provided separately in each chapter.

The following content is organized into four chapters. Chapter 2 examines the risk of death estimates associated with the neighborhood-poverty while comparing two models when using a single-value and time-varying approach in a non-spatial setting. Chapter 3 builds upon these findings and examines the risk of death estimates in geospatial settings, while integrating geographic mobility and estimating the spatial bias. In chapter 4, residential mobility data are used to capture not only the neighborhood-poverty but also the change in the landscape characteristics. Thereby, remote sensing-derived products are transformed into landscape metrics/indices and are integrated into non-spatial models in order to compare the risk of death estimates of the landscape characteristics after controlling for neighborhood SES. The conclusive chapter summarizes and discusses the major findings from my dissertation and provides directions for future research on cancer survival disparities.

CHAPTER 2:
SOCIOECONOMIC DISPARITIES IN COLON CANCER SURVIVAL:
REVISITING NEIGHBORHOOD POVERTY USING RESIDENTIAL HISTORIES.

Abstract

Residential histories linked to cancer registry data provide new opportunities to examine cancer outcomes by neighborhood socioeconomic status (SES). We examined differences in regional-stage colon cancer survival estimates comparing models using a single neighborhood SES at diagnosis to models using neighborhood SES from residential histories.

We linked regional-stage colon cancers from the New Jersey State Cancer Registry diagnosed from 2006-2011 to LexisNexis administrative data to obtain residential histories. We defined neighborhood SES as census tract poverty based on location at diagnosis, and across the follow-up period through 31 December 2016 based on residential histories (average, time-weighted average, time-varying). Using Cox proportional hazards regression, we estimated associations between colon cancer and census tract-poverty measurements (continuous and categorical), adjusted for age, gender, race/ethnicity, regional substage, and mover status.

Sixty-five percent of the sample were non-movers (one census tract); 35% (movers) changed tract at least once. Cases from tracts with >20% poverty changed residential tracts more often (42%) than cases from tracts with <5% poverty (32%). Hazard ratios (HRs) were generally similar in strength and direction across census tract-poverty measurements. In time-varying models, cases in the highest poverty category (>20%) had

a 30% higher risk of regional-stage colon cancer death than cases in the lowest category (<5%) (95% confidence interval [CI] 1.04-1.63).

Residential changes after regional-stage colon cancer diagnosis may be associated with a higher risk of colon cancer death among cases in high-poverty areas. This has important implications for post-diagnostic access to care for treatment and follow-up surveillance.

This chapter was published in present form in the *Epidemiology* journal.

Full citation:

Wiese, D., Stroup, A. M., Maiti, A., Harris, G., Lynch, S. M., Vucetic, S., & Henry, K. A. (2020). Socioeconomic Disparities in Colon Cancer Survival: Revisiting Neighborhood Poverty using Residential Histories. *Epidemiology*, 31(5), 728-735.
doi:10.1097/ede.0000000000001216

Introduction

Residential histories are valuable for longitudinal assessments of exposures to environmental hazards when evaluating the risk of cancer and other diseases (Francis P. Boscoe, 2011). Researchers can use residential histories to examine how environmental exposures at critical periods across the lifespan or the cumulative effects of exposures over time potentially impact cancer incidence and outcomes, such as stage at diagnosis or survival (Espejo-Herrera et al., 2016; Hystad et al., 2013; Geoffrey M. Jacquez et al., 2014; Ling et al., 2019; Roswall et al., 2016; Ruder & Bertke, 2017; Schullehner et al., 2018; S. P. Tsai et al., 2004; D. Zhang et al., 2019). For decades, researchers in the United States obtained residential histories for case–control or cohort studies through interviews or questionnaires. This approach is labor intensive, expensive, and subject to recall bias (Francis P. Boscoe, 2011; Geoffrey M. Jacquez et al., 2014; G. M. Jacquez et al., 2011). Recently, scientists have not only recommended that residential histories be routinely collected in case–control and cohort studies, but also in population-based disease surveillance systems, including cancer registries (Chittleborough, Baum, Taylor, & Hiller, 2006; Han et al., 2004; Geoffrey M Jacquez, Meliker, & Kaufmann, 2007; Urayama et al., 2009), that currently limit collection to residential address at the time of diagnosis. To manage the resource and scientific challenges of collecting residential histories, some investigators looked toward existing data sources and demonstrated the usefulness of residential histories collected from electronic medical records (S. Hurley et al., 2017) and of commercial and administrative databases to collect residential histories (Susan Hurley et al., 2017; G. M. Jacquez et al., 2011; Stinchcomb & Roeser, 2016; D. C. Wheeler & Wang, 2015).

While the main focus of using residential histories in cancer epidemiology has been to assess disease risk in relation to chemical exposures from water, soil, and air, residential histories can also be used in epidemiologic applications to examine cancer risks and outcomes associated with neighborhood factors that capture social or economic context (e.g., poverty and residential racial segregation)(Singh & Jemal, 2017). Area-based or neighborhood factors collected at the time of diagnosis, particularly neighborhood-level poverty, have been shown to be associated with cancer outcomes, including patient survival (Feinglass, Ryzewski, & Yang, 2015; 2015; Henry et al., 2009; Lian et al., 2011; Warner & Gomez, 2010). These studies have generally found that patients living in low-income/high-poverty areas at the time of diagnosis have lower survival rates compared to patients living in higher income/low-poverty areas. However, these studies used a single address based exclusively on residence at the time of diagnosis to measure neighborhood socioeconomic status (SES). This single-address approach is subject to a type of exposure estimation error and bias in the associations because it relies on only one neighborhood SES measurement and therefore could lead to an over- or underestimation of the true effect size (Brokamp, LeMasters, & Ryan, 2016). For studies examining patient survival, using a single address based on the patient's residence at the time of diagnosis introduces spatial uncertainty about their neighborhood SES because it does not account for possible changes due to residential mobility during the follow-up period. Therefore, it is unknown whether survival estimates by neighborhood SES would change if residential histories were used to measure a patient's neighborhood SES instead of using only the residential location at time of diagnosis.

In the present study, New Jersey State Cancer Registry (NJSCR) data were linked to administrative data from LexisNexis to obtain residential history on patients diagnosed with colon cancer in New Jersey from 2006 to 2011. The purpose was to determine whether survival estimates by neighborhood poverty are different if the patients' residential histories were used to measure poverty instead of the more commonly used residence at time of diagnosis.

Methods

Study Population

Colon cancer cases were obtained from the NJSCR, which is a population-based cancer registry established in October 1978 to monitor cancer among the more than 8.9 million residents of New Jersey (NJSCR, 2018). The NJSCR is nationally recognized for its high quality and timely submission of data to the Centers for Disease Control and Prevention, the National Cancer Institute, and the North American Association of Central Cancer Registries (NAACCR). All study activities were approved by the Rutgers University institutional review board.

The study population includes all adult New Jersey residents with their first, histologically confirmed regional stage colon cancer defined according to the International Classification of Diseases for Oncology, 3rd Edition (ICD-O3 C180-C189, C260; excluding histology codes 9050-9055, 9140, 9590-9992)(Percy, Holten, Muir, & Organization, 1990) diagnosed between 1 January, 2006, and 31 December, 2011 (N = 4,041). Individual-level variables provided by NJSCR include age at diagnosis (18+), gender (male, female), race (White, Black, Asian/Pacific Islander, Other), and ethnicity (Hispanic, Non-Hispanic). Cases with any stage at diagnosis other than regional

were excluded from analysis because of large disparities in survival time (American Cancer Society, 2019).

Vital status (1 = dead, 0 = alive) from NJSCR was captured through routine linkages with New Jersey Department of Health Office of Vital Statistics and Registry, the National Death Index, hospital discharge files, national Medicare and Medicaid files, Social Security Administration, and driver's license data from the New Jersey Motor Vehicle Commission. Patients were followed until 31 December, 2016. Cause of death was also provided by NJSCR (coded from Office of Vital Statistics and Registry and National Death Index data). Deaths from colon cancer were identified according to ICD-10-CM coding: C18 (Percy et al., 1990).

Residential History Data from LexisNexis

The NJSCR obtained residential histories of colon cancer cases through a data linkage with a commercial database developed by LexisNexis, Inc. (Miamisburg, Ohio, U.S.). LexisNexis collects consumer data, including credit checks, and in recent years has developed a specific resource for researchers to obtain residential histories for adults 18 and older (Susan Hurley et al., 2017). For each unique case record, LexisNexis returned up to 20 of the most recent addresses between 1946 and 2018 with documented start and stop dates at each address. The NJSCR used the NAACCR AGGIE Geocoder (Texas A&M, 2016) to geocode all residential addresses to their 2010 census tract boundaries ("neighborhood").

Neighborhood Poverty Data

A neighborhood SES measure, such as the poverty level of a census tract, captures conditions that may affect the health of individuals living in the same neighborhood. It has been shown to be independently associated with cancer outcomes, including stage at diagnosis and survival (Chien et al., 2013; Henry et al., 2009; Henry et al., 2014; Niu, Pawlish, & Roche, 2010). Further, while there are a number of neighborhood SES measures used in literature (Gomez et al., 2011), the percentage of the census tract population 18 years and older living below the federal poverty level (census tract-poverty) is one of the most commonly used socioeconomic status variables in cancer surveillance research (F.P. Boscoe, 2010; Henry et al., 2009). Census tract-poverty (CT-poverty) was obtained from publicly available U.S. Census and American Community Survey (ACS) data. For addresses between 2006 and 2010, we used U.S. Census 2010 and the ACS 5-year average data from 2006-2010, 2007-2011, 2008-2012, 2009-2013, and 2010-2014. ACS 2011-2015, 2012-2016, and 2013-2017 were used for addresses between 2011 and 2016. Each CT-poverty measurement was classified in four categories: <5% (low poverty), 5% - <10%, 10% - <20%, ≥ 20 high poverty) as suggested by other researchers (Krieger, Chen, Waterman, Rehkopf, & Subramanian, 2005; Krieger et al., 2003; Krieger et al., 2002).

Examining Residential Histories

Of the 4,041 regional stage colon cancer cases, 92 (2%) had no LexisNexis residential information. Residential histories for the remaining 3,949 (98%) and duration at each residence in the form of start and end dates were obtained from LexisNexis. However, as

previous studies have shown, there is a significant overlap in start/end dates due to forwarding of mail after a move (Stinchcomb & Roeser, 2016) or accounts remaining active after a move (e.g., bank accounts). In these scenarios, when multiple residences are available at a given time, we retained the most recently entered address (i.e. newest) similar to previous studies (S. Hurley et al., 2017; Ling et al., 2019; Stinchcomb & Roeser, 2016). Missing segments (gaps) of address history (N = 59; 0.2% of total time) were filled by imputing the address immediately preceding the gap. Then, we calculated the average time (months) of residency at each location by summing the number of unique census tracts divided by the total follow-up time. If patients remained at the same tract from diagnosis to the end of follow-up, they were classified as non-movers (i.e., no change in census tracts over time). If the tract changed at least once after diagnosis, we classified these patients as movers.

Generating Area-Level Poverty Measures (Neighborhood SES)

To calculate CT-poverty at each location, we split the entire follow-up period into time segments and calculated the amount of time (months) spent at each location. For each patient, every residential data point received a corresponding CT-poverty value. In situations where cases remained at a single location for the duration of follow-up, we assigned CT-poverty values that matched with those years. This allowed us to account for CT-poverty changes within the same tract, even if the patient did not move (e.g., due to gentrification). The average number of locations across all cases was 1.4 (range 1-10).

We developed four different poverty measures that are described in the Figure 2.

- Measure 1 – Standard approach to assign CT-poverty from the residential location of the patient at the time of diagnosis (DX CT-poverty).
- Measure 2 – Average CT-poverty, weighted by time spent in each census tract across all locations between diagnosis and the date of last contact (Follow-up-Time Weighted Poverty).
- Measure 3 – Overall Average (non-weighted) CT-poverty (Follow-up-Time Average Poverty).
- Measure 4 – Time-varying CT-poverty using all available residential data points (Time-Varying Poverty).

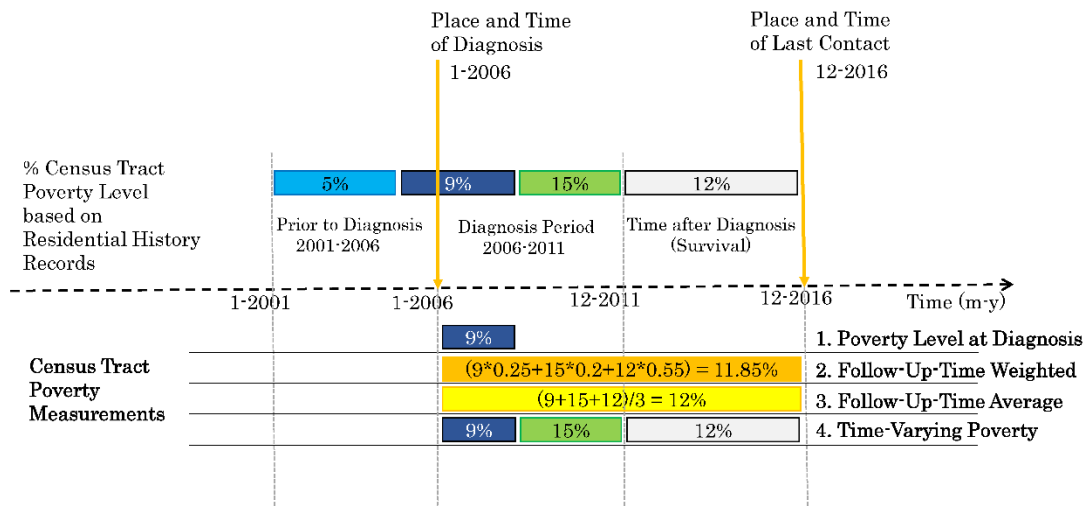


Figure 2: Description of the different CT-poverty measurements used in the analysis.

Statistical Methods

To compare the four poverty measurements, we conducted a Pearson correlation analysis.

Means, medians, ranges, and standard deviations were calculated for each measure. Then,

we examined the changes in CT-poverty over time during the follow-up period, and quantified residential mobility by number of changes in census tract. Patient survival times were calculated in months as the difference between the date of diagnosis and the date of last contact or death. Patients were censored at the date of death from causes other than colon cancer, the date the patient was lost to follow up, or at the end of the follow-up period, 31 December, 2016, whichever occurred first. For measure 4, which included time-varying CT-poverty values, we assigned start and end dates for every census tract, and calculated the survival time spent at each location, until the cumulative time in each location equaled the overall survival time. We defined each CT-poverty measure as either continuous or categorical (<5%, 5% - <10%, 10% - <20%, ≥20%).

Survival analyses were conducted using Cox Proportional Hazard (Cox PH) regression using R package *survsim* (Moriña & Navarro, 2014). For time-varying analysis, we applied method presented by Zhang et al. (Z. Zhang, Reinikainen, Adeleke, Pieterse, & Groothuis-Oudshoorn, 2018). For the models with continuous CT-poverty, we computed hazard ratios (HRs) and 95% confidence intervals (Cis) by exponentiating the parameter estimates. For the models with categorical CT-poverty, we used the lowest poverty category (<5%) as the reference, comparing low poverty census tracts to other categories. The proportional hazards assumption for all covariates in each model was confirmed through the examination of Schoenfeld residuals using the *cox.zph()* function in the R *survsim* (Mills, 2010; Moriña & Navarro, 2014; Z. Zhang et al., 2018). All models met the proportional hazard assumption. We adjusted models for gender, age, race/ethnicity, regional stage subcategories (direct extension only, regional lymph nodes only, direct extension and lymph nodes), and change in census tract (yes/no).

Results

Study Population

Table 2 summarizes the characteristics of the study population (n = 3,949). The average age at diagnosis was 65.8 (SD 13.8). There were fewer males (48%) than females (52%). About 74% of the patients were non-Hispanic White, 12% non-Hispanic Black, 8% Hispanic, 4% Asian/Pacific Islanders, and 2% were other non-Hispanic race. At the time of diagnosis, approximately 29% (26%) of the cases lived in census tracts with $\geq 10\%$ of the population living below the federal poverty level. Approximately 29% of all the patients died from colon cancer by the end of the follow-up period, 31 December, 2016. The mean of CT-poverty at diagnosis was 8.1% (SD 8.2 and similar for both the time-weighted average (8.6%, SD 7.6 and overall average (8.7%, SD 7.5) CT-poverty. The proportion of cases in each of the four CT-poverty categories was similar for both the time-weighted and average CT-poverty measures. The CT-poverty measures based on residence at diagnosis had a higher proportion of cases in the lowest and highest poverty categories compared to the time-weighted and average CT-poverty measures (Table 2).

Table 2: Study population characteristics

Characteristics	Overall (n = 3,949)
Age (years)	
Mean (SD)	65.8 (13.3)
Median [Min, Max]	68.0 [21.0, 85.0]
Gender	
Male, n(%)	1,878 (48)
Female, n(%)	2,071 (52)
Race/Ethnicity	
NH-White, n(%)	2,902 (74)
NH-Black, n(%)	488 (12)
Hispanic, n(%)	325 (8)
Asian/Pacific Islanders, n(%)	141 (4)
Other, n(%)	93 (2)

Table 2 continued

Regional Stage Subcategory	
Regional, direct extension only, n(%)	1,339 (34)
Regional, lymph nodes only, n(%)	1,268 (32)
Regional, both, n(%)	1,342 (34)
Vital status	
Censored, n(%)	2,805 (71)
Colon cancer death, n(%)	1,144 (29)
Survival time (months)	
Mean (SD)	66.2 (38)
Median [Min, Max]	70.0 [1.0, 139]
Census tract changes (“moves”)	
Census Tract at Date of Diagnosis Only, n(%)	2,587 (65)
1 Change in Residential Census Tract, n(%)	732 (19)
2 Changes in Residential Census Tract, n(%)	486 (12)
3+ Changes in Residential Census Tract, n(%)	144 (4)
Census tract-poverty at Diagnosis	
<5% (low), n(%)	1,816 (46)
5% - <10%, n(%)	1,110 (28)
10% - <20%, n(%)	688 (17)
≥20% (high), n(%)	335 (9)
Time-weighted census tract-poverty	
<5% (low), n(%)	1,554 (39)
5% - <10%, n(%)	1,267 (32)
10% - <20%, n(%)	814 (21)
≥20% (high), n(%)	314 (8.)
Average census tract-poverty	
<5% (low), n(%)	1,520 (39)
5% - <10%, n(%)	1,268 (32)
10% - <20%, n(%)	860 (21)
≥20% (high), n(%)	301 (8)
Census tract-poverty at diagnosis	
Mean % (SD)	8.1 (8.2)
Median [Min, Max]	5.5 [0.0, 62.9]
Time-weighted census tract-poverty	
Mean % (SD)	8.6 (7.6)
Median [Min, Max]	6.1 [0.0, 59.9]
Average census tract-poverty	
Mean % (SD)	8.7 (7.5)
Median [Min, Max]	6.3 [0.0, 59.9]

Note: NH=non-Hispanic; SD=Standard Deviation; Censored=alive at last known contact, alive at end of follow-up, death due to other causes

Mean follow-up (survival) time was 66.2 months (range 1-139). Sixty-five percent remained at the same census tract since diagnosis (non-movers); 19% changed census tract once, and 16% changed two or more times. Among the cases changing tracts one or more times 68% saw an absolute change in CT-poverty of less than 10 percentage points during the follow-up period. Comparing the categorical CT-poverty at diagnosis measure with the average CT-poverty measures based on residential histories, only about 14% would be classified into different poverty groups. The largest change in classification among the CT-poverty at diagnosis measure was from the low poverty category (Table 1). Cases living in census tract with the highest poverty rates at diagnosis changed their residential census tract more often (42%) compared to cases living in tracts with the lowest poverty rates (32%). Compared to Non-Hispanic Whites (32%), a higher proportion of Non-Hispanic Blacks (44%), Asian/Pacific Islanders (43%), and Hispanics (38%) changed their census tract during the follow-up period. (Table 3).

Table 3: Distribution of patient characteristics by the number of changes in residential census tracts

Characteristic	n	Census Tract at Date of Diagnosis Only, % ^b	1 Change in Residential Census Tract, % ^c	2 Changes in Residential Census Tract, % ^c	3+ Changes in Residential Census Tract, % ^c
Gender					
Male	1878	64	19	13	3
Female	2071	67	18	11	4
Race/Ethnicity					
NH-White	2902	68	17	12	3
NH-Black	488	56	22	14	8
Asian and Pacific Islanders	325	57	22	16	5
Hispanic	141	62	23	11	4
NH-Other	93	68	17	14	1

Table 3 continued

Census Tract Poverty at time of Diagnosis ^a	n	Census Tract at Date of Diagnosis Only, % ^b	1 Change in Residential Census Tract, % ^c	2 Changes in Residential Census Tract, % ^c	3+ Changes in Residential Census Tract, % ^c
< 5%	1816	68	17	12	3
5%-< 10%	1110	66	20	11	3
10%-< 20%	688	62	21	14	3
≥ 20%	335	57	21	16	6

^a Census tract poverty was classified based on the percentage of the population in the census tract living below poverty

^b Colon cancer cases remained in the same census tract from diagnosis throughout the follow-up period

^c Colon cancer cases changed census tracts n times since diagnosis during the follow-up period

NH=non-Hispanic

Colon Cancer Deaths

Across the four unique CT-poverty measures, the HRs did not vary substantially, ranging from only 1.006 to 1.009 (Table 4). The only model with an association between the continuous CT-poverty measures and risk of death from regional-stage colon cancer was the model that included CT-poverty as a time-varying measurement (HR 1.009; 95% CI 1.001-1.017). In the models with categorical time-weighted and average CT-poverty measures, none of the categories expressed substantial differences in hazard ratios. However, the model that included the time-varying CT-poverty measure as a categorical variable indicated a monotonic increase in risk with increasing CT-poverty across the categories, with one exception. After adjusting for change in CT, cases in the highest CT-poverty category (>20%) had a 30% greater risk of colon cancer death compared to cases in the lowest CT-poverty category (≤5%) (HR 1.30; 95% CI 1.04-1.63). The model that included the CT-poverty measure based on CT at time of diagnosis did not show a monotonic increase in risk across the CT-poverty categories, but it did indicate a statistically significant risk among regional-stage colon cancer in the second to highest CT-poverty category (10% - <20%) compared cases in the lowest CT-poverty category

($\leq 5\%$) (HR 1.33; 95% CI 1.12-1.57). This remained unchanged after controlling for move status. Finally, based on models with continuous measures of CT-poverty, we plotted the time-varying CT-poverty adjusted HRs against the adjusted HRs of the average CT-poverty measure and found no substantial differences between the two values.

Table 4: Comparison of HRs for different CT-poverty measures estimated from the Cox regression models.

