A Dissertation

entitled

The Influence of Social Environment, Physical Environment and Health Behaviors on Lung Cancer Mortality in Kentucky

by

Jayani Bothalage Done

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the

Doctor of Philosophy Degree in Spatially Integrated Social Sciences

Dr. Kevin Czajkowski, Committee Chair

Dr. Daniel J. Hammel, Committee Member

Dr. April Ames, Committee Member

Dr. Yanqing Xu, Committee Member

Dr. Yue Zhang, Committee Member

Dr. Amy Thompson, Dean College of Graduate Studies

The University of Toledo

May 2022

An Abstract of
The Influence of Social Environment, Physical Environment and Health Behaviors on Lung Cancer Mortality in Kentucky

by

Jayani Bothalage Done

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the Doctor of Philosophy Degree in Degree in Spatially Integrated Social Science

The University of Toledo

May 2022

Lung cancer has the highest mortality rate of all cancers worldwide. Generally, people from lower socioeconomic backgrounds have the highest incidence of lung cancer and mortality rates. Socioeconomic factors, physical environment, and health behaviors have been identified as crucial determinants of the incidence and mortality of cancer, survival rates, cancer stage at diagnosis, and treatment choices in the United States. Therefore, this study provides a comprehensive overview of lung cancer mortality trends, risk factor relationships, their influence, and disparity focusing on the state of Kentucky.

In this dissertation, Joinpoint regression model used to analyze recent changes in lung cancer incidence and mortality trends in Kentucky during 2000 - 2016. Incidence and mortality records were gathered from the Surveillance, Epidemiology, and End Results database (SEER). The result of Joinpoint analysis suggests overall, Kentucky lung cancer incidence trends show that progress is being made to reduce the lung cancer burden among residents of Kentucky.

The second phase of this paper investigates the relationship between socioeconomic variables and lung cancer mortality rate at the county scale using the Ordinary Least Squares Method (OLS) and Geographically Weighted Regression (GWR) method. The regression model results for all counties in Kentucky indicate the significant variables for lung cancer rates are adult smoking rate, high school graduation rate, median household income, and the number of coal mine employment. In non-Appalachian counties, results suggest that lung cancer rates positively correlate with adult smoking rates. Appalachian counties in Kentucky suggest lung cancer mortality rates increased with low graduation rates and low income.

Then the Geographical Detector technique was used to investigate the spatial distribution patterns of lung cancer mortality and suspected risk determinants. Adult smoking and median household income were the first two most important factors responsible for lung cancer mortality. The ecological detector finds that adult smoking rate, graduation rate, medium household income and uninsured rates substantially affect lung cancer mortality. The interactive detector demonstrated that the interaction of physically unhealthy days and the high school graduation rate nonlinearly enhanced lung cancer mortality. Also, the interactive effect between uninsured and high school graduation rates is nonlinear.

Finally, an overview of methods for summarizing socioeconomic and geographic disparities in health was compiled using the example of lung cancer. Results suggests, among males, mortality generally declined for all socioeconomic variables and geographic regions. But the magnitude of the decline was considerably more significant for men compared to women. On the other hand, the picture was more mixed among females.

Acknowledgments

I would like to express my sincere and warm gratitude to Dr. Kevin Czajkowski for providing me with the opportunity and resources to pursue my passion for research. His technical and editorial expertise and his constant guidance and encouragement were instrumental in completing this study. I am truly fortunate to have had him as my advisor. His mentorship has contributed immensely to my growth, both as an individual and a professional. I also would like to thank my committee members: Dr. Daniel J. Hammel, Dr. April Ames, Dr. Yanqing Xu, and Dr. Yue Zhang, for their valuable inputs and feedback that improved the content of this dissertation. A big thank you to my previous committee chair Dr. Yanquing Xu for her guidance and the current faculty in the Department of Geography and Planning. Furthermore, I would like to thank my friends at the University of Toledo: Dinesha Thejani, Lakshika Nishadhi, and Owusua Yamoah, for helping with the writing.