Model Type	CT-Poverty at Date of Diagnosis	Weighted CT-Poverty	Average CT-Poverty	Time-Varying CT-poverty
	HR(95% CI)	HR(95% CI)	HR(95% CI)	HR(95% CI)
CT-Poverty (continuous) ^a	1.005 (0.997-1.012)	1.003(0.994-1.011)	1.002(0.993-1.010)	1.007(1.010-1.015)
CT-Poverty (continuous) ^b	1.006 (0.999-1.014)	1.006 (0.998-1.014)	1.006 (0.997-1.014)	1.009 (1.001-1.017)
CT-Poverty (categorical) ^c				
<5%	Ref.	Ref.	Ref.	Ref.
5% - <10%	1.07(0.93-1.24)	0.91(0.79-1.05)	0.91(0.79-1.06)	1.06 (0.92-1.22)
10% - <20%	1.32(1.12-1.56)	1.03(0.87-1.22)	1.00(0.85-1.18)	1.13 (0.96-1.33)
$\geq 20\%$	1.07(0.84-1.36)	1.03(0.80-1.31)	1.01(0.79-1.30)	1.21 (0.97-1.51)
CT-Poverty (categorical) ^d				
< 5%	Ref.	Ref.	Ref.	Ref.
5% - <10%	1.07 (0.93-1.25)	0.97 (0.84-1.12)	0.98 (0.85-1.13)	1.09 (0.95 -1.26)
10% - <20%	1.33 (1.12-1.57)	1.15 (0.98-1.37)	1.16 (0.98-1.37)	1.17 (0.99-1.38)
$\geq 20\%$	1.12 (0.88-1.42)	1.12 (0.86-1.44)	1.11 (0.86-1.43)	1.30 (1.01-1.63)

^a model adjusted for age (categorical), sex, regional stage sub-categories, race/ethnicity, CT-poverty (continuous)

^b model adjusted for age (categorical), sex, regional stage sub-categories, race/ethnicity, changed CT (yes, no), CT-poverty (continuous)

^c model adjusted for age (categorical), sex, regional stage sub-categories, race/ethnicity, CT-poverty (categorical)

^d model adjusted for age (categorical), sex, regional stage sub-categories, race/ethnicity, changed CT (yes, no), CT-poverty (categorical)

Discussion

Previous studies that have utilized cancer registry data to examine cancer survival in relation to neighborhood socioeconomic measures have used a single location based on the residence location at time of diagnosis to assign the socioeconomic measures (Feinglass et al., 2015; 2015; Henry et al., 2009; Lian et al., 2011; Warner & Gomez,

2010; D. Wiese et al., 2019). Our study used a unique dataset of residential histories linked to population-based cancer registry data to examine whether regional stage colon cancer survival estimates would differ if area-based census tract-level-poverty were defined using residential histories versus the standard approach of using only the residence at diagnosis. This study contributes to our understanding of the residential mobility of colon cancer patients after diagnosis, and the impact of including residential histories when examining the association between colon cancer survival and neighborhood SES. Overall, we found that the HRs were generally similar in strength and direction regardless of whether the four CT-poverty measures were included as a continuous or as a categorical variable. However, models using an overall average or time-weighted average of CT-poverty failed to detect differences in the risk of colon cancer death. The reason for this finding is a result of averaging CT-poverty across the residential locations, which reduced the variance of these two measurements resulting in measurements that have less predictive power. Time-varying poverty models that maximized the use of all available data points for changing CT-poverty over time did, however, yield more precise estimates than the models using only the residence at diagnosis. We found that, after controlling for individual-level factors such as age, race/ethnicity, gender, and change in census tract, the time-varying model indicated that individuals living in high poverty areas (>20%) were 30% more likely to die from regional stage colon cancer compared to individuals in low poverty areas (< 5%). This finding is consistent with previous studies that used only residence at the diagnosis (Chien et al., 2013; Henry et al., 2009; Lian et al., 2011). The risk of death in this high-poverty group only increased in *post-hoc* stratified analysis when we limited the sample

to movers only (HR 1.62 95% CI 1.07-2.44). Thus, a possible association would have been missed without the utilization of time-varying data, suggesting that the use of residential history data warrants further study.

For cancer disparities research examining patient survival outcomes by neighborhood socioeconomic status, the relative importance of including residential histories versus a single location to reduce potential measurement bias (Oudin, Forsberg, Strömngren, Beelen, & Modig, 2012) will depend ultimately on factors such as population mobility (Geronimus, Bound, & Ro, 2014), the temporal and spatial patterns of the neighborhood measure of interest (Oudin et al., 2012) (e.g., pollution, segregation), and length of follow-up time. Our study population was mobile, with approximately 35% changing census tract one or more times after initial diagnosis. However, the majority who changed census tract moved to tracts with similar poverty levels, thus limiting any significant sources of measurement error between the models that included CT-poverty measured at diagnosis and the models that measured CT-poverty based on residential histories. The more precise findings for the models that included CT-poverty as a time-varying factor are likely a result of the greater statistical power from more data points directly incorporated from each residential CT location during the follow-up period.

More research is needed on the value of using residential histories versus a single address to assign neighborhood measures. Finding no major differences between the use of a single address versus residential histories in our study is likely a result of minimal socio-spatial mobility (i.e cases moving to tracts with similar poverty levels). For study populations with greater socio-spatial mobility or followed for longer periods of time it would be worthwhile using residential histories to avoid any type of exposure bias. Using

residential histories also provides researchers new avenues of inquiry to examine neighborhood SES effects that might not remain constant over time. Additionally, future work should consider census tract variables with different spatial and temporal patterns other than poverty, such as segregation, ethnic enclaves, and homeownership. There could be alternative measures that could reveal a clearer association with cancer survival. To our knowledge, this study is among the first to examine patient survival by neighborhood poverty utilizing statewide cancer registry data linked to a commercial database of residential histories. Linking cancer cases from a statewide cancer registry to commercial databases of residential histories opens up new and important avenues of inquiry to better understand the complex relationships between people and places, and residential mobility and cancer outcomes. Residential mobility has generally been regarded as a negative stressful event (Jelleyman & Spencer, 2008), while increased mobility is associated with poor health outcomes. To further understand the association between mobility and cancer outcomes, future studies using residential histories linked to cancer data should examine the trajectories of neighborhood socioeconomic status (e.g., upward, downward, stable high, and stable low) in relation to health outcomes similar to previous studies (D. Zhang et al., 2019). Residential histories can also provide great insight in studies where the actual distance to health services (e.g., colonoscopy providers) or food environments is important. For example, Richardson et al. used residential histories from participants in the U.S.-based Coronary Artery Risk Development in Young Adults study to examine longitudinal access to fast food and non-fast food restaurants, supermarkets, and convenience stores (Andrea S Richardson et al., 2014). Residential histories linked to cancer data can also be used to help answer

important questions about the role selective residential mobility plays in geographic patterns of cancer outcomes (W. A. Clark, Van Ham, & Coulter, 2014). Finally, spatial models that directly include geographic locations can use residential histories to more precisely estimate spatial risk of death estimates (Henry et al., 2009; D. Wiese et al., 2019).

This study contributes to research on the uncertain geographic context problem, which arises from the lack of knowledge about the precise spatial and temporal configuration of the physical and social factors that exert influence on the individual (Kwan, 2012). Cancer registries routinely collect only the cases location at time of diagnosis which is used in cancer surveillance and cancer disparities research to characterize neighborhood characteristics (e.g. poverty, pollution, segregation) or access to care (e.g. travel time to colonoscopy provider or diagnosing hospital). In the case of survival analysis the use of a single location at time of diagnosis creates uncertainty about patients' location and associated exposures during the follow-up period (J. R. Meliker & Sloan, 2011). When analysis is conducted using location at diagnosis, we assume that cases are not mobile and neighborhood measures like CT-poverty do not change over time. In this study we were able to explore differing temporal configurations of neighborhood poverty linked to cases. While we found minimal differences in survival estimates by CT-poverty when we included residential histories more research is needed to assess how geographic uncertainties about residential locations over time might impact model estimates especially when considering different neighborhood variables, populations, and when spatial location is directly included in the models.

There are several limitations to be noted. The study population was limited to New Jersey residents and, therefore, the results may not be generalizable to geographies with different demographic and socioeconomic profiles. Another limitation of our study is that we only followed patients over 10 years. We also restricted our study to regional stage colon cancer cases in order to minimize extreme variations in survival and limit sources of variation to residential history measures. Our findings may not be generalizable to colon cancer cases diagnosed in earlier (local) or later (metastatic) stages of the disease. It is possible that residential histories among local-stage colon cancer patients may play a more important role when evaluating survival among a group of patients who are expected to live longer and who may be at greater risk for subsequent cancers and other long-term effects. However, residential histories might also be informative for patients with metastatic disease given prior exposures to neighborhood SES are accounted for, particularly earlier in the disease process when access to care or other health services could have impacted being diagnosed with late versus early stage disease. Further, residential history may be useful and informative for elucidating health disparities because it helps account for residential mobility after diagnosis, given that access and continuity of care can be affected (Baugh & Verghese, 2013; Muralidhar et al., 2016).

The study also relied solely on residential histories collected from LexisNexis and did not include self-reported information that may be used to validate and/or augment LexisNexis data. However, previous studies found good concordance between LexisNexis addresses and addresses collected from study participants (85%-86%) (S. Hurley et al., 2017; D. C. Wheeler & Wang, 2015). In our study, concordance between the LexisNexis locations and the locations reported by the NJSCR was around 83% and

increased substantially to about 93% when comparisons were limited to a 6-month window before and after the diagnosis date. A small proportion (8%) of the locations from LexisNexis cases did not match any locations reported by the NJSCR. Discordant addresses may be a result of a number of factors including incorrect links at LexisNexis, incorrect geocodes assigned by NJSCR for both registry and LexisNexis residential addresses, incorrect addresses reported to the registry by hospitals and other reporting facilities, or geocodes assigned to addresses based on post office boxes. Although these issues could have contributed to non-differential misclassification, the extent of the bias in either direction would be minimal due to the low proportion of cases affected. On the other hand, LexisNexis is an objective source of residential history data that has been collected and maintained electronically over time and is not subject to other sources of bias that are known challenges with self-reported data (e.g., recall bias). Last, we did not have access to individual-level poverty or other socioeconomic measures. Individual-level poverty and socioeconomic status has been shown to play a substantial role in determining colon cancer survival outcomes (Aarts et al., 2010; Lian et al., 2011; Robinson, 2019).

Conclusion

In our head-to-head comparisons of survival estimates based on CT-poverty measured as a single location at diagnosis versus a series of time-varying longitudinal residential histories, we documented adjusted HRs that were generally similar in strength and direction. Our findings suggest that residential changes after regional-stage colon cancer diagnosis may be associated with a higher risk of colon cancer death among individuals

who live in high poverty areas. This has important clinical and public health implications in regards to post-diagnostic access to care for treatment and follow-up surveillance.

CHAPTER 3:
RESIDENTIAL MOBILITY AND GEOSPATIAL DISPARITIES
IN COLON CANCER SURVIVAL.

Abstract

Identifying geospatial cancer survival disparities is critical to focus interventions and prioritize efforts with limited resources. Incorporating residential mobility into spatial models may result in different geographic patterns of survival compared to the standard approach using a single location based on the patient's residence at the time of diagnosis. Data on 3,949 regional-stage colon cancer cases diagnosed from 2006-2011 and followed until December 31, 2016 were obtained from the New Jersey State Cancer Registry. Geographic disparity based on the spatial variance and effect sizes from a Bayesian spatial model using residence at diagnosis was compared with a time-varying spatial model using residential histories (adjusted for sex, gender, sub-stage, race/ethnicity and census tract (CT) poverty). Geographic estimates of risk of colon cancer death were mapped.

Most patients (65%) remained at the same residence, 22% changed CT and 12% moved out of state. The time-varying model produced a wider range of adjusted risk of colon cancer death (0.85-1.20 vs. 0.94-1.11) and resulted in greater geographic disparity statewide after adjustment (25.5% vs. 14.2%) compared to the model with only the residence at diagnosis.

Including residential mobility may allow for more precise estimates of spatial risk of death. Results based on the traditional approach using only residence at diagnosis were not substantially different for regional stage colon cancer in New Jersey.

Including residential histories opens up new avenues of inquiry to better understand the complex relationships between people and places, and the effect of residential mobility on cancer outcomes.

This chapter was published in present form in the *Cancer Epidemiology, Biomarkers and Prevention* journal.

Full citation:

Wiese, D., Stroup, A. M., Maiti, A., Harris, G., Lynch, S. M., Vucetic, S., & Henry, K. A. (2020). Residential mobility and geospatial disparities in colon cancer survival. *Cancer Epidemiology, Biomarkers and Prevention*, 29(11), 2119-2125. doi:10.1158/1055-9965.epi-20-0772

Introduction

Identifying cancer survival disparities is critical to focus interventions in specific populations and for clinical and cancer control programs to prioritize efforts with limited resources. Examining whether cancer survival varies geographically can also generate hypotheses about the underlying causes of survival disparities and help identify important demographic and neighborhood factors that play a role. However, it is not known whether incorporating the residential mobility after a cancer diagnosis into spatial models results in different geographic patterns of survival compared to the standard approach of using a single location based on the patient's residence at the time of diagnosis.

Numerous studies have examined associations between cancer survival and neighborhood-based socioeconomic measures, such as poverty. These studies have shown that patient survival is generally worse among patients in poor neighborhoods compared to patients living in more affluent neighborhoods (Feinglass et al., 2015; 2015; Henry et al., 2009; Lian et al., 2011; Warner & Gomez, 2010). Studies specifically examining spatial patterns in survival using geographic methods, such as spatial scan statistics (Henry et al., 2009; Niu, Roche, Pawlish, & Henry, 2013) and spatial models (Chien et al., 2013; Lian et al., 2011; D. Wiese et al., 2019), have also noted worse survival clustered in poor communities. To date, studies examining geographic disparities in cancer survival used the patient's residence at the time of diagnosis exclusively to map geographic distributions of patient survival or to define neighborhood characteristics. Limitations of using a single location have been well documented (Francis P. Boscoe, Ward, & Reynolds, 2004; G. M. Jacquez et al., 2005; M. Schootman et al., 2017; Tong, 2000; David C. Wheeler, Ward, & Waller, 2012), including potentially introducing

spatial uncertainty because it assumes the person remained at the same residence and in turn experienced uniform neighborhood-based contexts over time (e.g. unchanging CT-poverty). This may be a potential source of bias in non-spatial and spatial statistical models, calling into question both the accuracy and precision of survival estimates associated with geographic factors. The magnitude and direction of bias largely depends on the volume of residential mobility in the study population, the degree to which mobility results in changes in neighborhood context (e.g., higher or lower poverty rates), and whether the neighborhoods themselves are changing overtime (e.g., gentrification). If a large proportion of cases are moving to neighborhoods with similar characteristics, then bias would be minimal. Failure to account for residential socio-spatial mobility may also impede researchers' ability to account for self-selection into neighborhoods, which may change the overall risk of death in a neighborhood over time (Bergström & Van Ham, 2010; W. A. Clark et al., 2014).

The main reason researchers have not incorporated residential mobility into geographic studies of cancer survival or other outcomes (e.g., stage at diagnosis) is that cancer registries do not collect residential histories. The primary source of location data for cancer patients collected by registries is the residential address at the time of diagnosis. However, recent studies have demonstrated the potential utility of residential histories collected from sources such as electronic medical records (S. Hurley et al., 2017) and commercial and administrative databases to document residential mobility (Susan Hurley et al., 2017; G. M. Jacquez et al., 2011; Stinchcomb & Roeser, 2016; D. C. Wheeler & Wang, 2015). There are numerous examples of applying residential histories in environmental epidemiology for assessing longitudinal exposures (Brokamp et al., 2016;

Gallagher, Webster, Aschengrau, & Vieira, 2010; Hughes & Pruitt, 2017; Hystad et al., 2013; Jaymie R. Meliker et al., 2010) and risk and geographic clustering of cancers (G. M. Jacquez et al., 2006; J. R. Meliker & Jacquez, 2007; Nordsborg et al., 2014; Nordsborg et al., 2015; Sloan, Nordsborg, Jacquez, Raaschou-Nielsen, & Meliker, 2015), but far less use of residential histories to examine disparities in cancer outcomes, such as survival (D. Wiese, Stroup, A.M., Maiti, A., Harris, G., Lynch, S.M., Vucetic, S., Henry, K.A., 2020; D. Zhang et al., 2019). To our knowledge, residential histories have not been used by population-based cancer registries to examine geographic variation in cancer survival.

To address the uncertain geographic context problem as it relates to place-based disparities research (Sahar et al., 2019) and evaluate how uncertainty in a patient's residential location after diagnosis might impact survival estimates and geographic patterns, we incorporated residential histories into Bayesian spatial models to identify whether the survival of patients diagnosed with regional stage colon cancer in New Jersey varied geographically after adjusting for key individual-level risk factors and neighborhood poverty level. We compared differences in geographic variation in colon cancer survival between models using residence at the time of diagnosis alone to models that incorporated residential mobility, accounting for whether the patients moved out of state after diagnosis.

Materials and Methods

Study Population

Colon cancer cases were obtained from the New Jersey State Cancer Registry (NJSCR), which is a population-based cancer registry established in October 1978 to monitor

cancer among the more than 8.9 million residents of New Jersey (NJSCR, 2018). The study population includes all New Jersey residents 18 years and older diagnosed between January 1, 2006 and December 31, 2011 with histologically confirmed, first primary regional stage colon cancer as defined according to the International Classification of Diseases for Oncology, 3rd Edition (ICD-O3 C180-C189, C260; excluding histology codes 9050-9055, 9140, 9590-9992) (Percy et al., 1990) (N = 4,041). The Rutgers University institutional review board (IRB) approved all study activities.

Regional stage was defined according to Surveillance, Epidemiology and End Results (SEER) derived summary stage 2000 and includes regional, direct extension only, regional lymph nodes only, and regional, direct extension and regional lymph nodes. This definition includes AJCC 6th Edition (Greene et al., 2002) groupings IIA, IIB, IIIA, IIIB and IIC. The study sample was limited to regional stage colon cancers in order to control the influence that stage will have on the survival estimates. Doing so allows for ease of interpretation of the geographic disparities that may otherwise be attributable to spatial variation in stage.

Individual-level factors included age at diagnosis, gender (male, female), and race/ethnicity [Non-Hispanic (NH) White, NH Black, NH Asian/Pacific Islander (API), NH Other, and Hispanic (any race)]. Cases with any stage at diagnosis other than regional stage or had no survival time (i.e., only ascertained through death certificates or autopsy) were excluded from analysis (American Cancer Society, 2019). Vital status, including date of death and cause of death (if deceased) or date of last contact (if alive) is updated routinely by the NJSCR through linkages with state and national sources including death data from the New Jersey Department of Health Office of Vital Statistics and Registry

and the National Death Index, hospital discharge files, Centers for Medicare and Medicaid Services, Social Security Administration Services for Epidemiologic Researchers, and motor vehicle registration files. Cases were followed until their deaths or until December 31, 2016. Deaths attributed to colon cancer were coded based on ICD-10 code C18 (Percy et al., 1990).

Residential History Data from LexisNexis

Residential histories were acquired through a data linkage with a commercial database developed by LexisNexis, Inc. (Miamisburg, Ohio, US). LexisNexis (<https://www.lexisnexis.com/en-us/products/public-records.page>) has developed a specific resource for researchers to obtain residential histories for adults aged 18 and older (Susan Hurley et al., 2017; Geoffrey M Jacquez et al., 2010; Stinchcomb & Roeser, 2016; D. C. Wheeler & Wang, 2015). Up to 20 of the most recent addresses between 1946 and 2018 with documented start and stop dates were returned for each case. All residential addresses were geocoded to the 2010 census tract (CT) boundaries using the North American Association of Central Cancer Registries (NAACCR) AGGIE Geocoder (Texas A&M, 2016).

Of the 4,041 regional stage colon cancer cases, 98 (2.4%) had no residential information available. For the remaining 3,949 (97.6%), we applied a preprocessing technique described previously (D. Wiese, Stroup, A.M., Maiti, A., Harris, G., Lynch, S.M., Vucetic, S., Henry, K.A., 2020).

Neighborhood Socioeconomic Status Data

Similar to other studies (F.P. Boscoe, 2010; Henry et al., 2009), we used the percentage of the census tract population 18 years and older living below the federal poverty level as a measure of neighborhood deprivation (CT-poverty). CT-poverty was obtained from publicly available U.S. Census and American Community Survey (ACS) data. U.S. Census 2010 and the ACS 5-year average data 2006-2010, 2007-2011, 2008-2012, 2009-2013, and 2010-2014 were used for residencies between 2006-2010. ACS 2011-2015, 2012-2016, and 2013-2017 were used for residencies between 2011 and 2016. For each case, every residential record received a corresponding CT-poverty value based on the earliest date of the residential appearance in the data set. In situations when cases remained at a single census tract over multiple years, we assigned the annual CT-poverty values to capture changes within the neighborhood (e.g., changes due to gentrification). We developed two different poverty measurements. The first measurement used the standard CT-poverty from the location at the time of diagnosis (DX CT-poverty). For the second, all corresponding census tracts were included; therefore, each patient could have multiple New Jersey census tracts during the follow-up period (time-varying CT-poverty).

Statistical Methods

Patient survival times were calculated in months as the difference between the date of diagnosis and the date of last contact or death. Patients were censored at the date of death if they died from causes other than colon cancer, the date the patient was lost to follow-up, or at the end of the follow-up period (December 31, 2016), whichever occurred first.

In the time-varying model, cases were also censored at the time they moved out of the State of New Jersey; and, assigned start and end dates for every census tract location, and calculated the survival time spent at each census tract. In other words, start and end dates set the follow-up intervals and corresponding covariate values (e.g., CT-poverty), as well as the vital status at the end of every interval (1 = dead, 0 = alive).

Bayesian geospatial models were applied to survival time as an extension of conventional Cox regression survival models described by Kneib and Fahrmeir (Kneib & Fahrmeir, 2007). This model includes the geographic location of the patient's census tract, which is used as the spatial function and estimates the spatial effect. The spatial function provides a way to estimate the geographic variation in the risk of death after controlling for individual- and neighborhood-level covariates (Brezger & Lang, 2006). It measures the instantaneous event rate or the probability that an individual would experience an event (e.g., death from *regional stage colon cancer*) after diagnosis. The spatial function is based on stationary Gaussian random fields (bivariate penalized splines). P-spline smoothed spatial effects were incorporated into the model based on an adjacency matrix of geographic neighbors by census tracts (weights based on rook's case) (Adler, Kneib, Lang, Umlauf, & Zeileis, 2012; Belitz et al., 2015; Lang & Brezger, 2004). The risk of death is the exponentiated smoothed posterior mean for the census tract based on those living in that census tract. For the standard DX CT-poverty model, the spatial effect was estimated using the census tract from the time of the diagnosis. To implement the time-varying CT-poverty model, we redesigned the model as suggested by Belitz et al. (Belitz et al., 2015) for time-dependent covariates, while including only New Jersey census tracts.

Estimation of regression models is based on Markov chain Monte Carlo simulation techniques, corresponding to full Bayesian inference, and obtained by specifying prior distributions for all unknown parameters. For each model, 10,000 iterations were run, with the first 2,000 samples used as a burn-in. Every 20th sample from the remaining 8,000 samples was saved and used to construct the posterior distribution for each of the parameter estimates in the model. The 95% confidence intervals (CI) were calculated based on the posterior distribution of the 1,000 samples to identify significant risk of death and the CTs with significantly higher or lower estimates than the state average (< 1 lower risk, > 1 higher risk). All models were implemented with R using BayesX (Nikolaus Umlauf, Adler, et al., 2012), BayesXsrc (Adler et al., 2012), and R2BayesX (Nikolaus Umlauf, Kneib, Heinzl, Lang, & Zeileis, 2013; Nikolaus Umlauf, Kneib, Lang, Zeileis, & Umlauf, 2012) packages. The exponentiated spatial effects of each CT from the models were mapped to visualize the risk of death from regional stage colon cancer. Then, we calculated a difference map, highlighting differences between the risk of death estimates using CT-poverty from diagnosis and time-varying values. Additionally, we compared models using geographic disparity (GD) percentage, a method originally proposed by Chien et al. (Chien et al., 2013), to assess the geographic variance in regional stage colon cancer survival after accounting for DX CT-poverty and time-varying CT-poverty. It is calculated as the square root of the spatial variance, where higher GD suggests wider geographic variability unexplained by independent variables, and assumes larger geographic disparities in the study area. We also compared model's fit using Deviance Information Criterion (DIC). DIC is defined as the sum of the posterior expected

deviance and the effective number of parameters. A better model fit is assumed with a lower value of DIC (Jin, Carlin, & Banerjee, 2005).

Results

Study Population

Table 5 summarizes the characteristics of the study population (n=3,949). The average age at diagnosis was 66. There were fewer males (47.6%) than females (52.4%). Around three quarters (73.6%) of the study population were NH-White, 12.4% NH-Black, 8.1% Hispanic origin (any race), 3.6% NH-API, and 2.4% Other race. Approximately 27.5% of all the patients died from the colon cancer by the end of follow-up, with a median survival of 66 months (range 1-139).

During the follow-up period, 65.5% remained at their diagnosis census tract, 22.4% changed CTs within New Jersey, and 12.1% left New Jersey during the study period. Among those who moved, 18.5% only moved once, 12.3% moved twice, and 3.6% moved 3 or more times after cancer diagnosis. The average time spent at the census tract at diagnosis was 7.5 years. At the time of cancer diagnosis, nearly half (46%) lived in census tracts with <5% poverty and 25.9% lived in census tracts with poverty rates 10% or higher.

Table 5: Study population characteristics

Characteristics	Overall (n = 3,949)
Age	
Mean (SD)	65.8 (13.3)
Median [Min, Max]	68.0 [21.0, 85.0]
Gender	
Male	1,878 (47.6%)
Female	2,071 (52.4%)

Table 5 continued

Race/Ethnicity	
NH-White	2,902 (73.5%)
NH-Black	488 (12.4%)
Hispanic (any race)	325 (8.2%)
NH-API	141 (3.6%)
Other	93 (2.4%)
Regional Stage Subcategory	
Regional, direct extension only	1,339 (33.9%)
Regional, lymph nodes only	1,268 (32.1%)
Regional, both	1,342 (34.0%)
Vital Status	
Censored	2,862 (72.5%)
Colon Cancer Death	1,087 (27.5%)
Survival Time (months)	
Mean (SD)	62.3 (38.0)
Median [Min, Max]	66.0 [1.00, 139]
Census Tract Changes (Type of “moves”)	
CT at Date of Diagnosis Only	2,587 (65.5%)
Change in Residential CT within NJ	885 (22.4%)
Change in Residential CT outside NJ	477 (12.1%)
Census Tract Changes (Number of “moves”)	
Single residency	2, 587 (65.5%)
1 Change in Residential CT	732 (18.5%)
2 Changes in Residential CT	486 (12.3%)
3+ Changes in Residential CT	144 (3.6%)
CT-poverty at Diagnosis	
<5% (low)	1,816 (46.0%)
5% - <10%	1,110 (28.1%)
10% - <20%	688 (17.4%)
≥20% (high)	335 (8.5%)

Note: CT = Census Tract; NJ = New Jersey

SD = Standard Deviation; NH = Non-Hispanic; API= Asian/Pacific Islanders

Geographic Clustering and Spatial Effects

Table 6 summarizes the statewide range of risk of death and geographic disparities percentage for each of the spatial models (**Model a** using DX CT-poverty and **Model b** using time-varying CT-poverty). The DX CT-poverty model produced fewer CT-specific

risk of death estimates (N = 1,604) compared to the time-varying CT-poverty model with residential histories (N = 1,742). The time-varying model includes all the census tract locations where the cases lived, resulting in more census tracts where risk of death from regional stage colon cancer can be estimated. The adjusted risk estimates in Model a ranged from 0.94-1.11 – a 0.17 point range from low to high based on survival data for 1,604 CTs. Whereas, the adjusted risk estimates in Model b ranged from 0.83-1.2, which is a 0.37 point range from low to high based on survival data for 1,742 CTs. The GD percentage was 14.2% and 25.5% in Model a and Model b, respectively. The DIC indicated a better fit in Model a (14,269) compared to Model b (16,489). Fixed effects of CT-poverty were similar between the models and not significant at $p < 0.05$ level.

Table 6: Model Comparison

Models	Number of CTs	Range of Adjusted Risks of Death	GD	DIC	CT-Poverty Fixed Effect (95% CI)
Model a (DX CT-poverty)*	1,604	0.94-1.11	14.2%	14,269	1.003 (0.996-1.011)
Model b (Time-Varying CT-poverty)**	1,742	0.83-1.2	25.5%	16,489	1.003 (0.994-1.012)

Note:

* Adjusted for Sex, Age, Regional Sub-stage at Diagnosis, Race/Ethnicity, and CT-poverty at CT of Diagnosis.

** Adjusted for Sex, Age, Regional Sub-stage at Diagnosis, Race/Ethnicity, Change in CT (Mover Status), and Time-Varying CT-poverty using residential histories.

GD = Geographic Disparities; DIC = Deviance Information Criterion; CI = Confidence Interval

The adjusted risk of death from colon cancer from each model were mapped (Figure 3) to compare geographic distribution of adjusted risk estimates from Model a (Figure 3a) and Model b (Figure 3b). For both models, there were no statistically significant areas of higher or lower risk of colon cancer death. However, risk estimates for the southern and northwestern regions as well as the northeastern metropolitan area that borders New

York City clearly become more pronounced when residential mobility is included in the model. For example, in the southern region, the risk of colon cancer death ranged from 1.02-1.05 in the DX CT-poverty model and increased to 1.08-1.2 in the time-varying CT-poverty model. In the northeastern region, the area with the highest risk expands to a larger geographic area around the Newark metropolitan area. Similarly, in the northwestern region, areas with the lowest risk expands geographically. The range of risks in the lowest and upper most categories, however, widen from 0.93-0.95 (Figure 3a) to 0.83-0.95 (Figure 3b) and 1.08-1.11 (Figure 3a) to 1.08-1.2 (Figure 3b).

Figure 1c summarizes the difference in the risk of death estimates between the adjusted DX CT-poverty model (3a) and the adjusted time-varying CT-poverty model (3b). The negative values indicate areas where the risk of death estimates from Model b were higher than the risk estimates from Model a (DX CT-poverty < time-varying CT-poverty). Whereas, positive values indicate areas where the risk of death estimates from Model b were lower than the risk estimates from Model a (DX CT-poverty > time-varying CT-poverty). Overall, the range of differences in risk of death estimates were small, ranging from -0.14 and +0.10. There were two areas in North Jersey where the estimates in Model a and Model b change direction. In the area around parts of Somerset County, noted in the pink solid boundary, the risk of colon cancer death changes from low (< 1.0) in Model a to high (> 1.0) in Model b. Whereas, the area around Morris County and along the border of Essex County, in the solid gray boundary, the risk of colon cancer death changes from high (> 1.0) in Model a to low (< 1.0) in Model b.

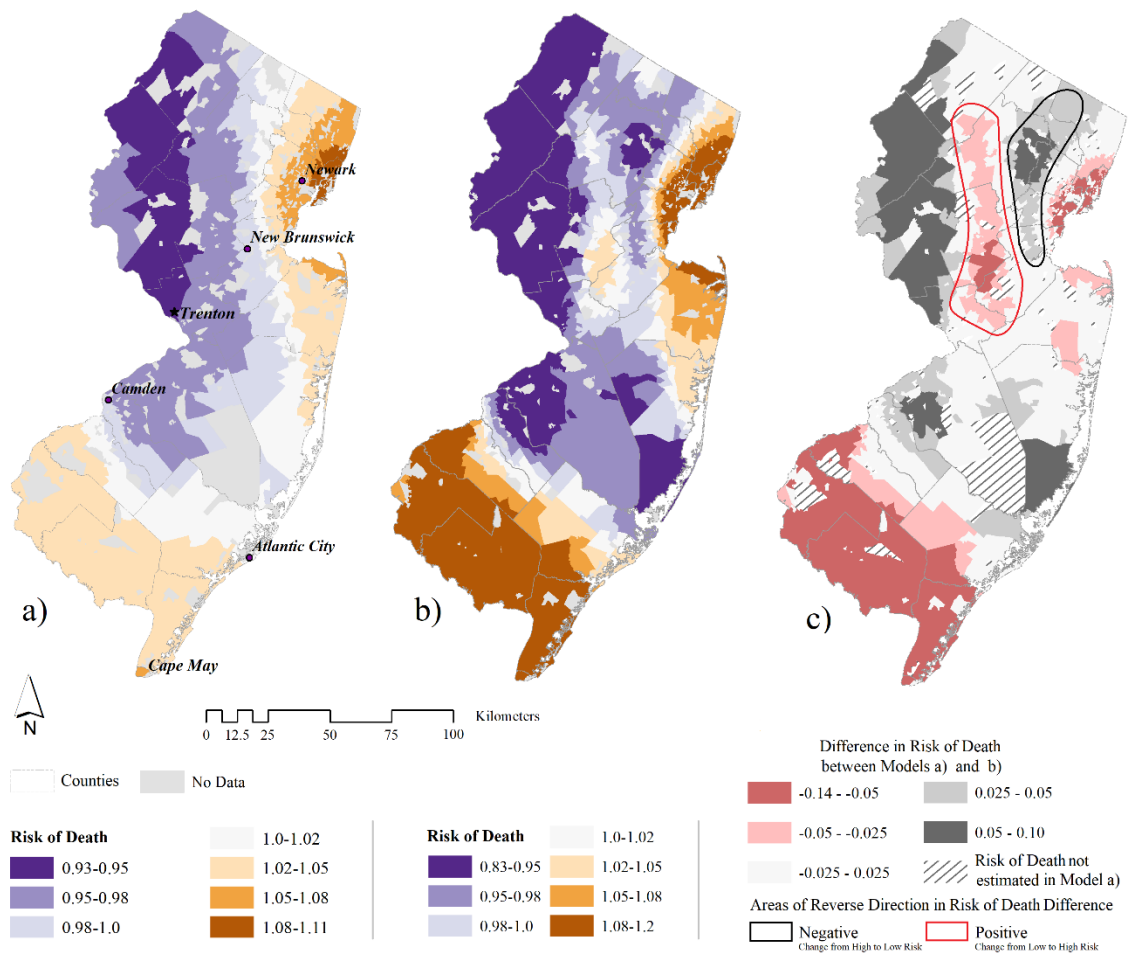


Figure 3: Geographic variation in risk of death estimates after adjustment for sex, age, race/ethnicity, sub-stage at diagnosis, and CT-poverty (a) using information from place of diagnosis and (b) using residential histories during the follow-up. Figure c shows the difference map of risk of death estimates between models a and b ($c = a - b$). Red and black circled areas show the CTs with risk of death differences that were in the opposite direction in the time-varying model compared to the diagnosis model.

Discussion

To our knowledge, this is the first geospatial cancer survival study that combined residential history data with cancer surveillance data from a population-based cancer registry. We used residential history data to characterize changes in neighborhood poverty level post-diagnosis and applied it to geospatial survival models. Cancer disparities research using population-based cancer registry data traditionally use

residential location at the time of diagnosis to assign neighborhood measures of socioeconomic or sociodemographic status and to conduct geographic analysis. The limitation of this approach is that it assumes patients remain at the same location during the follow-up period after diagnosis. This approach contributes to the uncertain geographic context problem, which arises from the lack of knowledge about the appropriate spatial and temporal configuration for assessing the influence of the environment on health outcomes (Kwan, 2012).

In this study, we used spatial modelling techniques to examine geographic variation in regional stage colon cancer survival and focused on whether the results based on the conventional “residence-at-diagnosis” approach to estimate area poverty would differ from results based on a time-varying model that incorporated changing poverty estimates from residential histories. After adjusting for age, gender, race/ethnicity, regional sub-stage, census tract change and CT-poverty, the time-varying geospatial model produced notable differences in the geographic patterns in risk of death as evidence by the increase in the effect sizes of high (South Jersey) and low (Northwest Jersey) risk and the expansion of affected geographic areas (Northeast Jersey). When we examined the spatial variance of each approach, we found that the standard model using poverty level at diagnosis alone explained more of the geographic disparity (i.e. lower spatial variance) than the time-varying model with changing CT-poverty levels; suggesting that standard approaches using residence at diagnosis explained a greater proportion of the geographic disparities in the risk of death from colon cancer in New Jersey. This finding could be a result of the length of the follow-up time, censoring of cases that moved out of the state during the follow-up period, and removing cases from census tracts as they move in and

out of the area over time. Although we followed cases for 11 years, this may not be long enough for residential mobility to have an impact on survival time. It is likely that the use of a single CT-poverty measure captured at the time of diagnosis is sufficient in this population. In the standard time-to-event models, all cases would have contributed survival time in the same census tract until the end of follow-up, death, or lost to follow-up. Whereas, in the time-varying model, survival time for each case will vary by the number of residential locations from diagnosis to the end of follow-up and the length of time in each location. Decreasing both the number of observations and follow-up time within census tracts may add to larger variance (i.e., geographic disparity) and loss in precision (i.e., changing effect sizes).

Our finding that, in some areas in New Jersey, the risk estimates of colon cancer death changed in direction from low risk to high risk or from high risk to low risk after incorporating residential history data requires further research to assess the accuracy of estimates in these geospatial models. However, despite this finding, the regional patterns and effect sizes were not entirely dissimilar between the two models. This is due to the relative stability of individuals post-diagnosis (65.5% remained in the same census tract). There was also minimal socio-spatial mobility as cases often moved to CTs with similar levels of poverty (68% of the cases that changed census tract during the follow-up period had an absolute change in CT-poverty of less than 10 percentage points). Indeed, researchers using either model would have come to the same general conclusion, but future work using simulated and real data should examine the impact of differential socio-spatial mobility on spatial and non-spatial models.

Our study had several strengths. We used high quality, population-based cancer surveillance data from the NJSCR, which conducts follow-up for all cases diagnosed in the State of New Jersey. With an average follow-up rate of 97%, our study population is less prone to bias due to loss to follow-up. We were also able to obtain residential histories on 97.6% of our study population, limiting selection bias. With residential history data, we were able to adjust our models for out migration of patients after diagnosis. Previous studies using population-based cancer registry data do not exclude cases moving out of the study area (Allemani et al., 2017; Niu et al., 2013), which may lead to an underestimation of the true risk of death among New Jersey residents. In this analysis, about 12% of the cases moved out of state before the end of follow-up. The majority of these cases had higher 5-year survival rates compared to those that remained in New Jersey (82% vs 72%); and, were from less impoverished areas (average poverty 7.6% vs 8.1%). Future research should look more carefully at potential biases in statewide survival estimates from not accounting for residential mobility. Our study also produced consistent results with previous studies as regards individual-level factors such as age, gender, race/ethnicity associated with survival outcomes (Du et al., 2007; Henry et al., 2009; Niu et al., 2010; White et al., 2017; White et al., 2010) as well the fixed-effect of area-based poverty as cases living in the high-poverty neighborhoods had a higher risk of death from colon (Chien et al., 2013; Henry et al., 2014) and other cancers (Niu et al., 2010; Mario Schootman, Jeffe, Lian, Gillanders, & Aft, 2009; Shariff-Marco et al., 2015; Shariff-Marco et al., 2014; D. Wiese et al., 2019).

Our study was limited to New Jersey residents diagnosed with regional stage colon cancer from 2006-2011, and may not be generalizable to other state populations with

different geographic, socioeconomic, and demographic profiles, nor to local or distant stage colon cancer or other cancer sites. Results may also be non-generalizable to patient populations that were diagnosed before or after our study cohort due to changes in both treatment regimens for colon cancer and area-level socioeconomic status. We also relied solely on residential histories collected from LexisNexis and did not include self-reported information that may be used to validate and/or augment LexisNexis data. However, recent studies found good concordance (82-92%) between LexisNexis addresses and addresses collected from study participants (Susan Hurley et al., 2017; G. M. Jacquez et al., 2011). For our study, data concordance between the LexisNexis locations and the cases' census tracts at time of diagnosis reported by the NJSCR was around 83%, and it increased to 93% when comparisons were limited to a 6-month window before and after the diagnosis date. Although these potential data errors could have contributed to non-differential misclassification, the extent of the bias in either direction would likely be minimal due to the low proportion of cases affected.

The results are also limited by the available follow-up time (maximum 11 years). Longer follow-up could increase the number of observed deaths as well as increase the number of potential moves over the follow-up period. If more people moved and we were looking at long-term survival (e.g., 20-25 years), then residential histories may have a larger impact on predicting survival and explaining the spatial variance. The study also did not include individual-level measures of poverty or income, which previous studies have shown to play a significant role in determining colon cancer survival outcomes (Aarts et al., 2010; Lian et al., 2011; Robinson, 2019)

Residential locations and CT-poverty prior to the diagnosis could have also impacted survival outcomes. We had access to residences prior to diagnosis, but in this study, we were limited by the specific spatial model employed, which required a start at the New Jersey diagnosis location. We are presently exploring alternative spatial models and approaches to assess the role of residential mobility prior to diagnosis on colon cancer survival outcomes.

Finally, the neighborhood geographic measure of poverty was based on census tracts. A change in geographic unit (census tract to block group) may result in different conclusions than those that are based on census tracts (i.e., modifiable areal unit problem) (Buzzelli, 2020; Pawitan & Steel, 2009; Sahar et al., 2019).

Conclusion

Including residential histories opens up new and important avenues of inquiry to better understand the complex relationships between people and places, and to evaluate the effects of residential mobility on cancer outcomes. Our findings suggest that residential mobility after regional stage colon cancer diagnosis has the potential to change geographic patterns of the risk of colon cancer death, increasing the spatial variance and ultimately leading to greater unexplained geographic disparities. Including residential mobility may also allow for more precise spatial risk of death estimates. However, in this study, which was limited in both geographic scope and cancer type, the results based on the traditional approach using only address at diagnosis were not substantially different from the approach using residential mobility data. Both approaches provide relevant results for understanding cancer survival disparities and prioritizing interventions.

CHAPTER 4:
MEASURING NEIGHBORHOOD LANDSCAPES: ASSOCIATIONS BETWEEN
NEIGHBORHOOD'S LANDSCAPE CHARACTERISTICS AND COLON CANCER
SURVIVAL.

Abstract

Abstract: Landscape characteristics have been shown to influence health outcomes, but few studies have examined their relationship with cancer survival. We used data from the National Land Cover Database to examine associations between regional-stage colon cancer survival and 27 different landscape metrics. The study population included all adult New Jersey residents diagnosed between 2006-2011 and followed until 31.12.2016 (N=3,949). Patient data were obtained from the New Jersey State Cancer Registry and linked to LexisNexis to obtain residential histories. Cox proportional hazard regression was used to estimate hazard ratios (HR) and 95% confidence intervals (CI95) for the different landscape metrics. An increasing proportion of high-intensity developed lands with 80-100% impervious surfaces per cell/pixel was significantly associated with the risk of colon cancer death (HR=1.006; CI95=1.002-1.01) after controlling for neighborhood-poverty and other individual-level factors. In contrast, an increase in the aggregation and connectivity of vegetation-dominated low-intensity developed lands with 20-<40% impervious surfaces per cell/pixel was significantly associated with the decrease in risk of death from colon cancer (HR=0.996; CI95=0.992-0.999). Reducing impervious surfaces in residential areas may increase the aesthetic value and provide

conditions more advantageous to a healthy lifestyle such as walking. Further research is needed to understand how these landscape characteristics impact survival.

This chapter was submitted to the special issue of International Journal of Environmental Research and Public Health on "Impacts of the Neighborhood Environment on Cancer Prevention, Outcomes and Disparities".

Wiese, D., Maiti, A., Harris, G., Lynch, S., Stroup, A., Vucetic, S., Gutierrez-Velez, V.H., Henry, K.A. (revised): Measuring Neighborhood Landscapes: Associations between neighborhood's landscape characteristics and colon cancer survival.
International Journal of Environmental Research and Public Health

Introduction

Neighborhood characteristics can capture social, physical, and economic conditions of the environment in which a person lives (A. V. Diez Roux, 2001). According to Northridge et al. (Northridge, Sclar, & Biswas, 2003), social context and the built environment are two intermediate factors that influence health and well-being at the individual and population level. The social context of a neighborhood is most often defined through its socioeconomic and demographic composition using census data (e.g., median income, % living below poverty) (Link & Phelan, 1995). The built environment includes physical characteristics of neighborhoods (i.e., land covers) that are human-made or modified (Rapoport, 2011) such as buildings, roads, housing conditions, parks sidewalks or greenspace that can potentially provide the setting for human activity (Renalds et al., 2010).

Therapeutic landscape theory (Gesler, 1992) can add depth to our understanding of how landscapes impact health and cancer outcomes (Matsuoka & Sullivan, 2011). Gesler's therapeutic landscape theory posits that social and spatial factors are interconnected, and modification of place can have various effects on human health. Moreover, people may have experienced different feelings and healing effects in various places caused by perceptions of landscape aesthetics (Gesler, 1992). Some studies linked the aesthetical value of the landscapes, analyzed through photography of micro-landscapes, to emotional load and restoration (Dramstad, Tveit, Fjellstad, & Fry, 2006; Palmer, 2004; Staats, Kieviet, & Hartig, 2003; van den Berg, Koole, & van der Wulp, 2003). Lee et al (Lee, Ellis, Kweon, & Hong, 2008) found that neighborhood satisfaction is related to landscape structure. Others suggest that configuration and composition of different types of land

cover and land use classes could be influential on place perception and emotions because “perception, cognition and evaluation are highly interrelated processes” (Kaplan, 1987), and perception of environmental landscape is important (among others) for movement, and social interactions (Ittelson, 1978; Iverson Nassauer, 1995; Zube, Sell, & Taylor, 1982). In addition, factors such as natural habitat fragmentation and impervious surfaces may also influence health outcomes considering the relation between urban design and mental health (Jackson, 2003).

For numerous cancer types, studies report significant associations between neighborhood socioeconomic conditions and cancer mortality and survival (Henry et al., 2009; Henry et al., 2014; Krieger et al., 1999; Lian et al., 2011; Niu et al., 2010; Singh, 2003; Singh & Jemal, 2017; Singh et al., 2011; L. Wang et al., 2011; D. Wiese et al., 2020a; D. Wiese et al., 2020b; Yang et al., 2019). Generally, these studies have found that people living in low socioeconomic-status neighborhoods have significantly higher mortality or shorter survival after diagnosis. The relationship is less consistent between the built environment and cancer mortality and survival. Among the studies that have examined these associations. For example, a study focused on greenspace availability and accessibility by James et al. (James, Hart, Banay, & Laden, 2016) found that higher levels of green vegetation were associated with decreased cancer mortality. Several other studies on lung cancer-specific mortality did not find any significant associations (Mitchell & Popham, 2008; E. Richardson, Pearce, Mitchell, Day, & Kingham, 2010; E. A. Richardson & Mitchell, 2010; E. A. Richardson et al., 2012). For breast cancer, Keegan et al. (Keegan et al., 2014) did not find any positive influence of park availability and survival, but did

report an increased risk of breast cancer death in areas with higher traffic/road density possibly through discouraging recreational-based physical activity.