Finally, I am deeply grateful to my beloved family, who accepted my difficult decision to pursue my education abroad and supported me every step of the way, despite the distance. My mom, dad, loving husband, and beloved sisters were my first teachers in life: thank you for your patience, presence, and love even from far away. I am incredibly fortunate to have you in my life.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AAPC	Average annual percentage changes
ACI	Absolute Concentration Index
AIC	Akaike Information Criterion
APC	Annual percentage change
ARC	Appalachian Regional Commission
BGV	Between-group variance
CDC	Disease Control and Prevention
GIS	Geographic information system
GWR	Geographically Weighted Regression
HD*Calc	Health Disparities Calculator
IDisp	Index of Disparity
LMC	lung cancer mortality
MLD	Mean Log Deviation
NSCLC	Non-small cell lung cancer
OLS	Ordinary least squares Method
P.M	Particulate matter
PD	Power determinants
Ph. D	Doctor of Philosophy
RCI	Relative Concentration Index
RD	Rate Difference
RII	Relative Index of Inequality
RR	Rate Ratio

SEER	Surveillance, Epidemiology, and End Results
SES	Socioeconomic status
SII	Slope Index of Inequality
SSH	Spatial Stratified Heterogeneity
T	Theil Index

Chapter 1

Introduction

This chapter aims to rationalize lung cancer mortality and risk factors in context and provide a clear justification for the studies described in this dissertation. First, this chapter will describe the epidemiology of lung cancer and the current position of lung cancer mortality. Second, the chapter will highlight the body of literature exploring lung cancer mortality trends, risk factor relationships, and their disparities. Finally, this chapter will discuss the objectives and research questions driving this research investigation.

1.1 Research Motivation and Needs

Lung cancer has the highest mortality rate of all cancers worldwide (Hovanec et al., 2018). Lung cancer causes the most cancer death in both men and women in the U.S. In 1987, lung cancer became the leading cause of cancer death in woman over breast cancer (American Cancer Association, 2021). Approximately 154,050 Americans are expected to die from lung cancer in 2018, accounting for roughly 25 percent of cancer mortality(Siegel, Miller, & Jemal, 2018). Lung cancer mortality peaked at 159,292 in 2005 and has decreased by 6.5 percent to 148,945 in 2016 (American Cancer Association, 2021). The age-adjusted lung cancer mortality rate is higher for men (46.7 per 100,000 persons) than women (31.9 per 100,000 persons). The ratio is similar for blacks (40.0 per 100,000 persons) and whites (39.2 per 100,000 persons) overall. However, black men have a considerably higher age-adjusted lung cancer mortality rate than white men, while black and white women have similar rates (American Cancer Association, 2021).

Socioeconomic factors such as education, income, poverty, and unemployment are essential elements of our health and well-being. Socioeconomic factors can also lead to lung cancer disparities according to the geographical distribution of where people live, work, study, and play (County Health Rankings & Roadmaps, 2021b). These referred to social determinants of health, including differences in physical environments, individual behaviors, social factors, access to health care services, employment status, economic state, and literacy levels. Socioeconomic status (SES) correlates to lung cancer in several research studies, with people from lower socioeconomic backgrounds having the highest incidence rates (Ekberg-Aronsson, Nilsson, Nilsson, Pehrsson, & Löfdahl, 2006). SES reflects one's situation in societal hierarchies and is generally measured by the interdependent dimensions of education, occupation, and income. SES is linked with the disease through multiple interacting pathways in material and social resources, physical and psycho-social stressors, and health-related behavior. SES is strongly associated with smoking behavior (Schaap, van Agt, & Kunst, 2008), the most critical risk factor in the etiology of lung cancer.

Cigarette smoking is considered the number one risk factor for lung cancer. In the United States, cigarette smoking is associated with 80% to 90% of lung cancer fatalities. Using different types of tobacco products such as cigars or pipes also raises the risk of lung cancer. Tobacco smoke is a combination of a toxic mixture of more than 7,000 chemicals. Many are poisons and harmful to health. At least seventy toxins are known to cause cancer in people or animals (centers for Disease Control and Prevention, 2021).

Smoke from other people's cigarettes, pipes, or cigars (secondhand smoke) also affects lung cancer. While a person breathes in secondhand smoke, it is similar as if they are smoking. From 2013 to 2014, one out of every four nonsmokers in the United States, including 14 million children, were exposed to secondhand smoke (centers for Disease Control and Prevention, 2021).

Following smoking, radon is the second primary cause of lung cancer in the United States. Radon is a naturally arising gas from rocks, soil, and water. Radon is invisible, tasteless, and odorless. When radon enters a home or building through cracks or holes, it can become trapped and accumulate in the air in the interior. People who reside or work in these homes and structures breathe in high levels of radon. Over long periods of time, radon can cause lung cancer (centers for Disease Control and Prevention, 2021).