Among the studies that have examined the relationship between the built environment and non-cancer health outcomes (e.g. cardiovascular and/or mental health illnesses), most have focused on greenspace. Mears et al. (Mears et al., 2020) summarized positive effects of the urban greenspace on health outcomes including the reduction of traffic pollution, a reduction of heat-island effects, as well as an increase of emotional recreation effect and physical or social activities. In one study, Bratman et al. (Bratman et al., 2019) examined evidence across the natural, social, and health sciences on the impacts of nature experience on mental health. They argue that the configuration and conception of greenspace is essential for human well-being. Kondo et al. (Kondo et al., 2018) also concluded that navigating through urban greenspaces causes more positive emotions compared to the built urban environment. Additionally, the quality of the greenspace (e.g. well-maintained parks) has a positive influence for psychological well-being (Mears & Brindley, 2019; Mears, Brindley, Jorgensen, Ersoy, & Maheswaran, 2019; Sugiyama, Carver, Koohsari, & Veitch, 2018; Van Dillen, de Vries, Groenewegen, & Spreeuwenberg, 2012). Moreover, urban gardens and well-maintained front yards may influence population health through the pathways of aesthetics and promote walking and outdoor physical activity (Hunter & Brown, 2012). Also, walkability, mix and type of businesses and land use composition of a neighborhood may encourage healthy behavior and “facilitate integration of habits into a daily lifestyle” (p.76)(Renalds et al., 2010). Given the importance of greenspace on health and a lack of studies focused specifically on cancer, more research is needed examining the relationship between the built

environment and cancer outcomes in order to better understand these relationships and underlying pathways (Gomez et al., 2015).

With the exception of a few studies (Demoury et al., 2017; James et al., 2016; Keegan et al., 2014), insufficient attention has been given to integration of both socio-environmental landscape on cancer outcomes. Considering that cancer survival has shown to be associated with mental well-being (Ashing-Giwa, Lim, & Tang, 2010; Cunningham, Sarfati, Stanley, Peterson, & Collings, 2015), integrating landscape and built-environmental characteristics may be useful for quantifying neighborhood quality and help measure the effect of neighborhood land cover configuration and composition on cancer outcomes.

Estimating the aesthetical value of a landscape and finding the best measure for defining the built environment can be challenging because there are an infinite number of ways to measure and operationalize the data in research (McCormack et al., 2019) and further distilling the essential components that impact health and health behaviors. Frequently, the built environment is integrated into modeling using census data such as housing density and quality (Gomez et al., 2015). However, recent technologies in image classification allow an estimation of neighborhood physical disorder (e.g. presence of abounded buildings, non-maintained roads) or greenspace availability and quality using remote sensing products, and street view in health research (Maharana & Nsoesie, 2018; Plascak et al., 2020; Rzotkiewicz, Pearson, Dougherty, Shortridge, & Wilson, 2018). Despite the increasing literature, however, when examining the role of greenspace on health outcomes, several gaps remain including “standardization” of appropriate measures (Kondo et al., 2018; Wendelboe-Nelson et al., 2019). Therefore, it is important

to develop clear concepts and metrics for quantifying and measuring the effects of neighborhood's landscape and the built environment on health outcomes in order to understand the underlying pathways.

Landscape metrics are a commonly used technique in landscape ecology – an environmental science approach of landscape characterization, evaluation, and design (Turner et al., 2001) quantify landscape characteristics and features. This concept is frequently applied in urban planning, biodiversity, species richness and conservation as well as in infectious disease epidemiology studies (Evelin Uuema, Antrop, Roosaare, Marja, & Mander, 2009; Evelyn Uuema et al., 2013). Landscape metrics are also common in urban aesthetics evaluation (Dramstad et al., 2006; Palmer, 2004). To date there have been no population-based cancer studies that integrate landscape metrics.

In this study we examine associations between regional-stage colon cancer survival and several landscape metrics that quantify neighborhood's built-environment using population-based cancer surveillance data from the New Jersey State Cancer Registry, residential histories obtained through data linkage with LexisNexis, and land cover and land use data from the National Land Cover Database.

Materials and Methods

Study Population

The New Jersey State Cancer Registry (NJSCR) provided all colon cancer cases. The NJSCR is a population-based cancer registry. It operates since October 1978 and monitors cancer among the more than 8.9 million residents of New Jersey (NJSCR, 2018). The study population includes all New Jersey residents 18 years and older with

histologically confirmed, first primary regional stage colon cancer as defined according to the International Classification of Diseases for Oncology, 3rd Edition (ICD-O3 C180-C189, C260; excluding histology codes 9050-9055, 9140, 9590-9992) (Percy et al., 1990) diagnosed between January 1, 2006 and December 31, 2011 (N = 4,041). To simplify the interpretation of the results and controlling for disparities between tumor stages, our study sample was limited to regional stage colon cancers. The study was reviewed and/or approved by Temple and Rutgers University Institutional Review Boards.

Individual-level factors included age at diagnosis, gender (male, female), and race/ethnicity [Non-Hispanic (NH) White, NH Black, NH Asian/Pacific Islander (API), NH Other, and Hispanic (any race)], vital status, including date of death and cause of death (if deceased) or date of last contact (if alive). Cases were followed until their deaths, relocation from the state of New Jersey or until December 31, 2016. Deaths attributed to *colon cancer* were coded based on ICD-10 code C18 (Percy et al., 1990). NJSCR constantly updates the database through linkages with state and national sources including death data from the New Jersey Department of Health Office of Vital Statistics and Registry and the National Death Index, hospital discharge files, Centers for Medicare and Medicaid Services, Social Security Administration Services for Epidemiologic Researchers, and motor vehicle registration files (NJSCR, 2018).

Residential Histories

The NJSCR linked the study population to the commercial residential history database developed by LexisNexis, Inc. (Miamisburg, Ohio, US; <https://www.lexisnexis.com/en-us/products/public-records.page>), (Susan Hurley et al., 2017; Geoffrey M Jacquez et al., 2010; Stinchcomb & Roeser, 2016; D. C. Wheeler & Wang, 2015). The majority 3,949

(97.6%) of regional stage colon cancer cases had residential information available for up to 20 of the most recent addresses between 1946 and 2018 with documented start and stop dates. A preprocessing technique described by Wiese et al. (D. Wiese et al., 2020b) was applied to establish a complete residential timeline for the time after diagnosis. Using the North American Association of Central Cancer Registries (NAACCR) AGGIE Geocoder (Texas A&M, 2016), all residential addresses were geocoded to the 2010 census tract (CT) boundaries.

Socioeconomic Variables

The socio-economic variables included a widely used census tract poverty (CT-poverty) variable defined as the proportion of population 18 years and older living below the Federal poverty level. Additionally, we included information on census tract measures of housing density and median year housing built. The required variables were obtained from publicly available U.S. Census and American Community Survey (ACS) data. U.S. Census 2010 and the ACS 5-year average data 2006-2010, 2007-2011, 2008-2012, 2009-2013, and 2010-2014 were used for residencies between 2006-2010. ACS 2011-2015, 2012-2016, and 2013-2017 were used for residencies between 2011 and 2016. For each case, all corresponding CTs during the follow-up period were included, and every residential record received a corresponding value based on the earliest date of the residential appearance in the data set. If a patient remained at the same CT after the diagnosis over multiple years, we assigned annual values to capture changes within the neighborhood that could be caused by the gentrification.

Environmental Variables

The neighborhood built-environment factors were measured by landscape metrics that capture the landscape characteristics such as land cover and land use composition and configuration of the neighborhood (e.g., census tract) considering local spatial patterns. The National Land Cover Database (NLCD) was used to extract land cover classes and proportion of imperviousness for the years 2006 (used for residential records 2006-2009), 2011 (used for residential records 2010-2014) and 2016 (used for residential records 2015-2016). The NLCD products are free available raster files of 30m spatial resolution that were classified using Landsat-based satellite imagery by the U.S. geological survey (Homer, Fry, & Barnes, 2012). The NLCD has several products: The land cover raster includes 16 categorical classes, and the imperviousness raster, which has a continuous scale of proportion of impervious surfaces per pixel. To reduce the number of categories for further analysis, we reclassified the original NLCD raster using R package *raster* (Hijmans, 2016) into 7 classes (Forest, Grass, Shrubs, Developed Lands (Open, Low-, Medium-, and High-Intensity) (Table 7). Land use information was obtained from the New Jersey Geographic Information Network (<https://njgin.nj.gov/njgin/edata/parcels/#!//>), which includes publicly available information on parcel use for the state of New Jersey as of the year 2019. The original shapefile (spatial polygon) was rasterized (i.e. converted) using R package *raster* (Hijmans, 2016) based on the parcel use category and reclassified by keeping only industrial and commercial lands. as high-intensity developed lands (NLCD class 24) with 80-100% imperviousness.

Table 7: Overview of the land cover classes after the reclassification of the NLCD dataset.

The NLCD Code	Reclassification	Commentary
21: Developed Open Space	U21: Developed Open Space	Max 20% imperviousness per cell/pixel
22: Developed Low Intensity	U22: Developed Low Intensity	20-49% imperviousness per cell/pixel
23: Developed Medium Intensity	U23: Developed Medium Intensity	50-79% imperviousness per cell/pixel
24: Developed High Intensity	U24: Developed High Intensity	80-100% imperviousness per cell/pixel
41: Deciduous Forest	Forest	Dominated by tree canopy and includes any type of parks and squares
42: Evergreen Forest		
43: Mixed Forest		
52: Shrub/Scrub	Shrubs	Dominated by shrubs; present on empty housing parcels
71: Grasslands/Herbaceous	Grassland	In urban areas, may assume a low-quality green space
81: Pasture/Hay		
82: Cultivated Crops		
90: Woody Wetland	Forest	Woody wetlands are common in southern New Jersey, and have large proportions of deciduous trees
95: Emergent Herbaceous Wetland	Grassland	Herbaceous (also grassy) wetlands are typical for many coastal regions.

Previous studies have already examined the NLCD’s classification of developed lands to estimate greenspace availability in urban areas, and have noted that open and low-intensity developed lands are suitable for identification of greenspaces and trees (Akpınar et al., 2016; Callaghan et al., 2019; E. A. Richardson et al., 2012; Wu & Jackson, 2017). Typically, areas with predominantly large housing parcels, roads surrounded by greenspaces, urban housing with larger backyards or isolated large roads would be classified as developed open land (NLCD class 21), assuming a large proportion of

greenspaces than of impervious/built surfaces (max. 20%) in the area. In contrast, a central business district, densely built inner-city housing or a largely expanded shopping mall would be classified

Landscape metrics on patch (square) and class levels (greenspaces) within a landscape (i.e., census tract) were calculated using R statistical software (R Core Team, 2013) (version 4.0.1), implementing packages *landscapemetrics* (Hesselbarth, Sciaini, With, Wiegand, & Nowosad, 2019) and *SDMTools* (VanDerWal et al., 2014). Landscape metrics include more than 50 measures. Because most variables are highly correlated or difficult to interpret and compare between landscapes because of the open, non-fixed range scale (Riitters et al., 1995; Turner et al., 2001), we developed a list of the influential landscape metrics as defined by previous studies (Mears & Brindley, 2019; Mears et al., 2019; Mears et al., 2020). Table 8 summarizes all landscape metrics that were considered in present study. Selected landscape metrics and the land cover classification were also mapped using QGIS v.3.10 (Figure 4).

Table 8: Overview of the area-based and individual variables

Variables	Land Covers	Definition	Commentary
Land Cover Class-Level Metrics			
Class Proportion	Forest, Grass, Shrubs, Industrial, Developed Lands (Open, Low, Medium, High Intensities)	Composition metric. Proportional coverage-% of the landscape covered by each type.	Used by (H Li & Reynolds, 1995; Habin Li & Reynolds, 1994) and recommended by (Mears et al., 2020) for green and water space; (W.-L. Tsai et al., 2018) recommend for Forest, Shrubs and Grass
Aggregation Index (AI)	Developed Lands (Open, Low, Medium, High Intensities), Industrial Areas	Configuration metric. Computed as an area-weighted mean class aggregation index, where each class is weighted by its proportional area in the landscape.	Redundant with several other metrics of proportion, cohesion and contiguity and may be a meaningful alternative (Turner et al., 2001)

Table 8 continued

Variables	Land Covers	Definition	Commentary
Splitting Index	Forest, Grass, Shrubs	Configuration metric. A large splitting index, results from land covers being split into many patches with an even size distribution.	Correlated with the aggregation index. Applied for green spaces only in relation to health outcomes (Mears et al., 2019)
Contiguity Index (CI)	Developed Lands (Open, Low, Medium, High Intensities), Forest, Grass	Configuration metric. Large contiguous patches will result in larger contiguity index values.	CI for green/tree land cover classes associated with health outcomes (Mears et al., 2019)
Landscape-Level Metrics			
Shannon Diversity Index	Based on all Land Cover Classes	Composition metric. The more classes and the more equally distributed, the higher the index.	Used for measuring the aesthetic value and diversity (Dramstad et al., 2006) Associated with health outcomes (Mears et al., 2019)
Patch Richness Density (PRD)	Based on all Land Cover Classes	Number of patches per hectare. High values indicate high dispersion	PRD for green areas and recreational lands associated with poor health (Mears et al., 2019)
Contagion Index	Based on all Land Cover Classes	Composition metric. High values indicate result from landscapes with a few large, contiguous patches and low dispersion and interspersions of patch types	
Average Proportion of Imperviousness	Census Tract Average based on NLCD dataset estimating imperviousness proportion per pixel	Composition metric. Highly negatively correlated with Tree Canopy proportions but is more accurate	Highly correlated (negative) with Tree Canopy Cover but more accurate (Greenfield, Nowak, & Walton, 2009; Nowak & Greenfield, 2010)
Census-Based Variables			
Poverty Level by Category	Percentage of population 18 and older living below federal poverty level.	Socio-economic status	Associated with cancer outcomes including survival and mortality
Median Year Structures Built	Median that the areas residential buildings were constructed.	Organized into categories.	Housing age and conditions are associated with health outcomes (Howden-Chapman, 2004; Jackson, 2003) and poverty (Shaw, 2004)
Housing Density	Number of structures per area unit (acre)	Continuous variable defined by census tract	Potential intermediate factor in health outcomes (Rauh, Landrigan, & Claudio, 2008)

Note: All developed lands are also referenced as U21=Open, U22=Low, U23=Medium, U24=High Intensities

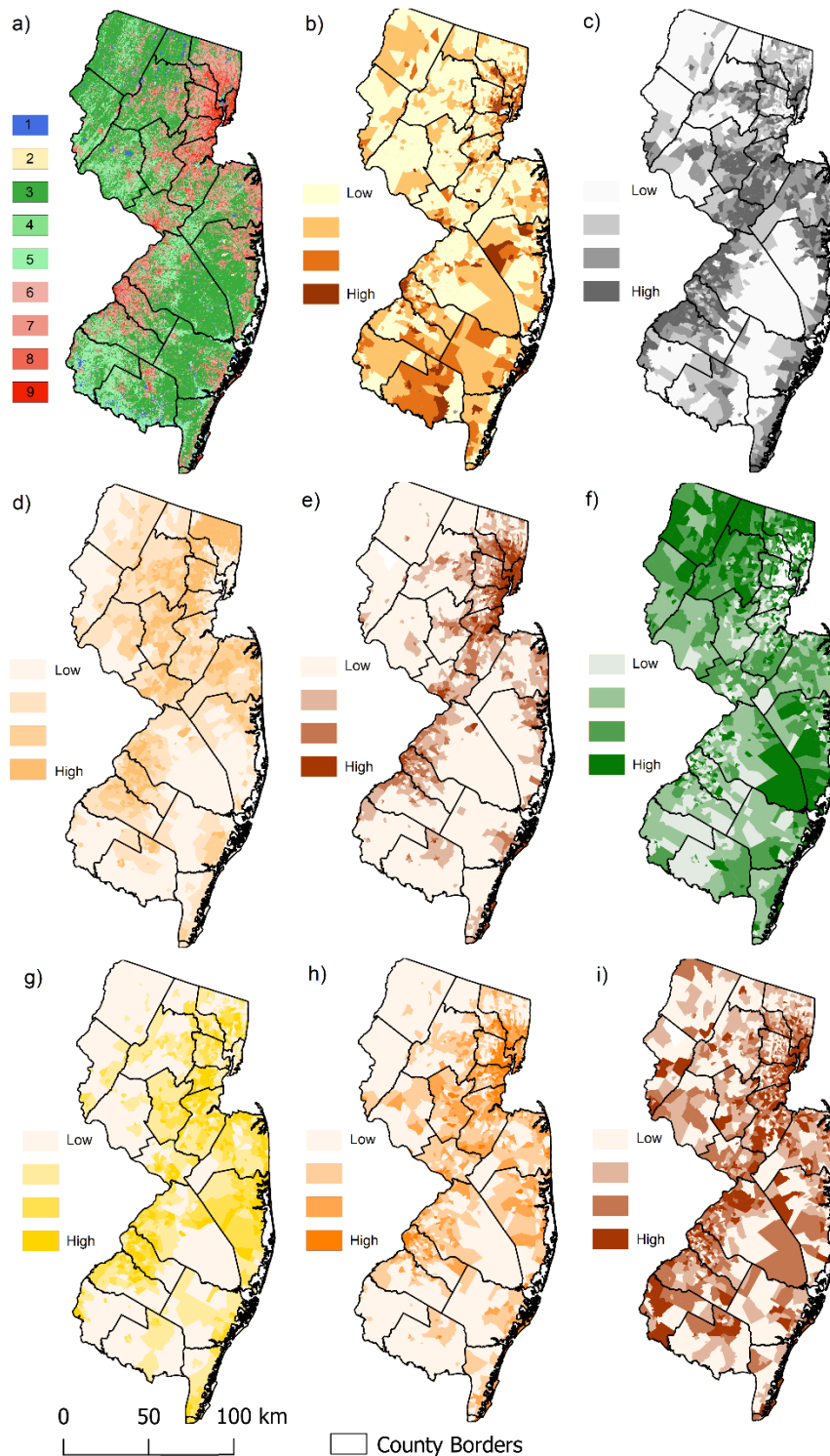


Figure 4: Land cover classification and landscape characteristics of New Jersey.

a) Reclassified land cover classification (1=Water, 2=Sand/Bare, 3=Forest, 4=Shrubs, 5=Grass, 6= Open, 7=Low, 8=Medium, 9=High intensity developed lands), b) Census tract poverty level by categories (0-5%, 5-10%, 10-20%, 20+%), c) Shannon Diversity Index (SHDI) by quartiles, d) Proportion of open developed lands by quartiles, e) Proportion of high-intensity developed lands by quartiles, f) Forest contiguity index by quartiles, g) Aggregation index (AI) of open developed lands by quartiles, h) Aggregation index (AI) of medium-intensity developed lands by quartiles, i) Aggregation index (AI) of high-intensity developed lands by quartiles

Statistical Analysis

The survival time for every patient was calculated in months as the difference between the date of diagnosis and the date of death or date of last contact. Cases missing survival time (i.e., only ascertained through death certificates or autopsy) were excluded from this analysis. Patients were censored at the date of death if they died from causes other than colon cancer, the date the patient was lost to follow-up, at the end of the follow-up period (31 December 2016) or at the time of the relocation from the State of New Jersey, whichever occurred first. Additionally, we calculated time intervals to every CT location, and assigned start and end dates. Every time interval received a corresponding value of each socioeconomic and environmental neighborhood variables, as well as the vital status of the patient (1-dead, 0-alive).

We designed a process for variable selection and evaluation with minor modifications based on methodology from a previous neighborhood wide association study or Nwas (Shannon M. Lynch, Mitra, Ross, et al., 2017). All methodological steps are summarized in Figure 5. First, we developed a series of univariate, crude models to estimate the effect of each selected landscape metric and other nSES on the duration of survival time using Cox proportional hazard regression for time-varying covariates (Z. Zhang et al., 2018). Cox proportional hazard regression is a widely used semi-parametric time-to-event modeling technique, where death is considered being the event. Cox proportional hazard regression allows incorporation of individual and area-level covariates. Additionally, it does not require to define any probabilities in advance, and is suitable for time-varying covariates (Mills, 2010). Variables were selected if they reached significance at $p < 0.05$.

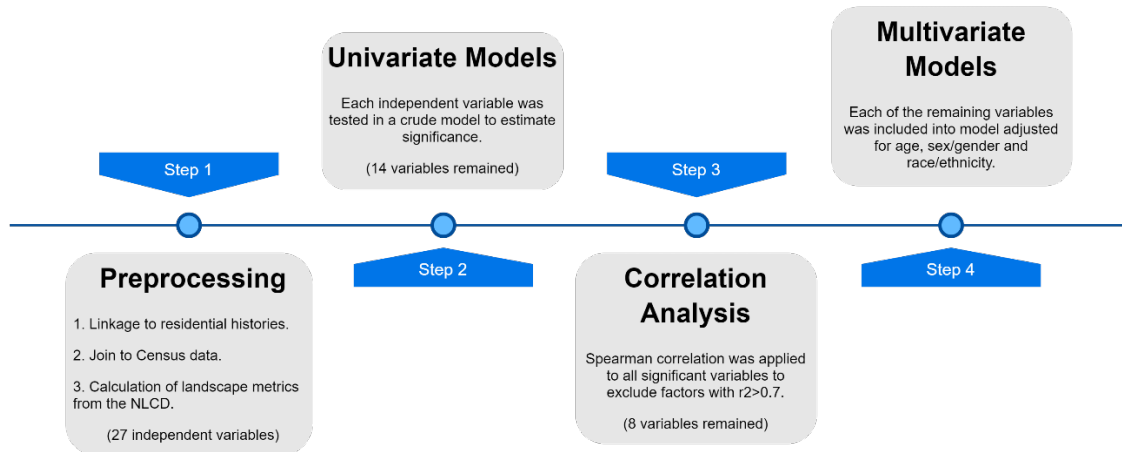


Figure 5: Workflow diagram

Second, we applied Spearman correlation analysis of the previously selected significant variables because many landscape metrics are highly correlated and redundant. After excluding all highly correlated variables ($r^2 > 0.7$), the number of variables was reduced from 14 to 8 (Figure 6). The remaining variables were aggregation indices of all four intensity levels of developed lands as well as proportion of high intensity developed lands, forest contiguity index, Shannon diversity index, and the patch richness density, CT-poverty.

We then developed a set of models that included all individual-level variables (sex, age and race/ethnicity, sub-stage at the diagnosis and mover status) and each nSES or landscape metric. To estimate the risk of death from colon cancer by each individual and area-based variable, all coefficients were exponentiated and expressed as Hazard Ratios (HR). For continuous variables, positive HRs indicate a positive association with the increase in risk of death. For categorical variables, the HRs are compared to the reference group. An HR=1 indicates that the risk is similar across all cases and groups (Zwiener, Blettner, & Hommel, 2011).

During the follow-up period, 65.5% remained at their diagnosis CT and 12.1% left New Jersey during the study period. Among those who moved, 18.5% only moved once, 12.3% moved twice, and 3.6% moved 3 or more times after cancer diagnosis. The average time spent at the CT at diagnosis was 7.5 years (Table 9).

Table 9: Study population characteristics

Characteristics	Overall (n = 3,949)
Age	
Mean (SD)	65.8 (13.3)
Median [Min, Max]	68.0 [21.0, 85.0]
Gender	
Male	1,878 (47.6%)
Female	2,071 (52.4%)
Race/Ethnicity	
NH-White	2,902 (73.5%)
NH-Black	488 (12.4%)
Hispanic (any race)	325 (8.2%)
NH-API	141 (3.6%)
Other	93 (2.4%)
Regional Stage Subcategory	
Regional, direct extension only	1,339 (33.9%)
Regional, lymph nodes only	1,268 (32.1%)
Regional, both	1,342 (34.0%)
Vital Status	
Censored	2,862 (72.5%)
Colon Cancer Death	1,087 (27.5%)
Survival Time (months)	
Mean (SD)	62.3 (38.0)
Median [Min, Max]	66.0 [1.00, 139]
CT Changes (Type of “moves”)	
CT at Date of Diagnosis Only	2,587 (65.5%)
Change in Residential CT within NJ	885 (22.4%)
Change in Residential CT outside NJ	477 (12.1%)

Note: CT = Census Tract; NJ = New Jersey
SD = Standard Deviation; NH = Non-Hispanic; API= Asian/Pacific Islanders

The distribution of colon cancer patients by neighborhood/landscape characteristics at the time of diagnosis showed that the majority of patients 73.9% (n=2919) lived in areas with a poverty level less than 10%. Additionally, most patients (82.5% n=3259) were living in neighborhoods with relatively low (<=30%) proportion of open developed lands (areas with large greenspace cover), while 89.8% (n=3547) patients lived in neighborhoods where more than 30% of the total landscape are dominated by high-intensity developed lands (less than 20% greenspaces). Twenty four percent (n=939) of all colon cancer patients were residents in neighborhood with no tree cover. The distribution of the study population for these and other landscape metric is summarized in Figure 7.

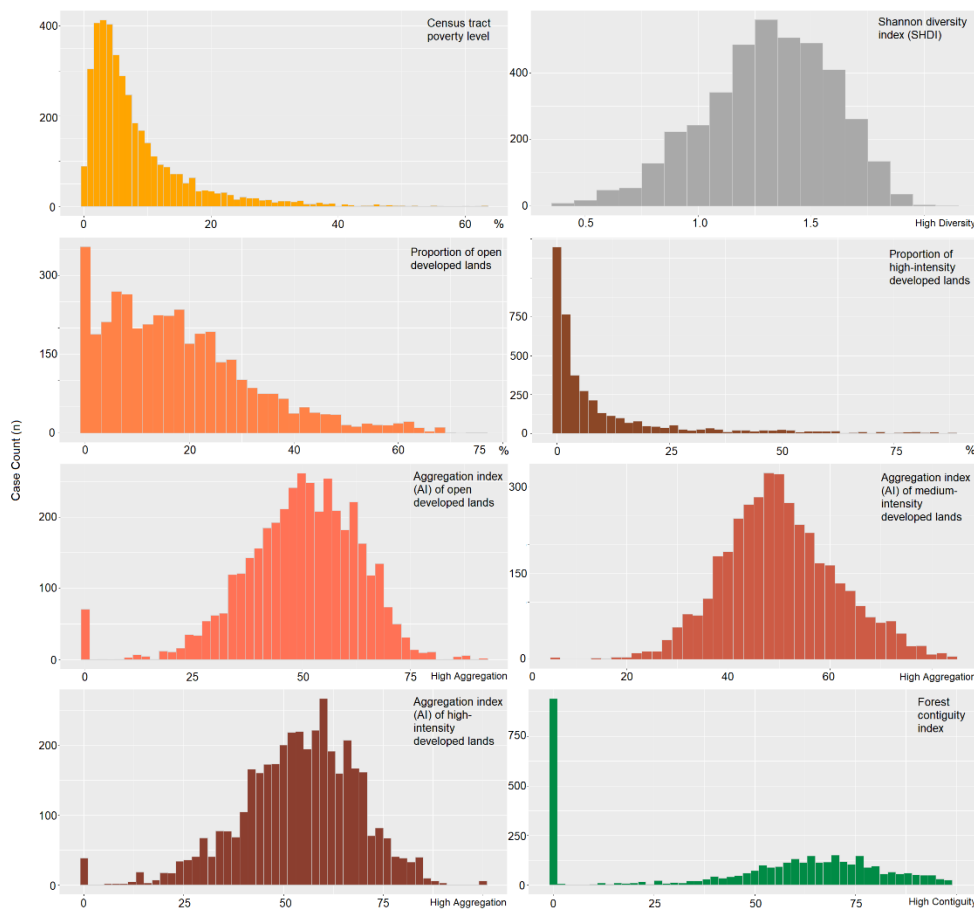


Figure 7: Distribution of colon cancer patients by neighborhood/landscape characteristics

Univariate Models

Figure 8 summarizes the model results for each variable. For every 10% increase in housing density, the risk of death increased by 5% (HR=1.005; 95% CI=1.002-1.009).

For CT poverty, cases living in CT with poverty levels from 10-20% had a higher risk of death than those living in CTs with poverty levels from 0-5% (HR=1.23; 95% CI=1.04-1.44).

Of the 27 landscape metrics, six were statistically significant: an increasing proportion of the high- and medium-intensity developed lands were positively associated with the risk of death increase (HR=1.007; 95% CI=1.003-1.01 and HR=1.008; 95% CI=1.005-1.011, respectively). Similarly, a positive association was found between an increasing aggregation index of high- and medium-intensity developed lands and risk of death (HR=1.005; 95% CI=1.001-1.009; HR=1.01; 95% CI=1.007-1.02, respectively). There was also a positive association between increasing patch richness density (HR=1.02; 95% CI=1.01-1.03) and elevated risk of death. Average imperviousness was also significant, indicating an approximate 7% risk increase for every 10% increase of the neighborhood's imperviousness (HR=1.007; 95% CI=1.004-1.011).

Seven landscape metrics were significantly associated with a decrease in risk of colon cancer death including aggregation of low-intensity developed lands (HR=0.996; 95% CI=0.991-0.999), proportion and aggregation of open develop lands (HR=0.992; 95% CI=0.988-0.997, HR=0.996; 95% CI=0.991-0.999, respectively), Shannon diversity index (HR=0.79; 95% CI=0.64-0.98), and proportion and contiguity (i.e., connectivity) of forest/trees (HR=0.995; 95% CI=0.992-0.998, HR=0.997; 95% CI=0.996-0.998, respectively).

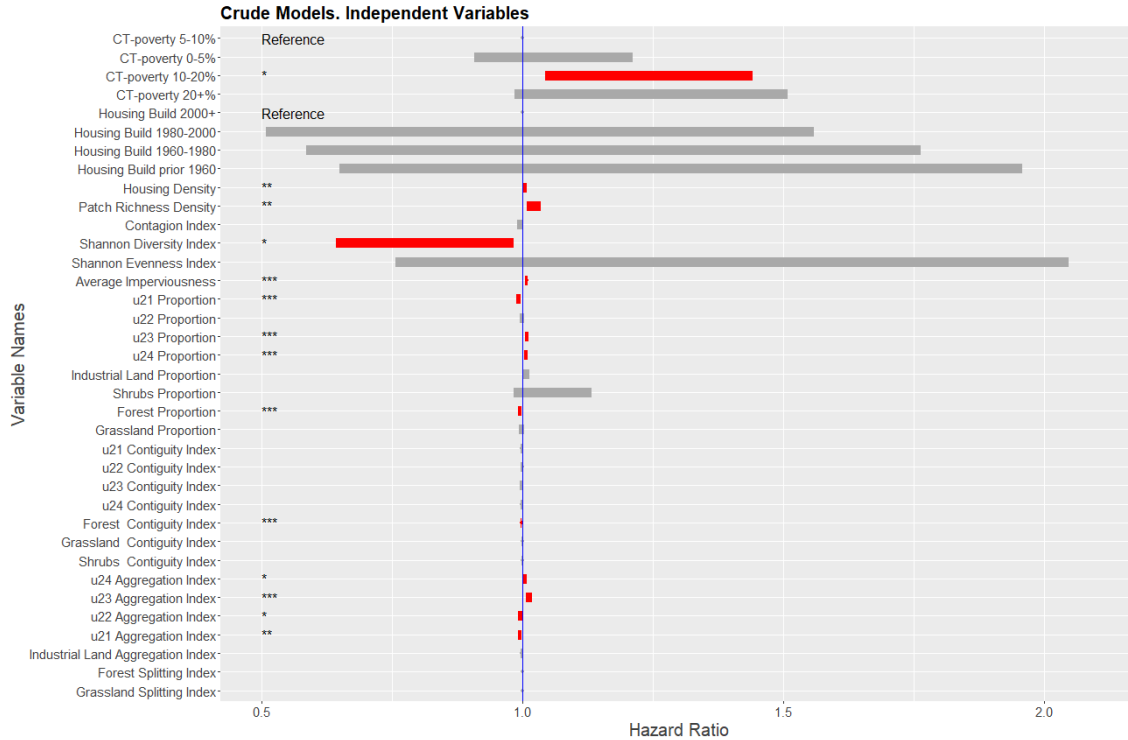


Figure 8: Hazard Ratios of neighborhood-level variables from the crude models, adjusted for each variable individually. Blue line indicates there is no difference in risk. Red bars indicate the 95% confidence interval of statistically significant variables.

Note: Significance levels $<0.05 = *$; $<0.01 = **$; $<0.001 = ***$

Land cover classes are based on the reclassified NLCD raster. U21=Open, U22=Low, U23=Medium, U24=High intensity developed lands

Multivariate Models

In the last step, we developed ten multivariate models adjusting for all individual level variables (age and sub-stage at the diagnosis, sex/gender, race/ethnicity, and mover status) and CT-poverty as well as each landscape metric at a time. The individual level HRs are summarized in Table 10.

Table 10: Hazard Rates of individual-level factors from the multivariate model.

Individual-Level Factors	Coefficient	Hazard Rate (95% Confidence Interval)	P-Value
Male	Reference		
Female	-0.21	0.81 (0.73-0.9)	<0.001

Table 10 continued

Individual-Level Factors	Coefficient	Hazard Rate (95% Confidence Interval)	P-Value
Age at diagnosis	0.02	1.025 (1.02-1.03)	<0.001
Regional. direct extensions only	Reference		
Regional. lymph nodes only	-0.39	0.68 (0.58-0.78)	<0.001
Regional. direct extension and lymph nodes	0.6	1.83 (1.62-2.07)	<0.001
NH-White	Reference		
NH-Black	0.45	1.56 (1.35-1.81)	<0.001
Hispanic	0.49	1.64 (1.37-1.96)	<0.001
Asian/Pacific Islanders	0.2	1.22 (0.92-1.62)	0.17
Others	0.13	1.14 (0.79-1.64)	0.45
Non-Movers	Reference		
Movers	1.15	3.17 (2.84-3.53)	<0.001

Note: Hazard Rates and 95% confidence intervals are exponentiated the parameter estimates/coefficients

In multivariable models, the proportion of high-intensity developed-lands had the strongest association with the colon cancer risk of death. For every 10% increase in the proportion the risk of death was increasing by 6% (HR=1.006; 95% CI=1.002-1.01). Additionally, increasing aggregation index (i.e., more compact areas) of high-intensity developed-lands was positively associated with risk of colon cancer death (HR=1.005; 95% CI=1.001-1.009). Similar associations were found for the increasing aggregation index of medium intensity developed lands (HR=1.009; 95% CI=1.003-1.015). In contrast, a 6% decrease in risk of colon cancer death was estimated for every 10% increase in the proportion of open developed lands (HR=0.994; 95% CI=0.988-0.999). Increasing aggregation index of low-intensity developed-lands (HR=0.995; 95% CI=0.99-0.999) and increasing forest contiguity index (HR=0.998; 95% CI=0.996-0.999) were all significantly negatively associated with the risk of colon cancer death. CT-poverty was no longer significant in multivariable models (Figure 9).

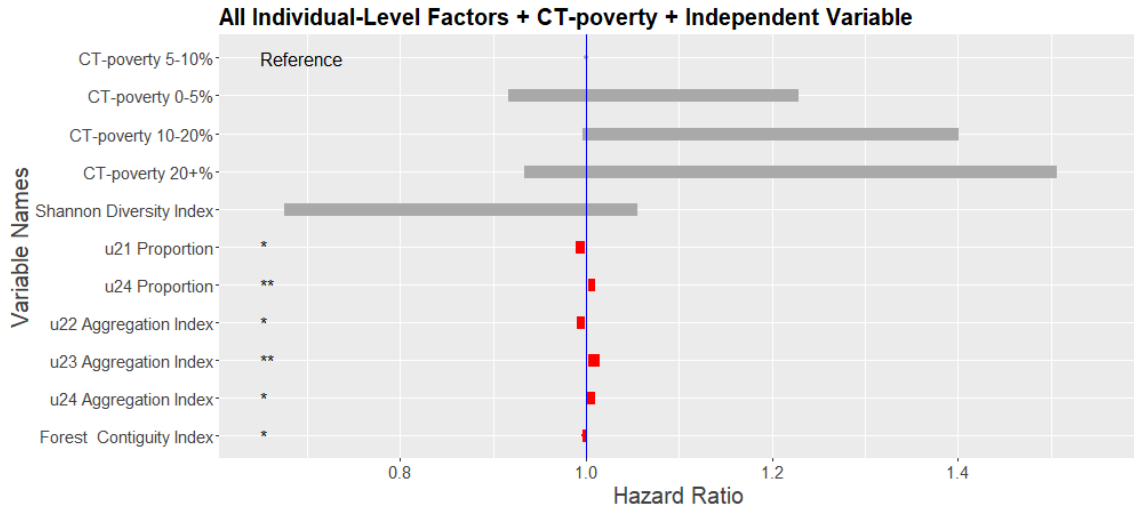


Figure 9: Hazard Ratios of neighborhood-level variables from the multivariate model adjusted for all individual-level factors and each neighborhood/landscape variable individually. Blue line indicates a zero-coefficient or no difference in risk. Red bars indicate the 95% confidence interval of statistically significant variables.

Note: Significance levels <math><0.05 = *</math>; <math><0.01 = **</math>; <math><0.001 = ***</math>

Land cover classes are based on the reclassified NLCD raster. U21=Open, U22=Low, U23=Medium, U24=High intensity developed lands.

Discussion

To date only a few studies have integrated landscape characteristics into cancer disparities research. Traditionally, socio-epidemiological neighborhood studies focus look on finding associations between the neighborhood factors and cancer survival using Census-based SES data. However, the physical environment (Rappaport, 2007), recreational activities (Banzhaf & Walsh, 2008), and urban design are important components in the selection of residency, especially among older adults (Litwak & Longino Jr, 1987). While experiences of place are recognized and there have been several attempts to incorporate the role of the landscape characteristics into cancer research (Gomez et al., 2015; Keegan et al., 2014), often it is done through the utilization of road networks or based on parcel use, greenspace availability and accessibility. Only few

attempts were made using satellite imagery-based products (Demoury et al., 2017; Iyer et al., 2020; O'Callaghan-Gordo et al., 2018).

We evaluated the relationship between the risk of death from colon cancer and several area-based landscape characteristics that attempt to describe the configuration and composition of the built environment after adjusting for individual-level factors and CT-poverty. This study adds to a body of literature on the effects of neighborhood landscape characteristics on cancer survival. Independent of CT-poverty and a few individual demographic and prognostic factors, we found a significant relationship between the risk of death from colon cancer and the proportion and aggregation of high-intensity developed-lands (i.e., areas dominated by buildings and roads). The risk of death increased as the proportion and aggregation of high-intensity developed-lands increased. The reasons for this relationship may be attributed to a low prevalence of greenspaces in areas with more high-intensity developed-lands. Less greenspace in these places could reduce access to recreational sites that promote physical activity. Furthermore, such places may evoke negative emotional feelings and psychological well-being that have been found to be associated with a lack of greenspace (Bratman et al., 2019). Creating landscapes that promote exercise and active transport such as biking and walking is particularly important to cancer patients. Many clinical trials have shown the positive effects that walking and physical activity can have on quality of life and survival time in cancer patients (Schmitz et al., 2021), and are known to reduce risk of colorectal cancer development (Brown, Winters-Stone, Lee, & Schmitz, 2012; Van Blarigan & Meyerhardt, 2015).

The relationship of worse survival with increasing high-intensity developed-lands independent of CT-poverty is likely related to what this metric is capturing. High-intensity developed lands are characterized by places that include less than 20% greenspace of total land area (Homer et al., 2012). An increasing proportion and aggregation of high-intensity developed lands is generally characterized by very compact and large areas within neighborhoods with few or no parks and where the majority of greenspace is a result of overgrown vegetation from abandoned lots. A study from Philadelphia reported that abandoned buildings and lots were associated with negative health outcomes among residents because of dangerous physical and social environments and sanitation/garbage issues (Garvin, Branas, Keddem, Sellman, & Cannuscio, 2013). High-intensity developed lands are also characterized by large parking lots and multiple lane roads that can reduce the aesthetic value of a neighborhood. This can lead to a reduction in walkability. Several previous studies reported the importance of aesthetic value and quality of the neighborhood environment on mental health and physical activity (Akpinar et al., 2016; Bratman et al., 2019; Chen et al., 2019; Hunter & Brown, 2012; Mears & Brindley, 2019; Mears et al., 2020; Wendelboe-Nelson et al., 2019). Additionally, we cannot exclude that this finding could also be a result of unmeasured confounding like environmental pollution in highly urbanized areas or food deserts. The better survival in neighborhoods with a higher proportion of open developed lands is likely a result of large and compact greenspaces found in these areas, which provide potential for recreational activities. Larger open developed lands also assume more contiguous areas with a large proportion of street tree canopy, urban gardens, or large backyards across the landscape (neighborhood).— Having vegetation along the roads,

green front and backyards could have positive emotional effects (Hunter & Brown, 2012; Mears et al., 2020) essential for many cancer survivors. Kondo et al (Kondo et al., 2018) also conclude that navigating through urban greenspaces compared to built urban spaces leads to more positive emotions. Additionally, this could suggest the availability of green infrastructure and neighborhood parks that would increase walkability and physical exercising (Hunter & Brown, 2012; Jennings, Baptiste, Jelks, & Skeete, 2017; Zuniga Teran, 2015).

We also found that increasing forest contiguity index was positively associated with a decrease in the risk of death. This suggests that lower shape complexity and lower interspersion of greenspaces (i.e. larger and better-connected greenspaces) may further decrease the risk of colon cancer death because of the availability of green corridors within cities. – Large contiguous green space leads to more varied use and extended use for physical activity and provides opportunity to strengthen social capital that improves survival. This aligns with an earlier study reporting a positive effect on general and mental health from having fewer, but larger patches/areas of greenspaces, rather than a high density of small patches (Mears et al., 2019). This finding confirms our earlier hypothesis that lack of greenspaces in the neighborhoods have a negative influence of colon cancer patients. As Mary Soderstrom (Soderstrom, 2008) argues, increasing street greenery and number of greenspaces can make a difference, and create a dense and pleasurable city at the same time.

While access to and availability of greenspace is recognized as an (predominantly positive) influential factor on health conditions, the reduction of impervious surfaces and spaces occupied by oversized residential-area roads and gigantic parking lots might be as

important as the creation of green infrastructure through zoning or re-use of abandon lots. Unfortunately, this practice is rare in the U.S., but more attention must be given to the aesthetics of the neighborhoods. Therefore, we agree with Kondo et al (Kondo et al., 2018) that “urban planners and public health professionals need evidence of the impacts of specific therapeutic or place-based interventions to help address public health issues facing their constituents”(p.22) and argue that integration of landscape metrics into health disparities research may provide the required evidence.

Remote sensing-based classification is valuable tool for land cover analysis and can be customized depending on the research question and study area. Additionally, there is a growing amount of data on a high spatial resolution allowing fine classifications.

However, the land cover classification schema developed by the NLCD is informative for analyzing urban and urbanized areas, and offers enormous opportunities for integration of land cover data into cancer research also on nationwide scale, including non-contiguous states and Puerto Rico (Homer et al., 2012). Additionally, it is available for several time points, which allows temporal analysis and application of longitudinal study design like in our example. On the other hand, calculation of various landscape metrics allows a straightforward integration of several measures of landscape characteristics. These could then be helpful for city planning and the establishment of specific cancer prevention and control strategies.

Moreover, while aesthetical value of a neighborhood is a very subjective measure and typically requires qualitative interviews or surveys, some landscape metrics offer an opportunity for quantification. The contiguity or aggregation indices used here may provide information about landscape configuration (Riitters et al., 1995; Turner et al.,

2001), and become an alternative to more complicated measures derived from Google street view which are time consuming and more expensive to process.

The present study has several limitations. While we found evidence for associations between the land cover configuration and risk of death from colon cancer, the results may not be generalizable to other states with different demographic, socioeconomic or landscape characteristics because the study population was limited to New Jersey only. Additionally, we did not have access to individual-level factors such as individual SES, general health conditions (e.g., obesity data) and behavior. Not only could these factors potentially confound the relationship between landscape characteristics and colon cancer survival, but accounting for individual-level factors may reduce the geographic variance explained by neighborhood and landscape characteristics (Doubeni et al., 2012).

Another limitation is that we restricted our study population to regional-stage colon cancer cases and followed patients for only 10 years after diagnosis. This was necessary to minimize extreme variations in survival and limit sources of variation in residential history measures. Therefore, analyzing colon cancer cases diagnosed in earlier (local) or later (metastatic) stages of the disease may not result in same conclusion.

Moreover, we utilized only residential histories collected from LexisNexis. We did not have access to self-reported information and could not validate and/or augment LexisNexis data. However, according to previous studies, the concordance between LexisNexis addresses and addresses collected from study participants (85%-86%) is high (S. Hurley et al., 2017; D. C. Wheeler & Wang, 2015). In our study, we could only validate the residential location at the time of diagnosis between the LexisNexis and the NJSCR. The concordance rate was approximately 83%. Opening to a 6-month window

before and after the diagnosis date substantially increased the concordance rate to 93%. Only 8% of the locations from LexisNexis cases did not match any locations reported by the NJSCR. Several factors such as incorrect links at LexisNexis, incorrect geocodes assigned by NJSCR for both registry and LexisNexis residential addresses, incorrect addresses reported to the registry by hospitals and other reporting facilities or geocodes assigned to addresses based on post office boxes could be the reason for address discordance. However, the extent of the bias in either direction would be minimal because of the low proportion of affected cases. In contrast to the residential histories from self-reported data, LexisNexis is an objective source of residential history data, and is not sensitive to potential recall bias.

Additionally, the application of landscape metrics in neighborhood research is not typical and is more common in natural landscapes for ecological analysis and modeling. In this study, we used Census tract boundaries defining neighborhoods or landscapes. The definition of a landscape as a Census tract is a subject of modified area unit problem. The selection of the landscape boundaries is essential for the calculation of landscape metrics and change in area size and shape will affect the values of multiple indices. Defining landscapes through a use of other administrative boundaries or grid system may result in different conclusions. However, Census tract is a common unit of analysis in public health research and allows an uncomplicated merge of data from various sources.

Lastly, the spatial resolution of the land cover classification raster was 30m and suggests that all features within an area of 900m² are generalized and defined as one class. The establishment of land cover classification with a fine spatial resolution would result in higher precision in the classification of the ground objects. However, 30m spatial

resolution is widely used in remote sensing discipline for land cover classification, and the utilized NLCD dataset is a well-known high-quality product.

The calculation of landscape metrics can be done on any spatial resolution, but values may vary with change in pixel size. More challenging is the selection of landscape metrics itself. There are many measures on various geographic levels. Thereby, most metrics are highly correlated and redundant. Previous research in landscape ecology suggests that a minimal number of metrics (e.g. number of patch types, mean edge/area ratio, contagion, average patch shape, fractal measurements) would be sufficient to quantify spatial heterogeneity (H Li & Reynolds, 1995; Habin Li & Reynolds, 1994; Riitters et al., 1995). However, in a public health context, the selection of landscape metrics should be done more careful, selecting meaningful variables that can be easily translated to the policy makers and urban planners. Selection of other landscape metrics could result in different associations or cause complication in result interpretation.