People, who work in places where asbestos is present (mills, mines, textile plants, sites where insulation is used, and shipyards), are more likely to develop lung cancer. In addition to other carcinogens (cancer-causing agents) found in some workplaces, other lung cancer risks, include uranium, arsenic, beryllium, cadmium, silica, vinyl chloride, nickel compounds, chromium compounds, coal products, mustard gas, chloromethyl ethers, and diesel exhaust (American Cancer Association, 2021). However, government and industry have taken actions to help protect workers from many of these exposures in recent years.

Air pollution, especially near severely trafficked roads, seems to raise the risk of lung cancer to some extent. This risk is considerably less than the risk produced by smoking, but some researchers estimate that about 5% of all mortality from lung cancer may be caused by outdoor air pollution (American Cancer Association, 2021).

Apart from the independent effects of different risk factors on lung cancer, complex combined effects might exist between various risk factors. The majority of researchers analyzed the independent influence of a single or a set of contextual factors on the lung cancer incidence or mortality rate (Klassen et al., 2019; Moore, Akinyemiju, & Wang, 2017). Research on the interactive effects of two or more risk factors is lacking. For example, physical environment features (e.g., radon. P.M 2.5), people's behavior and health conditions (e.g., smoking, uninsured rate, physically unhealthy days, etc.), and socio-economic factors (e.g., income, education, income, etc.). More notably, their mutual interactions are also significant underlying factors.

In addition to the high overall burden of cancer, lung cancer, and its risk factors have also differed systematically with social group status indicators such as race, sex, ethnicity, and socioeconomic status (Harper et al., 2008a). Such disparities are well-documented, and defeating cancer health disparities is a fundamental goal of the Healthy People 2010 Program (Davis, 2000) and one of the National Cancer Institute's vital strategic objectives (National Institutes of, 2006). US public health goals are to eliminate health disparities according to race, sexual orientation, education or income, disability, and geographic location (Davis, 2000). The measurement of improvement toward this goal has effects for prioritizing efforts aimed at reducing such disparities. As a result of the current policy emphasis on differences in cancer, it is essential to assess the level of progress toward disparity-related goals for two reasons. First, supervising disparities is a natural complement to monitoring overall progress in the fight against cancer and is crucial for identifying specific groups that may be experiencing a high burden of cancer-related illness. Second, examining disparities is important because it affords an opportunity to resolve observed trends with current etiologic justifications for the causes of social disparities in cancer (Krieger, 2005).

1.2 Objectives of the Study

Most of the US-based literature on public health and environmental justice includes extensive contributions by geographers, including numerous empirical studies of patterns of lung cancer mortality trends and socio-economic concern (Harper et al., 2008a; Hovanec et al., 2018).

While different social and economic risk factors, environmental factors, and health behaviors have been investigated, their impact on lung cancer constantly persists in public health and social concern. Although previous research studies have made significant steps towards understanding the influence and relationship of the imbalances in the geographic distribution of lung cancer mortality, they have been limited methodologically in four critical ways. This dissertation aims to focus on these limited four areas.

- The study's first phase aims to determine if there is a trend difference in analyzing age-adjusted cancer incidence and mortality rates trends. Also, this study evaluates trends in male and female incidence and mortality rates of lung cancer from 2002 until 2019 and their relationship to changes in diagnosis and treatment in recent decades. This method allows the user to interpret changes more accurately over time and, more importantly, identify the changes that occur between male and female.

- The second section seeks to examine how lung cancer mortality rates were distributed disproportionately concerning socioeconomic factors. Results from this section will demonstrate spatial relationships and explain the factors behind observed spatial patterns.

- The third phase of this study aims to calculate the mutual associations between a geographical phenomenon and relevant risk factors. This section reveals the spatial distribution patterns of lung cancer mortality and suspected determinants to help understand health risks factors. The underlying principle is to estimate the consistencies between the spatial distribution patterns of the studied geographical event (e.g., lung cancer mortality rate) and those of potential risk factors (e.g., education, income, smoking, etc.). Models propose four types of spatial variance analysis to assess combined effects that exist between different risk factors.

- Fourth, in recent years there has been a revival of interest in health disparity within public health. Health disparities have gained increasing attention from physicians and health policy experts and a renewed focus from federal health agencies. Lung cancer disproportionately impacts those from deprived groups. Those from more disadvantaged groups are not only more likely to be diagnosed with lung cancer, but they are also more prone to die from it when compared to their less deprived counterparts (Powell, 2019). This section aims to describe and empirically compare selected summary measures of health disparity in lung cancer mortality.

1.3 Research Questions

The detailed research questions investigated in this study are as follows:

1. What are the lung cancer incidence and mortality trends in Kentucky? And what are the disparities in male and female lung cancer trends?

2. Is there any statistical significance between lung cancer mortality and socio-economic factors across Kentucky? And what are the geographic patterns of lung cancer mortality in the Appalachian region versus the non-Appalachian region?

3. What is the spatial variation analysis of lung cancer in Kentucky? Are there any interactive effects on lung cancer risk factors? What are the highest lung cancer risk areas? What type of risk factors are mainly responsible for lung cancer? And what are their relative importance? Do lung cancer risk factors interact or lead to disease independently?

4. Is there lung cancer disparity across Kentucky?

This study utilizes the county-level data from the Surveillance, Epidemiology, and End Results (SEER) Program. SEER database collects cancer data from population-based cancer registries covering roughly 34.6 percent of the U.S. population. The SEER registries gather data on patient demographics, leading tumor site, tumor morphology, stage at diagnosis, and first plan of treatment, and they follow up with patients for vital status. The study also utilized data from numerous databases, such as the United States Census, County Health Ranking, U.S Environmental Protection Agency, Bureau of Labor Statistics, Kentucky Energy and Environment Cabinet - Department for Energy Development and Independence, and the database of Center for Disease Control.

Joinpoint regression, ordinary square regression method, Moran's I, and Geographically weighted regression method (GWR) were utilized to determine the trends of lung cancer and the relationship between risk factors and lung cancer mortality rates. The Geodetector model was used to identify the critical risk factors' combined effect on lung cancer mortality. Finally, a health disparity calculator (HD*Calc) was applied to measure lung cancer disparity using socio-economic factors and geographic region. The following flowchart demonstrates the sequence of steps of the four main sections studied in this dissertation. (Figure 1-1)