Conclusion

The associations between neighborhood SES and cancer survival are fairly established, where increased risk of death is often associated with high neighborhood poverty. However, neighborhood environment is not limited to socio-demographic factors because buildings, roads, greenspaces, and other human-made objects dominate landscapes. Especially in urban and urbanized areas. It is essential to understand the relationship between landscape characteristics and health outcomes in order to develop new policies in urban planning and design essential for population health, especially in urban and urbanized areas. Our results suggest that increasing proportions and connectivity of urban greenspaces may substantially decrease the risk of colon cancer death. This association

did not change even after adjusting for neighborhood SES, which is reported to be associated with a lack of greenspaces in urban and suburban areas (Wen, Zhang, Harris, Holt, & Croft, 2013), and reflected in our correlation analysis. The integration of remote sensing-based products, the NLCD and calculation of landscape metrics allow the exploration of undiscovered pathways between neighborhood characteristics and colon cancer survival and should be further evaluated in neighborhood studies in relation to other cancer sites and outcomes such as stage at the diagnosis. Additionally, further research is needed to understand how these specific landscape characteristics impact survival, and to evaluate opportunities for developing a socio-environmental deprivation index combining Census-based variables and land cover metrics in order to identify neighborhoods in need of interventions. Moreover, future studies should include additional neighborhood variables, especially related to walkability that could help to evaluate the association between neighborhood built environment and colon cancer survival.

CHAPETR 5:

CONCLUSION

This dissertation aimed to address a significant gap in geospatial cancer research related to describing a cancer case's neighborhood environment using only a single location based on the residential place at time of diagnosis. I utilized a unique dataset that included the residential histories of colon cancer cases diagnosed in New Jersey between 2006 and 2011. The colon cancer cases' residential histories were obtained through a data linkage with LexisNexis, a unique residential history database. By incorporating residential histories, I moved beyond a cross-sectional approach. More specifically, this dissertation examined how residential histories and socio-spatial mobility can change the case's geographic context over time and influence patient survival.

Geographers, cancer researchers and social epidemiologists in recent years have taken a keen interest in incorporating life course approaches when studying cancer risk and outcomes like the stage of disease at diagnosis and survival. There have been calls from leading scientists to incorporate longitudinal geographic factors and adding residential histories to cancer registry data. Incorporating residential histories is not a trivial task, as it requires coordination with state cancer registries and the use of more sophisticated non-spatial and spatial models to analyze these data. In this dissertation, I was able to demonstrate that linkages between residential histories data and state cancer registries are possible. I found the residential histories reported for more than 98% of the cases in our datasets. Furthermore, I was able to demonstrate in our three case studies how these data

can be used to expand the standard cross-sectional study design to examine geographic variation in colon cancer survival.

The findings from this dissertation provided new and important information about how changes in residence impact cancer survival. My first study examined whether the risk-of-death (i.e., survival) estimates by neighborhood poverty level were different if the patients' residential histories were used to measure poverty instead of the more commonly used residence at the time of diagnosis. I found that the hazard ratios were generally similar in strength and direction across models, regardless of whether the CT-poverty measures were included as a continuous or categorical variable. Finding no major differences in whether residential histories were included is likely a result of minimal socio-spatial mobility (i.e., cases generally moving to census tracts with similar poverty levels). While the study population was mobile, with approximately 35% changing census tract one or more times after diagnosis, most movers went to areas of similar poverty levels. However, this finding may not be valid for other neighborhood factors. More research is necessary on models when using alternative measures with different spatial and temporal patterns, such as segregation, ethnic enclaves, and homeownership, that could possibly reveal a clearer association with cancer survival.

Another important finding was related to the mover status of the patients. I found an increased risk of death among movers residing in high-poverty areas. Without residential histories, this finding would remain undiscovered, because there was no significant association between CT-poverty and the risk of death in the models when using residence at time of diagnosis only. Future studies should consider examining cancer disparities specifically among movers to better understand this relationship.

My second study focused on whether including the residential histories of the colon cancer cases in the spatial models changed the geographic patterns of the risk-of-death estimates compared to the models where only the location at the time of diagnosis was used. I found that the time-varying model produced a wider range of adjusted risk-of-death estimates compared to the model with only the residence at the time of diagnosis (HR = 0.85 - 1.20 vs. HR = 0.94 - 1.11). This also resulted in greater statewide geographic disparity (25.5% vs. 14.2%) after adjustment for individual-level factors and residential mobility. Although the general results based on the traditional approach of using only residence at the time of diagnosis were not substantially different from the time-varying model, the inclusion of residential histories may allow for more precise estimates of spatial risk of death.

When I did not account for residential mobility, the geographic risk of death might be underestimated in several areas. Residential histories allowed me to evaluate the proportion of geographic bias in spatial risk of death estimates, suggesting that approximately 12% of the initial study population have moved the state where they had been residing at the time of diagnosis during the follow-up period. Thereby, there are variations in the destination, depending on the SES of the patient. In general, those moving out of state were arriving from more affluent areas and had better 10-year relative survival rates than those who moved in state and non-movers. Therefore, the relocation of wealthy patients with longer survival times to areas outside New Jersey may result in the overall decrease of the state-wide survival rates.

Finally, the third chapter focused on whether landscape characteristics (e.g., built environment) were associated with the risk of death from colon cancer independent of

individual-level factors, including residential mobility and neighborhood poverty. In the literature, the definition of the neighborhood effects frequently assumes associations between neighborhood SES or deprivation and some health outcome. While neighborhood SES (e.g., census tract poverty level) is often an important predictor of health disparities, our results suggest that a neighborhood's landscape characteristics related to the built environment and land-cover patterns are also influential and require further attention. For example, I found that an increasing proportion of high-intensity-developed lands almost doubled the risk of death, while an increase in the aggregation and connectivity of vegetation-dominated, low-intensity-developed lands reduced the risk of death by 4% for every 10% increase in greenspace cover. These relationships might be a result of the availability and extent of greenspaces in urban and urbanized neighborhoods. Less greenspaces in these areas could reduce access to recreational sites that promote physical activity. Furthermore, places that lack greenspaces may evoke negative emotions and decrease psychological well-being.

The reduction in impervious surfaces in residential areas may increase their aesthetic value and provide conditions more advantageous to a healthy lifestyle such as walking, which could improve patient survival. To address this, landscape design and land-cover zoning of numerous cities must be revisited. Several international urban-planning projects are known for their success in urban revitalization and greening, which increased the neighborhood living quality and met several sustainability goals, including providing healthy environments (Kemp & Stephani, 2013). Creating more public space for people than for cars (Houstoun, 2013) and increasing walkability in American cities (Soderstrom, 2008) may help to promote more active and healthy lifestyles, increase

diversity of neighborhoods, prevent the concentration of poverty, and therefore reduce cancer mortality. A study from Pittsburgh (Pennsylvania) showed that improving street walkability and aesthetics increase park/greenspace use, even in low-income neighborhoods (Andrea S. Richardson et al., 2020).

Another contribution from my dissertation relates to how data fusion and adaptation of methodologies from other disciplines can increase the potential for researchers to better understand the role of the neighborhood environment and the underlying pathways that impact cancer survival. For example, I showed that the integration of land-cover data into neighborhood studies is feasible and could expand our knowledge about neighborhood effects in colon cancer survival. While previous studies have produced many different deprivation measures (Gomez et al., 2011; Gomez et al., 2015), landscape ecology offers alternatives that may help develop a more comprehensive socio-environmental deprivation index, which would capture the complexity of interactions between human and built environments.

Future research should focus on more frequent integration of residential histories or at least design time-varying models by including annual values of neighborhood factors (e.g., census tract poverty) based on the place of residence at the time of diagnosis.

Additionally, the landscape metrics introduced in this dissertation should be further evaluated in other study areas and cancer sites and outcomes to confirm and test these associations.

While each chapter includes a summary of major limitations, it is worthwhile mentioning a few of those here. First, patients were followed for only approximately 10 years after the diagnosis. This could have influenced the trajectory patterns of the socio-spatial

mobility of patients. During this time, a majority (65%) of patients remained in the same census tract. Following patients for a longer time may result in different findings.

Second, all case studies relied solely on residential histories collected from LexisNexis and did not include self-reported residential histories that may be used to validate and/or augment LexisNexis data. While studies have noted good concordance (85% - 86%) with self-reported data (S. Hurley et al., 2017; D. C. Wheeler & Wang, 2015), errors in the data could remain without a thorough evaluation. The residential histories timeline was developed based on the newest-wins algorithm, which assumes that when cases overlap temporally, the most recent address is used. Additionally, all residential locations were based on census tracts and may not capture short-distance moves within census tracts.

Future studies should repeat these analyses using the residential address.

Finally, the landscape metrics are calculated on the level of the census tract. However, considering the variation in area size and perimeter among census tracts, results may vary when other landscape definitions (e.g., census blocks or counties) are used. This issue is commonly referred to as the modifiable areal unit problem (Buzzelli, 2020; Pawitan & Steel, 2009; Sahar et al., 2019). Additionally, in general, administrative boundaries are not an optimal landscape definition, because most individuals would move across borders and perceive their neighborhoods differently. Developing a spatial grid system could be a more objective landscape definition and make the landscape metrics more comparable throughout the study area.

This dissertation expands our knowledge about the geographic colon cancer disparities in New Jersey and provides examples of the integration of residential histories and remote sensing-based products into cancer disparities research. While the presented studies are

among the first to include this unique dataset, in the upcoming years, studies will increasingly use population-based cancer registry data linked to commercial residential histories database such as LexisNexis. Including residential histories opens new avenues of inquiry to better understand the complex relationships between people and places, and the effect of residential mobility on cancer outcomes. The integration of non-census-based neighborhood variables will become more frequent with the advances in land-cover classification and feature extraction techniques. Combining multiple socio-demographic and environmental domains to estimate the neighborhood effect on cancer outcomes will increase the potential to understand the underlying pathways.

REFERENCES CITED

- Aarts, M. J., Lemmens, V. E. P. P., Louwman, M. W. J., Kunst, A. E., & Coebergh, J. W. W. (2010). Socioeconomic status and changing inequalities in colorectal cancer? A review of the associations with risk, treatment and outcome. *European Journal of Cancer*, 46(15), 2681-2695. doi:<https://doi.org/10.1016/j.ejca.2010.04.026>
- American Cancer Society. (2019). *Cancer Facts & Figures 2019*. Retrieved from Atlanta.
- American Cancer Society. (2020). *Colorectal Cancer Facts & Figures 2020-2022*. Retrieved from Atlanta.
- Adler, D., Kneib, T., Lang, S., Umlauf, N., & Zeileis, A. (2012). BayesXsrc: R Package Distribution of the BayesX C++ Sources. *R package version*, 2, 1-1.
- Akpinar, A., Barbosa-Leiker, C., & Brooks, K. R. (2016). Does green space matter? Exploring relationships between green space type and health indicators. *Urban Forestry & Urban Greening*, 20, 407-418. doi:<https://doi.org/10.1016/j.ufug.2016.10.013>
- Allemani, C., Harewood, R., Johnson, C. J., Carreira, H., Spika, D., Bonaventure, A., . . . Coleman, M. P. (2017). Population-based cancer survival in the United States: Data, quality control, and statistical methods. *Cancer*, 123(S24), 4982-4993. doi:10.1002/cncr.31025
- Ashing-Giwa, K. T., Lim, J.-w., & Tang, J. (2010). Surviving cervical cancer: does health-related quality of life influence survival? *Gynecologic oncology*, 118(1), 35-42.
- Audy, J. R. (1954). A Biological Approach to Medical Geography. *British Medical Journal*.
- Banzhaf, H. S., & Walsh, R. P. (2008). Do people vote with their feet? An empirical test of Tiebout. *American Economic Review*, 98(3), 843-863.
- Baugh, D. K., & Verghese, S. (2013). Migration patterns for Medicaid enrollees 2005-2007. *Medicare Medicaid Res Rev*, 3(4). doi:10.5600/mmrr.003.04.b02
- Belitz, C., Brezger, A., Klein, N., Kneib, T., Lang, S., & Umlauf, N. (2015). BayesX: Methodology Manual; 2015. *There is no corresponding record for this reference*.
- Bergström, L., & Van Ham, M. (2010). Understanding neighbourhood effects: Selection bias and residential mobility.

- Berkman, L. F., Kawachi, I., & Glymour, M. M. (2014). *Social epidemiology*: Oxford University Press.
- Beyer, K. M. M., Tiwari, C., & Rushton, G. (2012). Five Essential Properties of Disease Maps. *Annals of the Association of American Geographers*, 102(5), 1067-1075. doi:10.1080/00045608.2012.659940
- Böheim, R., & Taylor, M. P. (2007). From the dark end of the street to the bright side of the road? The wage returns to migration in Britain. *Labour Economics*, 14(1), 99-117.
- Bolan, M. (1997). The mobility experience and neighborhood attachment. *Demography*, 34(2), 225-237.
- Boscoe, F. P. (2010). Towards the use of a census tract poverty indicator variable in cancer surveillance. *J Registry Manag*, 37(4), 148-151.
- Boscoe, F. P. (2011). The Use of Residential History in Environmental Health Studies. In J. A. Maantay & S. McLafferty (Eds.), *Geospatial Analysis of Environmental Health* (pp. 93-110). Dordrecht: Springer Netherlands.
- Boscoe, F. P., Ward, M. H., & Reynolds, P. (2004). Current practices in spatial analysis of cancer data: data characteristics and data sources for geographic studies of cancer. *International Journal of Health Geographics*, 3(1), 28. doi:10.1186/1476-072X-3-28
- Bourdieu, P. (1986). The Forms of Capital. In J. Richardson (Ed.), *Handbook of Theory and Research for the Sociology of Education* ((pp. 241-258). Westport, CT: Greenwood.
- Bratman, G. N., Anderson, C. B., Berman, M. G., Cochran, B., De Vries, S., Flanders, J., . . . Hartig, T. (2019). Nature and mental health: An ecosystem service perspective. *Science advances*, 5(7), eaax0903.
- Brett, J. M. (1982). Job transfer and well-being. *Journal of Applied Psychology*, 67(4), 450.
- Brezger, A., & Lang, S. (2006). Generalized structured additive regression based on Bayesian P-splines. *Computational Statistics & Data Analysis*, 50(4), 967-991.
- Brokamp, C., LeMasters, G. K., & Ryan, P. H. (2016). Residential mobility impacts exposure assessment and community socioeconomic characteristics in longitudinal epidemiology studies. *Journal Of Exposure Science And Environmental Epidemiology*, 26, 428. doi:10.1038/jes.2016.10

- Brown, J. C., Winters-Stone, K., Lee, A., & Schmitz, K. H. (2012). Cancer, physical activity, and exercise. *Compr Physiol*, 2(4), 2775-2809. doi:10.1002/cphy.c120005
- Buzzelli, M. (2020). Modifiable Areal Unit Problem. *International Encyclopedia of Human Geography*, 169-173. doi:10.1016/B978-0-08-102295-5.10406-8
- Cabo, J., Shu, X., Shu, X. O., Parikh, A., & Bailey, C. (2020). Treatment at Academic Centers Decreases Insurance-Based Survival Disparities in Colon Cancer. *J Surg Res*, 245, 265-272. doi:10.1016/j.jss.2019.07.059
- Callaghan, C. T., Bino, G., Major, R. E., Martin, J. M., Lyons, M. B., & Kingsford, R. T. (2019). Heterogeneous urban green areas are bird diversity hotspots: insights using continental-scale citizen science data. *Landscape Ecology*, 34(6), 1231-1246. doi:10.1007/s10980-019-00851-6
- Carmichael, H., Cowan, M., McIntyre, R., & Velopulos, C. (2020). Disparities in colorectal cancer mortality for rural populations in the United States: Does screening matter? *The American Journal of Surgery*, 219(6), 988-992. doi:<https://doi.org/10.1016/j.amjsurg.2019.09.027>
- Chen, Y., Stephens, M., & Jones, C. A. (2019). Does residents' satisfaction with the neighbourhood environment relate to residents' self-rated health? Evidence from Beijing. *International Journal of Environmental Research and Public Health*, 16(24), 5051.
- Chien, L. C., Schootman, M., & Pruitt, S. L. (2013). The modifying effect of patient location on stage-specific survival following colorectal cancer using geosurvival models. *Cancer Causes Control*, 24(3), 473-484. doi:10.1007/s10552-012-0134-4
- Chittleborough, C. R., Baum, F., Taylor, A., & Hiller, J. E. (2006). A life-course approach to measuring socioeconomic position in population health surveillance systems. *Journal of Epidemiology & Community Health*, 60(11), 981-992.
- Clark, W. A., Van Ham, M., & Coulter, R. (2014). Spatial mobility and social outcomes. *Journal of Housing and the Built Environment*, 29(4), 699-727.
- Clark, W. A. V., & Morrison, P. S. (2012). Socio-spatial Mobility and Residential Sorting: Evidence from a Large-scale Survey. *Urban Studies*, 49(15), 3253-3270. doi:10.1177/0042098012442418
- Cole, A. M., Jackson, J. E., & Doescher, M. (2013). Colorectal Cancer Screening Disparities for Rural Minorities in the United States. *Journal of Primary Care & Community Health*, 4(2), 106-111. doi:10.1177/2150131912463244
- Cornwell, E. Y., & Waite, L. J. (2009). Social disconnectedness, perceived isolation, and health among older adults. *Journal of Health and Social Behavior*, 50(1), 31-48.

- Coulter, R., & Scott, J. (2015). What Motivates Residential Mobility? Re-examining Self-Reported Reasons for Desiring and Making Residential Moves. *Population, Space and Place*, 21(4), 354-371. doi:10.1002/psp.1863
- Cunningham, R., Sarfati, D., Stanley, J., Peterson, D., & Collings, S. (2015). Cancer survival in the context of mental illness: a national cohort study. *General Hospital Psychiatry*, 37(6), 501-506. doi:<https://doi.org/10.1016/j.genhosppsych.2015.06.003>
- Demoury, C., Thierry, B., Richard, H., Sigler, B., Kestens, Y., & Parent, M.-E. (2017). Residential greenness and risk of prostate cancer: A case-control study in Montreal, Canada. *Environment international*, 98, 129-136.
- Diez Roux, A. V. (2001). Investigating neighborhood and area effects on health. *American Journal of Public Health*, 91(11), 1783-1789.
- Diez Roux, A. V., & Mair, C. (2010). Neighborhoods and health. *Annals of the New York Academy of Sciences*, 1186(1), 125-145. doi:10.1111/j.1749-6632.2009.05333.x
- Doubeni, C. A., Major, J. M., Laiyemo, A. O., Schootman, M., Zauber, A. G., Hollenbeck, A. R., . . . Allison, J. (2012). Contribution of Behavioral Risk Factors and Obesity to Socioeconomic Differences in Colorectal Cancer Incidence. *JNCI: Journal of the National Cancer Institute*, 104(18), 1353-1362. doi:10.1093/jnci/djs346
- Douven, W., & Scholten, H. J. (1995). Spatial analysis in health research *The Added Value of Geographical Information Systems in Public and Environmental Health* (pp. 117-133): Springer.
- Dramstad, W. E., Tveit, M. S., Fjellstad, W., & Fry, G. L. (2006). Relationships between visual landscape preferences and map-based indicators of landscape structure. *Landscape and Urban Planning*, 78(4), 465-474.
- Du, X. L., Fang, S., Vernon, S. W., El-Serag, H., Shih, Y. T., Davila, J., & Rasmus, M. L. (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. doi:10.1002/cncr.22826
- Dunn, E. C., Winning, A., Zaika, N., & Subramanian, S. V. (2014). Does poor health predict moving, move quality, and desire to move?: A study examining neighborhood selection in US adolescents and adults. *Health Place*, 30, 154-164. doi:10.1016/j.healthplace.2014.08.007
- Emch, M., Root, E. D., & Carrel, M. (2017). *Health and Medical Geography* (4th ed.). Guilford Publications. New York, London

- Eriksson, M. (2011). Social capital and health--implications for health promotion. *Glob Health Action*, 4, 5611. doi:10.3402/gha.v4i0.5611
- Espejo-Herrera, N., Gràcia-Lavedan, E., Boldo, E., Aragonés, N., Pérez-Gómez, B., Pollán, M., . . . La Vecchia, C. (2016). Colorectal cancer risk and nitrate exposure through drinking water and diet. *International journal of cancer*, 139(2), 334-346.
- Fahrmeir, L., & Lang, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(2), 201-220.
- Feinglass, J., Rydzewski, N., & Yang, A. (2015). The socioeconomic gradient in all-cause mortality for women with breast cancer: findings from the 1998 to 2006 National Cancer Data Base with follow-up through 2011. *Annals of Epidemiology*, 25(8), 549-555. doi:10.1016/j.annepidem.2015.02.006
- Gallagher, L. G., Webster, T. F., Aschengrau, A., & Vieira, V. M. (2010). Using residential history and groundwater modeling to examine drinking water exposure and breast cancer. *Environ Health Perspect*, 118(6), 749-755. doi:10.1289/ehp.0901547
- Garvin, E., Branas, C., Keddem, S., Sellman, J., & Cannuscio, C. (2013). More Than Just An Eyesore: Local Insights And Solutions on Vacant Land And Urban Health. *Journal of Urban Health*, 90(3), 412-426. doi:10.1007/s11524-012-9782-7
- Geronimus, A. T., Bound, J., & Ro, A. (2014). Residential mobility across local areas in the United States and the geographic distribution of the healthy population. *Demography*, 51(3), 777-809.
- Gesler, W. M. (1992). Therapeutic landscapes: Medical issues in light of the new cultural geography. *Social Science & Medicine*, 34(7), 735-746. doi:10.1016/0277-9536(92)90360-3
- Gillespie, B. J. (2016). *Household mobility in America: Patterns, processes, and outcomes*: Springer.
- Gilliland, J. A. (1998). Modeling residential mobility in Montreal, 1860–1900. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 31(1), 27-42.
- Glass, G. E. (2000). Update: spatial aspects of epidemiology: the interface with medical geography. *Epidemiol Rev*, 22(1), 136-139.
- Glennerster, H., Lupton, R., Noden, P., & Power, A. (1999). Poverty, Social Exclusion and Neighbourhood: Studying the area bases of social exclusion. *LSE STICERD Research Paper No. CASE022*.