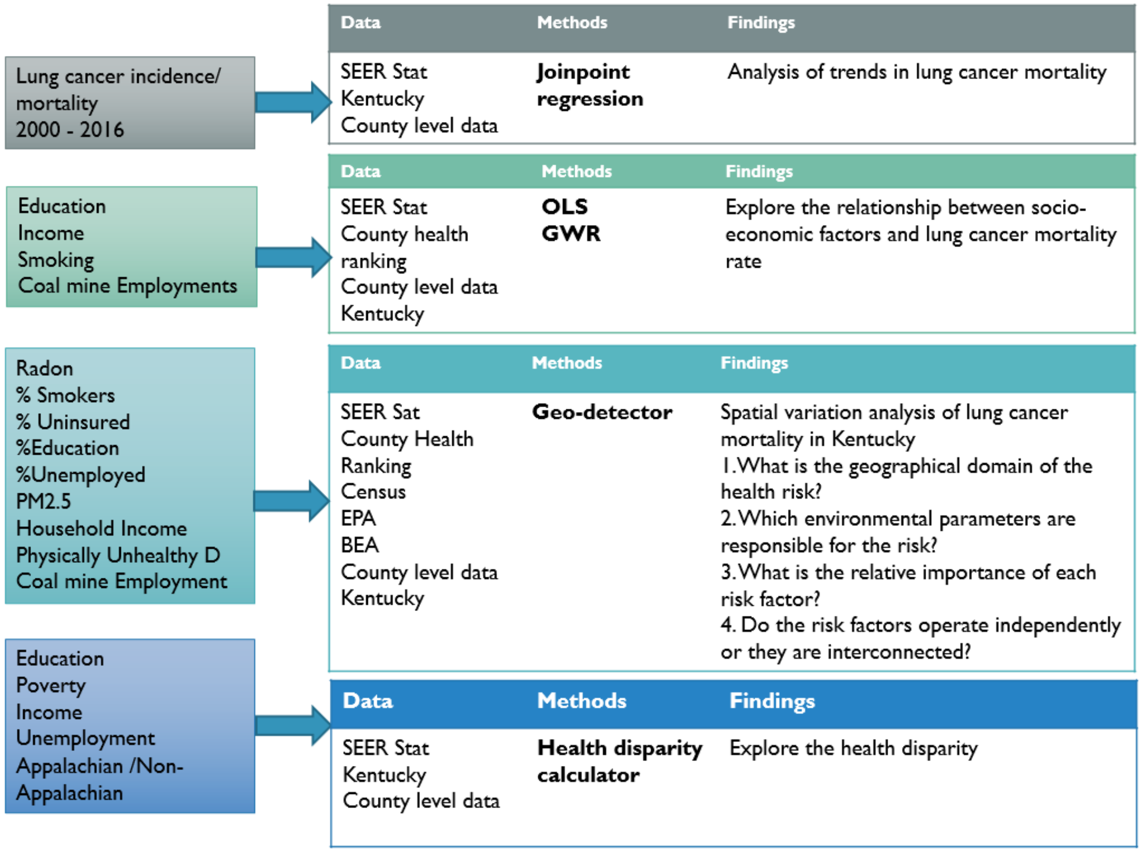


Figure 1-1: A sequence of steps used

1.4 Significance of the Study

Compared to genetic and medical literature, social environment, physical environment, and health behavior relationship, inequality, and disparity in lung cancer mortality have received limited attention. Furthermore, the inter-relationships of the multiple measures of socioeconomic status and their interaction with risk factors have limited consideration. Although previous research studies have made necessary steps towards understanding the impact and relationship of the imbalances in the geographic distribution of lung cancer mortality, they have been systematically limited in four critical ways. This dissertation aims to focus on these limited four areas.

First, previous analyses of lung cancer mortality and incidence were based on models of death rates or incidence within one time, if rates increase or decrease with time at a constant rate. Also, the interest of male and female lung cancer trends had inadequate attention in previous research work. To overcome this problem, the first phase of the study aims to analyze recent changes in male and female lung cancer incidence and mortality trends in Kentucky from 2002 through 2019 using Joinpoint regression models.

Second, lung cancer mortality and its relationship with risk factors have been studied by different researchers. But disparities in incidence and interaction of risk factors in diverse geographic areas has limited attention. Thus, lung cancer risk factors that produce measurable effects on lung cancer mortality have been identified regionally throughout the state. The primary focus was Appalachia due to its extremely high rates of mortality. However, the study also addressed mortality patterns across the entire state to understand why Appalachian counties have higher lung cancer mortality and to explore possible cause(s).

Third, apart from the independent effects of various risk factors on lung cancer, complex interactive outcomes might exist between different risk factors. Previous research analyzed the independent influence of a single or a set of contextual factors on lung cancer incidence and mortality rate (Klassen et al., 2019; Moore et al., 2017); however, the interactive impacts of two or more risk factors has not been studied. For example, physical environment features (e.g., radon. P.M 2.5), people's behavior and health conditions (e.g., sex, smoking, uninsured rate, physically unhealthy days, etc.), and socio-economic factors (e.g., income, education, etc.) have all been studied. More notably, their mutual interactions are also critical underlying factors. Therefore, the third section of this study focuses on interactive effects of lung cancer risk factors.

Forth, several previous research studies have focused on disparities between specific groups (e.g., Black/White, poor/rich) and applied measures such as rate ratios to calculate the difference. However, when considering disparities across multiple subgroups and how those may change over time and by gender is lacking in current research. Therefore, this study adopted a statistical perspective compatible with the Healthy People 2010 (Davis, 2000) framework, which seeks to exclude disparities across the entire range of subgroups defined by characteristics such as socioeconomic position and gender.

The intellectual merit of this dissertation lies in its potential to enhance our understanding of social, physical, and health behavior relationships, disparities, and interaction of different lung cancer risk factors on lung cancer mortality in Kentucky.

This dissertation will benefit society by pointing out gaps in understanding of measurable effects of the community, health, and environmental policy and science. This dissertation will provide helpful insight for advocacy and building policy on lung cancer mortality. Policies need to focus more broadly on upstream causes. Traditionally, these policies have been focused on downstream behaviors (e.g., public space smoking ban). Still, upstream approaches should base fundamental political decisions on distribution of income, education, and health facilities.

1.5 Study Area

In 2009, Kentucky had the highest cigarette smoking rate in the United States, at about 25.6% of the adult population (W. Jay Christian, Bin Huang, John Rinehart, & Claudia Hopenhayn, 2011). Age-adjusted lung cancer incidence and mortality rates in Kentucky are also among the highest ranking in the nation, at 97.7 and 74.6 per 100,000 people, respectively, in 2007 (centers for Disease Control and Prevention).

These statistics vary widely across the 120 counties in Kentucky; however, counties in the southeastern portion of the state generally have higher smoking and lung cancer incidence (Appalachian Regional Commission). The geographic distributions of the Appalachian and non-Appalachian counties are illustrated in Figure 1-2. Fifty-four counties belong to the Appalachian region, and 66 counties belong to the non-Appalachian area. Many of these counties belong to central Appalachia, a subregion of Appalachia well-known for its high poverty and low educational achievement (Twiss & Mueller, 2004).

The Appalachian region incorporates counties in 13 states from New York to Mississippi and has a higher ratio of lung cancer than the general U.S. population. Central Appalachia (West Virginia, eastern Kentucky, and adjacent parts of Tennessee and Virginia) has the highest lung cancer rates in the region and the nation (Lengerich et al., 2005). However, a recent multi-scale study suggests that high lung cancer mortality rates in coal-mining areas of Central Appalachia cannot be determined by tobacco use alone (Hendryx, O'Donnell, & Horn, 2008). Therefore, more research needs to be done on understand this phenomenon.

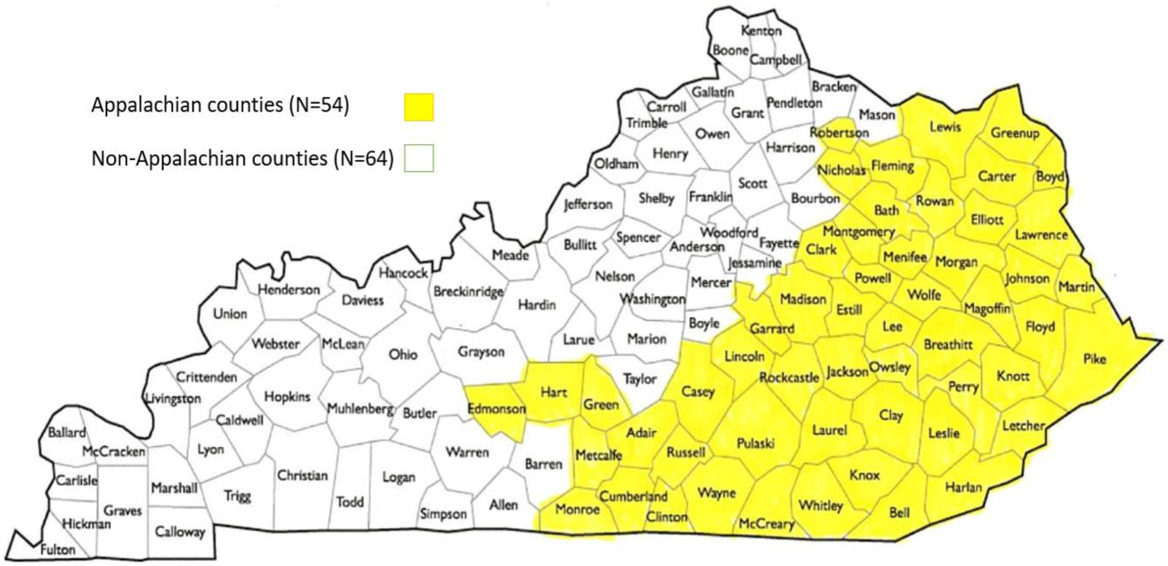


Figure 1-2: The geographic distribution of the Appalachian and non-Appalachian counties in Kentucky. Source: The Appalachian Regional Commission (ARC).

1.6 Structure of the Dissertation

This dissertation investigates the association of social environment, physical environment, and health behavior with lung cancer mortality in Kentucky. Chapter one sets out the context, the background to social, physical, and health behaviors, and its definition, describing lung cancer's key risk factors, including different approaches to measuring various aspects of socioeconomic inequalities and their association with lung cancers mortality rate. It also identifies the debates in the literature and sets out the goals of this dissertation.

Chapter two provides a detailed literature review of the evidence of inequalities in lung cancer mortality. This section offers narrative literature for social environment risk factors, the physical environment and health behaviors, and their relationship with lung cancer mortality.

Chapter three illustrates the data sources used in this study and four methodological aspects: Joinpoint regression method, ordinary square regression method, geographically weighted regression method, Geodetector method, and Health disparity calculator.