- Gomez, S. L., Glaser, S. L., McClure, L. A., Shema, S. J., Kealey, M., Keegan, T. H. M., & Satariano, W. A. (2011). The California Neighborhoods Data System: a new resource for examining the impact of neighborhood characteristics on cancer incidence and outcomes in populations. *Cancer Causes & Control*, 22(4), 631-647. doi:10.1007/s10552-011-9736-5
- Gomez, S. L., Shariff-Marco, S., DeRouen, M., Keegan, T. H. M., Yen, I. H., Mujahid, M., . . . Glaser, S. L. (2015). The impact of neighborhood social and built environment factors across the cancer continuum: Current research, methodological considerations, and future directions. *Cancer*, 121(14), 2314-2330. doi:10.1002/cncr.29345
- Greene, F. L., Balch, C. M., Fleming, I. D., Fritz, A., Haller, D. G., Morrow, M., & Page, D. L. (2002). *AJCC cancer staging handbook: TNM classification of malignant tumors*: Springer Science & Business Media.
- Greenfield, E. J., Nowak, D. J., & Walton, J. T. (2009). Assessment of 2001 NLCD percent tree and impervious cover estimates. *Photogrammetric Engineering & Remote Sensing*, 75(11), 1279-1286.
- Haggar, F. A., & Boushey, R. P. (2009). Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clinics in colon and rectal surgery*, 22(4), 191-197. doi:10.1055/s-0029-1242458
- Han, D., Rogerson, P. A., Nie, J., Bonner, M. R., Vena, J. E., Vito, D., . . . Freudenheim, J. L. (2004). Geographic clustering of residence in early life and subsequent risk of breast cancer (United States). *Cancer Causes & Control*, 15(9), 921-929.
- Henry, K. A., Niu, X., & Boscoe, F. P. (2009). Geographic disparities in colorectal cancer survival. *International Journal of Health Geographics*, 8(1), 48. doi:10.1186/1476-072x-8-48
- Henry, K. A., Sherman, R. L., McDonald, K., Johnson, C. J., Lin, G., Stroup, A. M., & Boscoe, F. P. (2014). Associations of census-tract poverty with subsite-specific colorectal cancer incidence rates and stage of disease at diagnosis in the United States. *J Cancer Epidemiol*, 2014, 823484. doi:10.1155/2014/823484
- Henry, K. A., Stroup, A. M., Warner, E. L., & Kepka, D. (2016). Geographic Factors and Human Papillomavirus (HPV) Vaccination Initiation among Adolescent Girls in the United States. *Cancer Epidemiology Biomarkers & Prevention*, 25(2), 309-317. doi:10.1158/1055-9965.epi-15-0658
- Hesselbarth, M. H., Sciaini, M., With, K. A., Wiegand, K., & Nowosad, J. (2019). landscapemetrics: an open-source R tool to calculate landscape metrics. *Ecography*, 42(10), 1648-1657.

- Hijmans, R. J., J. van Etten, J. Cheng, M. Mattiuzzi, M. Sumner, J. A. Greenberg, O. P. Lamigueiro, A. Bevan, E. B. Racine, and A. Shortridge. (Producer). (2016). Package 'raster'. *R package*. Retrieved from <https://cran.r-project.org/web/packages/raster/index.html>
- Homer, C. H., Fry, J. A., & Barnes, C. A. (2012). The national land cover database. *US Geological Survey Fact Sheet, 3020(4)*, 1-4.
- Houstoun, L. (2013). Manchester, England, Creates more public places for people rather than cars. In R. L. Kemp & C. J. Stephani (Eds.), *Global Models of Urban Planning: Best Practices Outside the United States*. Jefferson NC, London: McFarland.
- Howden-Chapman, P. (2004). Housing standards: a glossary of housing and health. *Journal of Epidemiology and Community Health, 58(3)*, 162-168. doi:10.1136/jech.2003.011569
- Hughes, A. E., & Pruitt, S. L. (2017). The utility of EMR address histories for assessing neighborhood exposures. *Ann Epidemiol, 27(1)*, 20-26. doi:10.1016/j.annepidem.2016.07.016
- Hulchanski, J. D. (2010). The three cities within Toronto. Cities Centre Press, University of Toronto. https://homelesshub.ca/sites/default/files/Three-Cities-Within-Toronto_2010_Final.pdf
- Hunter, M. C. R., & Brown, D. G. (2012). Spatial contagion: Gardening along the street in residential neighborhoods. *Landscape and Urban Planning, 105(4)*, 407-416.
- Hurley, S., Hertz, A., Nelson, D. O., Layefsky, M., Von Behren, J., Bernstein, L., . . . Reynolds, P. (2017). Tracing a path to the past: exploring the use of commercial credit reporting data to construct residential histories for epidemiologic studies of environmental exposures. *American Journal of Epidemiology, 185(3)*, 238-246.
- Hurley, S., Hertz, A., Nelson, D. O., Layefsky, M., Von Behren, J., Bernstein, L., . . . Reynolds, P. (2017). Tracing a Path to the Past: Exploring the Use of Commercial Credit Reporting Data to Construct Residential Histories for Epidemiologic Studies of Environmental Exposures. *Am J Epidemiol, 185(3)*, 238-246. doi:10.1093/aje/kww108
- Hystad, P., Demers, P. A., Johnson, K. C., Carpiano, R. M., & Brauer, M. (2013). Long-term residential exposure to air pollution and lung cancer risk. *Epidemiology, 24(5)*, 762-772. doi:10.1097/EDE.0b013e3182949ae7
- Ittelson, W. H. (1978). Environmental Perception and Urban Experience. *Environment and Behavior, 10(2)*, 193-213. doi:10.1177/0013916578102004

- Iverson Nassauer, J. (1995). Culture and changing landscape structure. *Landscape Ecology*, 10(4), 229-237. doi:10.1007/BF00129257
- Iyer, H. S., Valeri, L., James, P., Chen, J. T., Hart, J. E., Laden, F., . . . Rebbeck, T. R. (2020). The contribution of residential greenness to mortality among men with prostate cancer: a registry-based cohort study of Black and White men. *Environmental epidemiology (Philadelphia, Pa.)*, 4(2), e087-e087. doi:10.1097/EE9.0000000000000087
- Jackson, L. E. (2003). The relationship of urban design to human health and condition. *Landscape and Urban Planning*, 64(4), 191-200.
- Jacobs, J. A. (1990). The sex segregation of occupations as a circulating system. In R. L. Breiger (Ed.), *Social mobility and social structure* (pp. 27-58). Cambridge, New York, Port Chester, Melbourne, Sydney: Cambridge University Press.
- Jacquez, G. M., Barlow, J., Rommel, R., Kaufmann, A., Rienti, M., AvRuskin, G., & Rasul, J. (2014). Residential Mobility and Breast Cancer in Marin County, California, USA. *International Journal of Environmental Research and Public Health*, 11(1), 271-295. doi:10.3390/ijerph110100271
- Jacquez, G. M., Kaufmann, A., Meliker, J., Goovaerts, P., AvRuskin, G., & Nriagu, J. (2005). Global, local and focused geographic clustering for case-control data with residential histories. *Environ Health*, 4(1), 4. doi:10.1186/1476-069X-4-4
- Jacquez, G. M., Meliker, J., & Kaufmann, A. (2007). In search of induction and latency periods: Space-time interaction accounting for residential mobility, risk factors and covariates. *International Journal of Health Geographics*, 6(1), 35.
- Jacquez, G. M., Meliker, J. R., Avruskin, G. A., Goovaerts, P., Kaufmann, A., Wilson, M. L., & Nriagu, J. (2006). Case-control geographic clustering for residential histories accounting for risk factors and covariates. *Int J Health Geogr*, 5, 32. doi:1476-072X-5-32 [pii]10.1186/1476-072X-5-32
- Jacquez, G. M., Slotnick, M. J., Meliker, J. R., AvRuskin, G., Copeland, G., & Nriagu, J. (2010). Accuracy of commercially available residential histories for epidemiologic studies. *American Journal of Epidemiology*, 173(2), 236-243.
- Jacquez, G. M., Slotnick, M. J., Meliker, J. R., AvRuskin, G., Copeland, G., & Nriagu, J. (2011). Accuracy of commercially available residential histories for epidemiologic studies. *Am J Epidemiol*, 173(2), 236-243. doi:kwq350 [pii]10.1093/aje/kwq350
- James, P., Hart, J. E., Banay, R. F., & Laden, F. (2016). Exposure to greenness and mortality in a nationwide prospective cohort study of women. *Environmental Health Perspectives*, 124(9), 1344-1352.

- Jelleyman, T., & Spencer, N. (2008). Residential mobility in childhood and health outcomes: a systematic review. *Journal of Epidemiology & Community Health*, 62(7), 584-592.
- Jennings, V., Baptiste, A. K., Jelks, O., & Skeete, R. (2017). Urban green space and the pursuit of health equity in parts of the United States. *International Journal of Environmental Research and Public Health*, 14(11), 1432.
- Jerant, A. F., Fenton, J. J., & Franks, P. (2008). Determinants of racial/ethnic colorectal cancer screening disparities. *Archives of internal medicine*, 168(12), 1317-1324.
- Jin, X., Carlin, B. P., & Banerjee, S. (2005). Generalized Hierarchical Multivariate CAR Models for Areal Data. *Biometrics*, 61(4), 950-961. doi:10.1111/j.1541-0420.2005.00359.x
- Kafadar, K. (1996). Smoothing geographical data, particularly rates of disease. *Statistics in Medicine*, 15(23), 2539-2560.
- Kafadar, K. (1997). Geographic trends in prostate cancer mortality: an application of spatial smoothers and the need for adjustment. *Annals of Epidemiology*, 7(1), 35-45.
- Kafadar, K. (1999). Simultaneous smoothing and adjusting mortality rates in US counties: melanoma in white females and white males. *Statistics in Medicine*, 18(23), 3167-3188.
- Kafadar, K. (2014). Smoothing. *Wiley StatsRef: Statistics Reference Online*.
- Kaplan, S. (1987). Aesthetics, Affect, and Cognition: Environmental Preference from an Evolutionary Perspective. *Environment and Behavior*, 19(1), 3-32. doi:10.1177/0013916587191001
- Keegan, T. H. M., Shariff-Marco, S., Sangaramoorthy, M., Koo, J., Hertz, A., Schupp, C. W., . . . Gomez, S. L. (2014). Neighborhood influences on recreational physical activity and survival after breast cancer. *Cancer Causes & Control*, 25(10), 1295-1308. doi:10.1007/s10552-014-0431-1
- Kemp, R. L., & Stephani, C. J. (2013). *Global Models of Urban Planning: Best Practices Outside the United States*. Jefferson NC, London: McFarland.
- Kneib, T., & Fahrmeir, L. (2007). A Mixed Model Approach for Geoaddditive Hazard Regression. *Scandinavian Journal of Statistics*, 34(1), 207-228.
- Kondo, M. C., Fluehr, J. M., McKeon, T., & Branas, C. C. (2018). Urban green space and its impact on human health. *International Journal of Environmental Research and Public Health*, 15(3), 445.

- Krieger, N., Chen, J. T., Waterman, P. D., Rehkopf, D. H., & Subramanian, S. V. (2005). Painting a Truer Picture of US Socioeconomic and Racial/Ethnic Health Inequalities: The Public Health Disparities Geocoding Project. *American Journal of Public Health, 95*(2), 312-323. doi:10.2105/ajph.2003.032482
- Krieger, N., Chen, J. T., Waterman, P. D., Soobader, M.-J., Subramanian, S., & Carson, R. (2003). Choosing area based socioeconomic measures to monitor social inequalities in low birth weight and childhood lead poisoning: The Public Health Disparities Geocoding Project (US). *Journal of Epidemiology & Community Health, 57*(3), 186-199.
- Krieger, N., Chen, J. T., Waterman, P. D., Soobader, M.-J., Subramanian, S. V., & Carson, R. (2002). Geocoding and Monitoring of US Socioeconomic Inequalities in Mortality and Cancer Incidence: Does the Choice of Area-based Measure and Geographic Level Matter?: The Public Health Disparities Geocoding Project. *American Journal of Epidemiology, 156*(5), 471-482. doi:10.1093/aje/kwf068
- Krieger, N., Quesenberry, C., Peng, T., Horn-Ross, P., Stewart, S., Brown, S., . . . Alvarez-Martinez, L. (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.
- Kwan, M.-P. (2012). The Uncertain Geographic Context Problem. *Annals of the Association of American Geographers, 102*(5), 958-968. doi:10.1080/00045608.2012.687349
- Labianca, R., Beretta, G. D., Kildani, B., Milesi, L., Merlin, F., Mosconi, S., . . . Wils, J. (2010). Colon cancer. *Critical reviews in oncology/hematology, 74*(2), 106-133. doi:<https://doi.org/10.1016/j.critrevonc.2010.01.010>
- Lang, S., & Brezger, A. (2004). Bayesian P-splines. *Journal of computational and graphical statistics, 13*(1), 183-212.
- Lawson, A., Biggeri, A., Böhning, D., Lesaffre, E., Viel, J.-F., & Bertollini, R. (1999). *Disease mapping and risk assessment for public health*: Wiley New York.
- Lawson, A. B. (2001). Disease map reconstruction. *Statistics in Medicine, 20*(14), 2183-2204.
- Lawson, A. B., Williams, F. L., & Williams, F. (2001). *An introductory guide to disease mapping*: John Wiley.
- Lee, S.-W., Ellis, C. D., Kweon, B.-S., & Hong, S.-K. (2008). Relationship between landscape structure and neighborhood satisfaction in urbanized areas. *Landscape and Urban Planning, 85*(1), 60-70.

- Li, H., & Reynolds, J. (1995). On definition and quantification of heterogeneity. *Oikos*, 280-284.
- Li, H., & Reynolds, J. F. (1994). A simulation experiment to quantify spatial heterogeneity in categorical maps. *Ecology*, 75(8), 2446-2455.
- Lian, M., Schootman, M., Doubeni, C. A., Park, Y., Major, J. M., Stone, R. A., . . . Schatzkin, A. (2011). Geographic variation in colorectal cancer survival and the role of small-area socioeconomic deprivation: a multilevel survival analysis of the NIH-AARP Diet and Health Study Cohort. *Am J Epidemiol*, 174(7), 828-838. doi:10.1093/aje/kwr162
- Lin, K. C., Huang, H. C., Bai, Y. M., & Kuo, P. C. (2012). Lifetime residential mobility history and self-rated health at midlife. *J Epidemiol*, 22(2), 113-122. doi:10.2188/jea.je20110055
- Ling, C., Heck, J. E., Cockburn, M., Liew, Z., Marcotte, E., & Ritz, B. (2019). Residential mobility in early childhood and the impact on misclassification in pesticide exposures. *Environmental Research*, 173, 212-220. doi:<https://doi.org/10.1016/j.envres.2019.03.039>
- Link, B. G., & Phelan, J. (1995). Social Conditions As Fundamental Causes of Disease. *Journal of Health and Social Behavior*, 80-94. doi:10.2307/2626958
- Litwak, E., & Longino Jr, C. F. (1987). Migration patterns among the elderly: A developmental perspective. *The Gerontologist*, 27(3), 266-272.
- Lix, L. M., Hinds, A., DeVerteuil, G., Robinson, J. R., Walker, J., & Roos, L. L. (2006). Residential mobility and severe mental illness: a population-based analysis. *Administration and Policy in Mental Health and Mental Health Services Research*, 33(2), 160.
- Long, L. (1988). *Migration and residential mobility in the United States*: Russell Sage Foundation.
- Low, E. E., Demb, J., Liu, L., Earles, A., Bustamante, R., Williams, C. D., . . . Gupta, S. (2020). Risk Factors for Early-Onset Colorectal Cancer. *Gastroenterology*, 159(2), 492-501.e497. doi:<https://doi.org/10.1053/j.gastro.2020.01.004>
- Lynch, S. M., Mitra, N., Ravichandran, K., Mitchell, J., Spangler, E., Zhou, W., . . . Rebbeck, T. R. (2017). Telomere Length and Neighborhood Circumstances: Evaluating Biological Response to Unfavorable Exposures. *Cancer Epidemiology Biomarkers & Prevention*, 26(4), 553-560. doi:10.1158/1055-9965.epi-16-0554
- Lynch, S. M., Mitra, N., Ross, M., Newcomb, C., Dailey, K., Jackson, T., . . . Rebbeck, T. R. (2017). A Neighborhood-Wide Association Study (NWAS): Example of

prostate cancer aggressiveness. *PLoS One*, 12(3), e0174548.
doi:10.1371/journal.pone.0174548

Lynch, S. M., & Rebbeck, T. R. (2013). Bridging the Gap between Biologic, Individual, and Macroenvironmental Factors in Cancer: A Multilevel Approach. *Cancer Epidemiology Biomarkers & Prevention*, 22(4), 485-495. doi:10.1158/1055-9965.Epi-13-0010

Maharana, A., & Nsoesie, E. O. (2018). Use of Deep Learning to Examine the Association of the Built Environment With Prevalence of Neighborhood Adult Obesity. *JAMA Network Open*, 1(4), e181535-e181535. doi:10.1001/jamanetworkopen.2018.1535

Mandelblatt, J., Andrews, H., Kao, R., Wallace, R., & Kerner, J. (1996). The late-stage diagnosis of colorectal cancer: demographic and socioeconomic factors. *American Journal of Public Health*, 86(12), 1794-1797.

Matsuoka, R., & Sullivan, W. (2011). Human Psychological Community Health. In I. Douglas, D. Goode, M. C. Houck, & D. Maddox (Eds.), *The Routledge Handbook of Urban Ecology*. London, UNITED KINGDOM: Taylor & Francis Group.

Mauri, G., Sartore-Bianchi, A., Russo, A.-G., Marsoni, S., Bardelli, A., & Siena, S. (2019). Early-onset colorectal cancer in young individuals. *Molecular Oncology*, 13(2), 109-131. doi:<https://doi.org/10.1002/1878-0261.12417>

May, J. M. (1950). Medical Geography: Its Methods and Objectives. *Geographical Review*, 40(1), 9-41.

McCormack, G. R., Cabaj, J., Orpana, H., Lukic, R., Blackstaffe, A., Goopy, S., . . . Chapman, J. (2019). Evidence synthesis A scoping review on the relations between urban form and health: A focus on Canadian quantitative evidence. *Health promotion and chronic disease prevention in Canada: research, policy and practice*, 39(5), 187.

McGrath, P., & Rawson, N. (2013). The experience of relocation for specialist treatment for Indigenous women diagnosed with vulvar cancer in East Arnhem Land. *Journal of Psychosocial Oncology*, 31(5), 540-555.

McLeroy, K. R., Bibeau, D., Steckler, A., & Glanz, K. (1988). An ecological perspective on health promotion programs. *Health Educ Q*, 15(4), 351-377.
doi:10.1177/109019818801500401

- Mears, M., & Brindley, P. (2019). Measuring urban greenspace distribution equity: the importance of appropriate methodological approaches. *ISPRS International Journal of Geo-Information*, 8(6), 286.
- Mears, M., Brindley, P., Jorgensen, A., Ersoy, E., & Maheswaran, R. (2019). Greenspace spatial characteristics and human health in an urban environment: An epidemiological study using landscape metrics in Sheffield, UK. *Ecological Indicators*, 106, 105464.
- Mears, M., Brindley, P., Jorgensen, A., & Maheswaran, R. (2020). Population-level linkages between urban greenspace and health inequality: The case for using multiple indicators of neighbourhood greenspace. *Health & Place*, 102284. doi:<https://doi.org/10.1016/j.healthplace.2020.102284>
- Meen, G. (2009). Modelling Local Spatial Poverty Traps in England. *Housing Studies*, 24(1), 127-147. doi:10.1080/02673030802547413
- Meliker, J. R., & Jacquez, G. M. (2007). Space-time clustering of case-control data with residential histories: insights into empirical induction periods, age-specific susceptibility, and calendar year-specific effects. *Stoch Environ Res Risk Assess*, 21(5), 625-634. doi:10.1007/s00477-007-0140-3
- Meliker, J. R., & Sloan, C. D. (2011). Spatio-temporal epidemiology: principles and opportunities. *Spat Spatiotemporal Epidemiol*, 2(1), 1-9. doi:10.1016/j.sste.2010.10.001
- Meliker, J. R., Slotnick, M. J., AvRuskin, G. A., Schottenfeld, D., Jacquez, G. M., Wilson, M. L., . . . Nriagu, J. O. (2010). Lifetime exposure to arsenic in drinking water and bladder cancer: a population-based case-control study in Michigan, USA. *Cancer Causes & Control*, 21(5), 745-757. doi:10.1007/s10552-010-9503-z
- Mills, M. (2010). *Introducing survival and event history analysis*: SAGE Publications. London, Thousand Oaks, New Delhi, Signapore.
- Mitchell, R., & Popham, F. (2008). Effect of exposure to natural environment on health inequalities: an observational population study. *The Lancet*, 372(9650), 1655-1660.
- Moriña, D., & Navarro, A. (2014). The R package survsim for the simulation of simple and complex survival data. *Journal of Statistical Software*, 59(2), 1-20.
- Muralidhar, V., Nguyen, P. L., & Tucker-Seeley, R. D. (2016). Recent relocation and decreased survival following a cancer diagnosis. *Preventive Medicine*, 89, 245-250. doi:<https://doi.org/10.1016/j.ypmed.2016.06.007>

- Murphy, C. C., Wallace, K., Sandler, R. S., & Baron, J. A. (2019). Racial Disparities in Incidence of Young-Onset Colorectal Cancer and Patient Survival. *Gastroenterology*, *156*(4), 958-965. doi:<https://doi.org/10.1053/j.gastro.2018.11.060>
- Niu, X., Pawlish, K. S., & Roche, L. M. (2010). Cancer survival disparities by race/ethnicity and socioeconomic status in New Jersey. *J Health Care Poor Underserved*, *21*(1), 144-160. doi:10.1353/hpu.0.0263
- Niu, X., Roche, L. M., Pawlish, K. S., & Henry, K. A. (2013). Cancer survival disparities by health insurance status. *Cancer Med*, *2*(3), 403-411. doi:10.1002/cam4.84
- NJSCR. (2018). NJ State Cancer Registry. Retrieved from <https://www.state.nj.us/health/ces/reporting-entities/njscr/>
- Nordsborg, R. B., Meliker, J. R., Ersboll, A. K., Jacquez, G. M., Poulsen, A. H., & Raaschou-Nielsen, O. (2014). Space-time clusters of breast cancer using residential histories: a Danish case-control study. *BMC Cancer*, *14*, 255. doi:10.1186/1471-2407-14-255
- Nordsborg, R. B., Sloan, C. D., Shahid, H., Jacquez, G. M., De Roos, A. J., Cerhan, J. R., . . . Meliker, J. R. (2015). Investigation of spatio-temporal cancer clusters using residential histories in a case-control study of non-Hodgkin lymphoma in the United States. *Environ Health*, *14*, 48. doi:10.1186/s12940-015-0034-7
- Northridge, M. E., Sclar, E. D., & Biswas, P. (2003). Sorting out the connections between the built environment and health: a conceptual framework for navigating pathways and planning healthy cities. *Journal of Urban Health*, *80*(4), 556-568.
- Nowak, D. J., & Greenfield, E. J. (2010). Evaluating the National Land Cover Database tree canopy and impervious cover estimates across the conterminous United States: a comparison with photo-interpreted estimates. *Environmental Management*, *46*(3), 378-390.
- O'Callaghan-Gordo, C., Pollan, M., Molina de la Torre, A. J., Dierssen-Sotos, T., Ardanaz, E., Fernández-Tardón, G., . . . Nieuwenhuijsen, M. (2018). *Green Spaces and Colon Cancer Risk: A Case-Control in Spain*. Paper presented at the ISEE Conference Abstracts.
- Oliver, M. N., Matthews, K. A., Siadat, M., Hauck, F. R., & Pickle, L. W. (2005). Geographic bias related to geocoding in epidemiologic studies. *International Journal of Health Geographics*, *4*(1), 29.
- Oudin, A., Forsberg, B., Strömgren, M., Beelen, R., & Modig, L. (2012). Impact of residential mobility on exposure assessment in longitudinal air pollution studies: a

sensitivity analysis within the ESCAPE project. *The Scientific World Journal*, 2012.

Padgett, J. F. (1990). Mobility as control: Congressmen through committees. In R. L. Breiger (Ed.), *Social mobility and social structure* (pp. 27-58). Cambridge, New York, Port Chester, Melbourne, Sydney: Cambridge University Press.