Chapter four evaluates the result of lung cancer incidence and mortality trend by gender through 2002 – 2019, using the county-level data from Surveillance, Epidemiology, and End Results (SEER) Program. This section identifies and explains the changes in different periods throughout trends in data, and also illustrate recent changes in male and female lung cancer mortality trends in Kentucky.

Chapter five demonstrates the result of the association of several lung cancer risk factors with cancer mortality, such as the importance of each risk factor and their statistical association between lung cancer mortality and geographic pattern of lung cancer mortality in the Appalachian and non-Appalachian regions. The independent variables include four socioeconomic factors: adult smoking rate, medium household income, high school graduation rate, and the number of coal mine employment.

Chapter six justifies the spatial variation of lung cancer mortality in Kentucky. This section demonstrates the four geographical detectors-based assessment on spatial variation analysis of the geographical strata to assess social, physical and health behavior risks on lung cancer mortality. In addition, this section explains the lung cancer risk areas, which risk factors are responsible for lung cancer mortality, relative importance between the risk factors, and their interaction with each other.

Chapter seven reveal the result of different measures of lung cancer disparity in Kentucky. This section explains the influence of six measures of relative disparity and four measures of absolute disparity.

Chapter eight brings together and discusses the results of all four included studies, compares the results with the existing body of literature, and describes the implications and limitations of the study.

Chapter 2

Background and Literature Review

This chapter aims to prepare detailed literature on the association of social environment, physical environment, and health behaviors with lung cancer mortality in Kentucky. The first section of the study discusses lung cancer incidence and mortality trends and recent changes in the United States and Kentucky.

The second section demonstrates the evidence of the previous research for the relationship of social environment, physical environment, and health behavior risk factors with lung cancer mortality, including different approaches to measure socioeconomic inequalities. Then, the third section of this chapter investigates previous research on spatial heterogeneity and its influence on lung cancer risk factors. Finally, this chapter provides a detailed narrative literature review of the evidence of disparity in lung cancer mortality, identifies the debates in the literature, and provides rationale for the research objectives.

2.1 Worldwide Lung Cancer Estimates of Incidence and Mortality

Lung cancer remains a leading worldwide health problem, accounting for more than a sixth of cancer deaths. Lung cancer is considered the most common malignant neoplasm globally (12.8% of all new cancer cases and 17.8% of cancer deaths) (Hoffman, Mauer, & Vokes, 2000). The global geographical patterns in lung cancer deaths strongly follow those in incidence because of poor survival and the high mortality rate of this disease (Figure 2-1 and 2-2). Worldwide, lung cancer is the primary cause of cancer death in men and the second-leading cause in women (Schabath & Cote, 2019). In 2018, a projected 1.8 million deaths happened (1.2 million in men and 576,100 in women), reporting 1 in 5 cancer deaths worldwide (Bray et al., 2018).

In men, higher incidence rates are observed in Western Europe and North America. In women, the highest rates are found in North America and Northwestern Europe. In Europe, lung cancer is the leading cause of cancer mortality in men and it is the third leading cause of death in women (Levi, Lucchini, La Vecchia, & Negri, 1999).

The geographical variations by country/region and between men and women are primarily attributed to historical patterns in tobacco smoking and the maturity of the tobacco epidemic (Bray et al., 2018). According to Figure 2-1, Lung cancer mortality among males is highest in Western Asia, Eastern Europe, Northern Africa, and particular countries in Eastern Asia and lowest in most of Africa (Schabath & Cote, 2019).