Palmer, J. F. (2004). Using spatial metrics to predict scenic perception in a changing landscape: Dennis, Massachusetts. *Landscape and Urban Planning*, 69(2), 201-218. doi:<https://doi.org/10.1016/j.landurbplan.2003.08.010>

Pawitan, G., & Steel, D. (2009). Exploring the MAUP from a spatial perspective.

Percy, C., Holten, V. v., Muir, C. S., & Organization, W. H. (1990). *International classification of diseases for oncology*. Retrieved from Geneva:

Plascak, J. J., Rundle, A. G., Babel, R. A., Llanos, A. A., LaBelle, C. M., Stroup, A. M., & Mooney, S. J. (2020). Drop-and-spin virtual neighborhood auditing: assessing built environment for linkage to health studies. *American journal of preventive medicine*, 58(1), 152-160.

Power, A., & Wilson, W. J. (2000). Social exclusion and the future of cities. *LSE STICERD Research Paper No. CASE035*.

Pulte, D., Jansen, L., & Brenner, H. (2018). Disparities in Colon Cancer Survival by Insurance Type: A Population-Based Analysis. *Dis Colon Rectum*, 61(5), 538-546. doi:10.1097/dcr.0000000000001068

Purcell, P. J. (2020). Geographic Mobility and Annual Earnings in the United States. *Soc. Sec. Bull.*, 80, 1.

Pyle, G. F. (1976). Human Health Problems - Spatial Perspectives - Foundations to Medical Geography - Introduction. *Economic Geography*, 52(2), 95-102. doi:10.2307/143356

R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>

Rapoport, A. (2011). *The Mutual Interaction of People and Their Built Environment*: De Gruyter Mouton.

Rappaport, J. (2007). Moving to nice weather. *Regional Science and Urban Economics*, 37(3), 375-398.

- Rauh, V. A., Landrigan, P. J., & Claudio, L. (2008). Housing and health: intersection of poverty and environmental exposures. *Annals of the New York Academy of Sciences*, 1136(1), 276-288.
- Renalds, A., Smith, T. H., & Hale, P. J. (2010). A systematic review of built environment and health. *Family & community health*, 33(1), 68-78.
- Richardson, A. S., Ghosh-Dastidar, M., Collins, R. L., Hunter, G. P., Troxel, W. M., Colabianchi, N., . . . Dubowitz, T. (2020). Improved Street Walkability, Incivilities, and Esthetics Are Associated with Greater Park Use in Two Low-Income Neighborhoods. *Journal of Urban Health*, 97(2), 204-212. doi:10.1007/s11524-019-00416-7
- Richardson, A. S., Meyer, K. A., Howard, A. G., Boone-Heinonen, J., Popkin, B. M., Evenson, K. R., . . . Gordon-Larsen, P. (2014). Neighborhood socioeconomic status and food environment: a 20-year longitudinal latent class analysis among CARDIA participants. *Health & Place*, 30, 145-153.
- Richardson, E., Pearce, J., Mitchell, R., Day, P., & Kingham, S. (2010). The association between green space and cause-specific mortality in urban New Zealand: an ecological analysis of green space utility. *BMC Public Health*, 10(1), 240.
- Richardson, E. A., & Mitchell, R. (2010). Gender differences in relationships between urban green space and health in the United Kingdom. *Social Science & Medicine*, 71(3), 568-575.
- Richardson, E. A., Mitchell, R., Hartig, T., De Vries, S., Astell-Burt, T., & Frumkin, H. (2012). Green cities and health: a question of scale? *J Epidemiol Community Health*, 66(2), 160-165.
- Richter, M., Günther, S., & Herke, M. (2018). Lebenslaufforschung in der Epidemiologie und den Sozialwissenschaften. Gemeinsam einsam? [Life Course Epidemiology and the Social Sciences: Living Apart Together?]. *Gesundheitswesen*, 43(06), 522-531. doi:10.1055/s-0042-108642
- Riitters, K. H., O'Neill, R., Hunsaker, C., Wickham, J. D., Yankee, D., Timmins, S., . . . Jackson, B. (1995). A factor analysis of landscape pattern and structure metrics. *Landscape Ecology*, 10(1), 23-39.
- Robinson, J. R. M. (2019). *Effects of individual and neighborhood socioeconomic status on outcomes following a colorectal cancer diagnosis*.
- Roche, L. M., Niu, X., Stroup, A. M., & Henry, K. A. (2017). Disparities in Female Breast Cancer Stage at Diagnosis in New Jersey: A Spatial-Temporal Analysis. *Journal of public health management and practice : JPHMP*, 23(5), 477-486. doi:10.1097/PHH.0000000000000524

- Rosenberg, M. (2015). Health geography II. *Progress in Human Geography*, 40(4), 546-554. doi:10.1177/0309132515581094
- Roswall, N., Bidstrup, P. E., Raaschou-Nielsen, O., Jensen, S. S., Olsen, A., & Sørensen, M. (2016). Residential road traffic noise exposure and survival after breast cancer—A cohort study. *Environmental Research*, 151, 814-820.
- Ruder, A. M., & Bertke, S. J. (2017). Cancer incidence among boat-building workers exposed to styrene. *American Journal of Industrial Medicine*, 60(7), 651-657. doi:10.1002/ajim.22735
- Rushton, G. (2003). Public health, GIS, and spatial analytic tools. *Annual Review of Public Health*, 24(1), 43-56.
- Rutter, C. M., Knudsen, A. B., Lin, J. S., & Bouskill, K. E. (2021). Black and White Differences in Colorectal Cancer Screening and Screening Outcomes: A Narrative Review. *Cancer Epidemiology Biomarkers & Prevention*, 30(1), 3-12. doi:10.1158/1055-9965.epi-19-1537
- Rzotkiewicz, A., Pearson, A. L., Dougherty, B. V., Shortridge, A., & Wilson, N. (2018). Systematic review of the use of Google Street View in health research: Major themes, strengths, weaknesses and possibilities for future research. *Health & Place*, 52, 240-246. doi:<https://doi.org/10.1016/j.healthplace.2018.07.001>
- Sahar, L., Foster, S. L., Sherman, R. L., Henry, K. A., Goldberg, D. W., Stinchcomb, D. G., & Bauer, J. E. (2019). GIScience and cancer: State of the art and trends for cancer surveillance and epidemiology. *Cancer*, 125(15), 2544-2560. doi:10.1002/cncr.32052
- Sampson, R. J. (2003). *Neighborhood-level context and health: lessons from sociology*: Nueva York, Oxford University Press.
- Schmitz, K. H., Stout, N. L., Maitin-Shepard, M., Campbell, A., Schwartz, A. L., Grimmett, C., . . . Sokolof, J. M. (2021). Moving through cancer: Setting the agenda to make exercise standard in oncology practice. *Cancer*, 127(3), 476-484. doi:<https://doi.org/10.1002/cncr.33245>
- Schootman, M., Gomez, S. L., Henry, K. A., Paskett, E. D., Ellison, G. L., Oh, A., . . . Berrigan, D. A. (2017). Geospatial Approaches to Cancer Control and Population Sciences. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 26(4), 472-475. doi:10.1158/1055-9965.EPI-17-0104
- Schootman, M., Jeffe, D. B., Lian, M., Gillanders, W. E., & Aft, R. (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. doi:10.1093/aje/kwn369
- Schullehner, J., Hansen, B., Thygesen, M., Pedersen, C. B., & Sigsgaard, T. (2018). Nitrate in drinking water and colorectal cancer risk: A nationwide population-based cohort study. *International journal of cancer*, 143(1), 73-79.
- Sell, R. R. (1983). Analyzing migration decisions: the first step-whose decisions? *Demography*, 20(3), 299-311.
- Shariff-Marco, S., Gomez, S. L., Sangaramoorthy, M., Yang, J., Koo, J., Hertz, A., . . . Keegan, T. H. M. (2015). Impact of neighborhoods and body size on survival after breast cancer diagnosis. *Health & Place*, 36, 162-172. doi:<https://doi.org/10.1016/j.healthplace.2015.10.003>
- Shariff-Marco, S., Yang, J., John, E. M., Sangaramoorthy, M., Hertz, A., Koo, J., . . . Gomez, S. L. (2014). Impact of Neighborhood and Individual Socioeconomic Status on Survival after Breast Cancer Varies by Race/Ethnicity: The Neighborhood and Breast Cancer Study. *Cancer Epidemiology Biomarkers & Prevention*, 23(5), 793-811. doi:10.1158/1055-9965.epi-13-0924
- Shaw, M. (2004). Housing and public health. *Annu. Rev. Public Health*, 25, 397-418.
- Shvetsov, Y. B., Shariff-Marco, S., Yang, J., Conroy, S. M., Canchola, A. J., Albright, C. L., . . . Cheng, I. (2020). Association of change in the neighborhood obesogenic environment with colorectal cancer risk: The Multiethnic Cohort Study. *SSM - Population Health*, 10, 100532. doi:<https://doi.org/10.1016/j.ssmph.2019.100532>
- Singh, G. K. (2003). Area Deprivation and Widening Inequalities in US Mortality, 1969–1998. *American Journal of Public Health*, 93(7), 1137-1143.
- Singh, G. K., & Jemal, A. (2017). Socioeconomic and Racial/Ethnic Disparities in Cancer Mortality, Incidence, and Survival in the United States, 1950–2014: Over Six Decades of Changing Patterns and Widening Inequalities. *Journal of Environmental and Public Health*, 2017, 19. doi:10.1155/2017/2819372
- Singh, G. K., Williams, S. D., Siahpush, M., & Mulhollen, A. (2011). Socioeconomic, Rural-Urban, and Racial Inequalities in US Cancer Mortality: Part I-All Cancers and Lung Cancer and Part II-Colorectal, Prostate, Breast, and Cervical Cancers. *J Cancer Epidemiol*, 2011, 107497-107497. doi:10.1155/2011/107497
- Skobba, K., & Goetz, E. G. (2013). Mobility decisions of very low-income households. *Cityscape*, 155-171.

- Sloan, C. D., Nordsborg, R. B., Jacquez, G. M., Raaschou-Nielsen, O., & Meliker, J. R. (2015). Space-time analysis of testicular cancer clusters using residential histories: a case-control study in Denmark. *PLoS One*, *10*(3), e0120285. doi:10.1371/journal.pone.0120285
- Soderstrom, M. (2008). *The walkable city: from Haussmann's boulevards to Jane Jacobs' streets and beyond*. Montreal: Vehicule Press.
- Staats, H., Kieviet, A., & Hartig, T. (2003). Where to recover from attentional fatigue: An expectancy-value analysis of environmental preference. *Journal of Environmental Psychology*, *23*(2), 147-157. doi:[https://doi.org/10.1016/S0272-4944\(02\)00112-3](https://doi.org/10.1016/S0272-4944(02)00112-3)
- Steinbrecher, A., Fish, K., Clarke, C. A., West, D. W., Gomez, S. L., & Cheng, I. (2012). Examining the association between socioeconomic status and invasive colorectal cancer incidence and mortality in California. *Cancer Epidemiology and Prevention Biomarkers*, *21*(10), 1814-1822.
- Stinchcomb, D., & Roeser, A. (2016). *NCI/SEER Residential History Project Technical Report*. Rockville, MD, Westat, Inc.
- Stokols, D., & Shumaker, S. A. (1982). The Psychological Context of Residential Mobility and Well-Being. *Journal of Social Issues*, *38*(3), 149-171.
- Stokols, D., Shumaker, S. A., & Martinez, J. (1983). Residential mobility and personal well-being. *Journal of Environmental Psychology*, *3*(1), 5-19.
- Strauss, A. L. (2009). *The contexts of social mobility: Ideology and theory*. New Brunswick, London: Transaction Publishers.
- Sugiyama, T., Carver, A., Koohsari, M. J., & Veitch, J. (2018). Advantages of public green spaces in enhancing population health. *Landscape and Urban Planning*, *178*, 12-17.
- Sui, D. Z. (2007). Geographic Information Systems and Medical Geography: Toward a New Synergy. *Geography Compass*, *1*(3), 556-582. doi:10.1111/j.1749-8198.2007.00027.x
- Texas A&M, N. a. N. (2016). *NAACCR Geocoder Data Dictionary*. Retrieved from <http://www.naacr.org/wp-content/uploads/2016/11/NAACCR-Geocoder-Data-Dictionary.pdf>
- Therneau, T. M., & Lumley, T. (2015). Package 'survival'. *R Top Doc*, *128*, 112.
- Thernstrom, S. (1968). Notes on the Historical Study of Social Mobility. *Comparative Studies in Society and History*, *10*(2), 162-172.

- Tong, S. (2000). Migration bias in ecologic studies. *European journal of epidemiology*, 16(4), 365-369.
- Townsend, P. (2002). Poverty, social exclusion and social polarisation: the need to construct an international welfare state. *World poverty: New policies to defeat an old enemy*, 3-24.
- Trivers, K. F., Shaw, K. M., Sabatino, S. A., Shapiro, J. A., & Coates, R. J. (2008). Trends in colorectal cancer screening disparities in people aged 50–64 years, 2000–2005. *American journal of preventive medicine*, 35(3), 185-193.
- Tsai, S. P., Chen, V. W., Fox, E. E., Wendt, J. K., Wu, X. C., Foster, D. E., & Fraser, A. E. (2004). Cancer incidence among refinery and petrochemical employees in Louisiana, 1983–1999. *Annals of Epidemiology*, 14(9), 722-730.
- Tsai, W.-L., McHale, M. R., Jennings, V., Marquet, O., Hipp, J. A., Leung, Y.-F., & Floyd, M. F. (2018). Relationships between characteristics of urban green land cover and mental health in US metropolitan areas. *International Journal of Environmental Research and Public Health*, 15(2), 340.
- Turner, M. G., Gardner, R. H., O'Neill, R. V., & O'Neill, R. V. (2001). *Landscape ecology in theory and practice* (Vol. 401): Springer.
- Umlauf, N., Adler, D., Kneib, T., Lang, S., & Zeileis, A. (2012). *Structured additive regression models: An R interface to BayesX*. Retrieved from
- Umlauf, N., Kneib, T., Heinzl, F., Lang, S., & Zeileis, A. (2013). R2BayesX: Estimate structured additive regression models with BayesX. *R package version 0.1-2*.
- Umlauf, N., Kneib, T., Lang, S., Zeileis, A., & Umlauf, M. N. (2012). Package 'R2BayesX'.
- Urayama, K. Y., Von Behren, J., Reynolds, P., Hertz, A., Does, M., & Buffler, P. A. (2009). Factors associated with residential mobility in children with leukemia: implications for assigning exposures. *Annals of Epidemiology*, 19(11), 834-840.
- Uuemaa, E., Antrop, M., Roosaare, J., Marja, R., & Mander, Ü. (2009). Landscape metrics and indices: an overview of their use in landscape research. *Living reviews in landscape research*, 3(1), 1-28.
- Uuemaa, E., Mander, Ü., & Marja, R. (2013). Trends in the use of landscape spatial metrics as landscape indicators: a review. *Ecological Indicators*, 28, 100-106.

- van Beurden, A. U., & de Lepper, M. J. (1995). The Integration of Information in Geographical Information Systems *The Added Value of Geographical Information Systems in Public and Environmental Health* (pp. 71-86): Springer.
- Van Blarigan, E. L., & Meyerhardt, J. A. (2015). Role of physical activity and diet after colorectal cancer diagnosis. *J Clin Oncol*, *33*(16), 1825-1834.
doi:10.1200/jco.2014.59.7799
- van den Berg, A. E., Koole, S. L., & van der Wulp, N. Y. (2003). Environmental preference and restoration: (How) are they related? *Journal of Environmental Psychology*, *23*(2), 135-146. doi:[https://doi.org/10.1016/S0272-4944\(02\)00111-1](https://doi.org/10.1016/S0272-4944(02)00111-1)
- Van Dillen, S. M., de Vries, S., Groenewegen, P. P., & Spreeuwenberg, P. (2012). Greenspace in urban neighbourhoods and residents' health: adding quality to quantity. *J Epidemiol Community Health*, *66*(6), e8-e8.
- van Ham, M. (2012). Housing behaviour. In D. F. Clapham, W. A. Clark, & K. Gibb (Eds.), *The Sage handbook of housing studies* (pp. 47-65). London: Sage.
- VanDerWal, J., Falconi, L., Januchowski, S., Shoo, L., Storlie, C., & VanDerWal, M. J. (2014). Package 'SDMTools'. *R package*.
- Wan, N., Zhan, F. B., Lu, Y., & Tiefenbacher, J. P. (2012). Access to healthcare and disparities in colorectal cancer survival in Texas. *Health & Place*, *18*(2), 321-329.
doi:<https://doi.org/10.1016/j.healthplace.2011.10.007>
- Wan, N., Zhan, F. B., Zou, B., & Chow, E. (2012). A relative spatial access assessment approach for analyzing potential spatial access to colorectal cancer services in Texas. *Applied Geography*, *32*(2), 291-299.
doi:<https://doi.org/10.1016/j.apgeog.2011.05.001>
- Wan, N., Zou, B., & Sternberg, T. (2012). A three-step floating catchment area method for analyzing spatial access to health services. *International Journal of Geographical Information Science*, *26*(6), 1073-1089.
doi:10.1080/13658816.2011.624987
- Wang, F. (2012). Measurement, Optimization, and Impact of Health Care Accessibility: A Methodological Review. *Annals of the Association of American Geographers*, *102*(5), 1104-1112. doi:10.1080/00045608.2012.657146
- Wang, F., & Luo, W. (2005). Assessing spatial and nonspatial factors for healthcare access: towards an integrated approach to defining health professional shortage areas. *Health & Place*, *11*(2), 131-146.
- Wang, L., Wilson, S. E., Stewart, D. B., & Hollenbeak, C. S. (2011). Marital status and colon cancer outcomes in US Surveillance, Epidemiology and End Results

registries: does marriage affect cancer survival by gender and stage? *Cancer epidemiology*, 35(5), 417-422.

- Warnecke, R. B., Oh, A., Breen, N., Gehlert, S., Paskett, E., Tucker, K. L., . . . Hiatt, R. A. (2008). Approaching Health Disparities From a Population Perspective: The National Institutes of Health Centers for Population Health and Health Disparities. *American Journal of Public Health*, 98(9), 1608-1615. doi:10.2105/AJPH.2006.102525
- Warner, E. T., & Gomez, S. L. (2010). Impact of neighborhood racial composition and metropolitan residential segregation on disparities in breast cancer stage at diagnosis and survival between black and white women in California. *J Community Health*, 35(4), 398-408. doi:10.1007/s10900-010-9265-2
- Wen, M., Zhang, X., Harris, C. D., Holt, J. B., & Croft, J. B. (2013). Spatial Disparities in the Distribution of Parks and Green Spaces in the USA. *Annals of Behavioral Medicine*, 45(suppl_1), S18-S27. doi:10.1007/s12160-012-9426-x
- Wendelboe-Nelson, C., Kelly, S., Kennedy, M., & Cherrie, J. W. (2019). A scoping review mapping research on green space and associated mental health benefits. *International Journal of Environmental Research and Public Health*, 16(12), 2081.
- Wheeler, D. C., & Wang, A. (2015). Assessment of Residential History Generation Using a Public-Record Database. *Int J Environ Res Public Health*, 12(9), 11670-11682. doi:10.3390/ijerph120911670
- Wheeler, D. C., Ward, M. H., & Waller, L. A. (2012). Spatial-Temporal Analysis of Cancer Risk in Epidemiologic Studies with Residential Histories. *Annals of the Association of American Geographers*, 102(5), 1049-1057. doi:10.1080/00045608.2012.671131
- White, A., Joseph, D., Rim, S. H., Johnson, C. J., Coleman, M. P., & Allemani, C. (2017). Colon cancer survival in the United States by race and stage (2001-2009): Findings from the CONCORD-2 study. *Cancer*, 123 Suppl 24(Suppl 24), 5014-5036. doi:10.1002/cncr.31076
- White, A., Vernon, S. W., Franzini, L., & Du, X. L. (2010). Racial disparities in colorectal cancer survival: to what extent are racial disparities explained by differences in treatment, tumor characteristics, or hospital characteristics? *Cancer*, 116(19), 4622-4631. doi:10.1002/cncr.25395
- Wiese, D., Rodriguez Escobar, J., Hsu, Y., Kulathinal, R. J., & Hayes-Conroy, A. (2018). The fluidity of biosocial identity and the effects of place, space, and time. *Soc Sci Med*, 198, 46-52. doi:10.1016/j.socscimed.2017.12.023

- Wiese, D., Stroup, A. M., Crosbie, A., Lynch, S. M., & Henry, K. A. (2019). The Impact of Neighborhood Economic and Racial Inequalities on the Spatial Variation of Breast Cancer Survival in New Jersey. *Cancer Epidemiology Biomarkers & Prevention*. doi:10.1158/1055-9965.epi-19-0416
- Wiese, D., Stroup, A. M., Maiti, A., Harris, G., Lynch, S. M., Vucetic, S., & Henry, K. A. (2020a). Residential mobility and geospatial disparities in colon cancer survival. *Cancer Epidemiology, Biomarkers and Prevention*, 29(11), 2119-2125. doi:10.1158/1055-9965.epi-20-0772
- Wiese, D., Stroup, A. M., Maiti, A., Harris, G., Lynch, S. M., Vucetic, S., & Henry, K. A. (2020b). Socioeconomic Disparities in Colon Cancer Survival: Revisiting Neighborhood Poverty using Residential Histories. *Epidemiology*, 31(5), 728-735. doi:10.1097/ede.0000000000001216
- Wu, J., & Jackson, L. (2017). Inverse relationship between urban green space and childhood autism in California elementary school districts. *Environment international*, 107, 140-146.
- Yang, C.-C., Cheng, L.-C., Lin, Y.-W., Wang, S.-C., Ke, T.-M., Huang, C.-I., . . . Tai, M.-H. (2019). The impact of marital status on survival in patients with surgically treated colon cancer. *Medicine*, 98(11), e14856. doi:10.1097/md.00000000000014856
- Zahnd, W. E., & McLafferty, S. L. (2017). Contextual effects and cancer outcomes in the United States: a systematic review of characteristics in multilevel analyses. *Ann Epidemiol*, 27(11), 739-748 e733. doi:10.1016/j.annepidem.2017.10.002
- Zhang, D., Matthews, C. E., Powell-Wiley, T. M., & Xiao, Q. (2019). Ten-year change in neighborhood socioeconomic status and colorectal cancer. *Cancer*, 125(4), 610-617. doi:10.1002/cncr.31832
- Zhang, Z., Reinikainen, J., Adeleke, K. A., Pieterse, M. E., & Groothuis-Oudshoorn, C. G. M. (2018). Time-varying covariates and coefficients in Cox regression models. *Annals of translational medicine*, 6(7), 121-121. doi:10.21037/atm.2018.02.12
- Zube, E. H., Sell, J. L., & Taylor, J. G. (1982). Landscape perception: Research, application and theory. *Landscape Planning*, 9(1), 1-33. doi:[https://doi.org/10.1016/0304-3924\(82\)90009-0](https://doi.org/10.1016/0304-3924(82)90009-0)
- Zuniga Teran, A. A. (2015). From Neighborhoods To Wellbeing And Conservation: Enhancing The Use Of Greenspace Through Walkability: The University of Arizona.

Zwiener, I., Blettner, M., & Hommel, G. (2011). Survival analysis: part 15 of a series on evaluation of scientific publications. *Deutsches Arzteblatt international*, 108(10), 163-169. doi:10.3238/arztebl.2010.0163