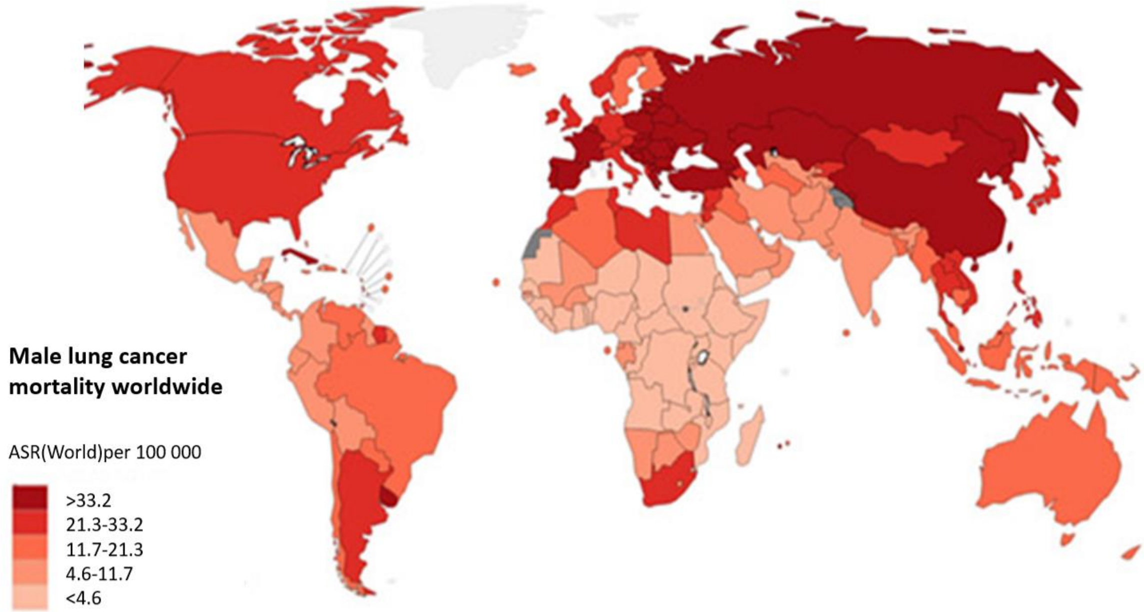


Figure 2-1: Male lung cancer mortality worldwide. Data source: Global Cancer Observatory (GLOBOCAN) 2018

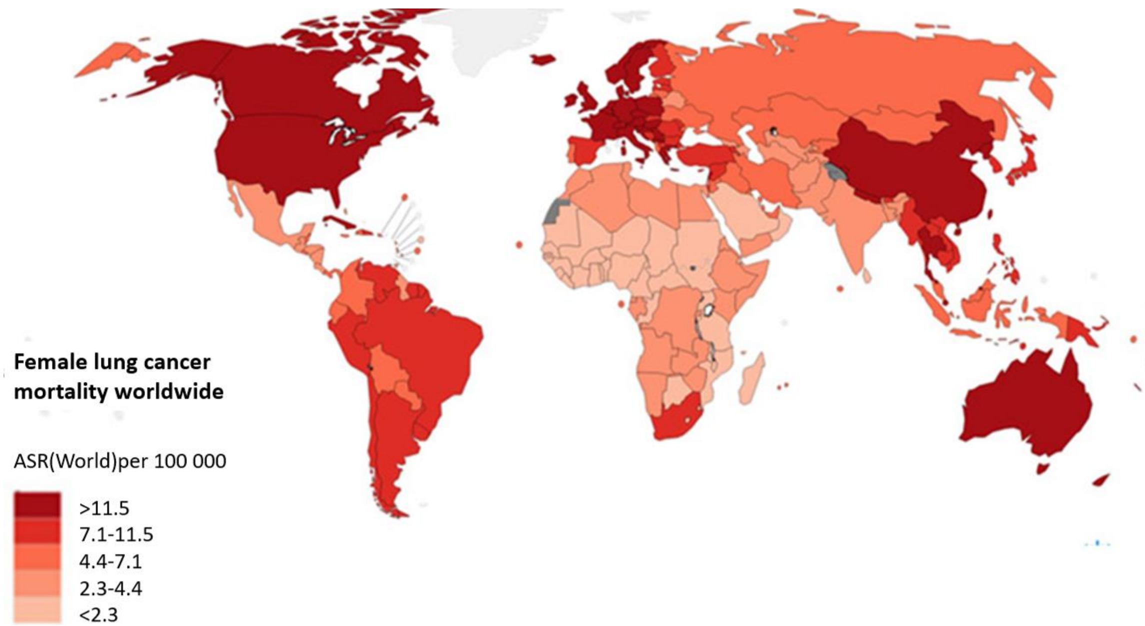


Figure 2-2: Female lung cancer mortality worldwide. Data source: Global Cancer Observatory (GLOBOCAN) 2018

According to Figure 2-2, lung cancer mortality among females is highest in North America, Western Europe, Northern Europe, and Australia/New Zealand and lowest in most African countries (Schabath & Cote, 2019).

2.1.1 Lung Cancer Estimates in the United States

The proportion of adenocarcinomas cancers is increasing in North America and some parts of Europe (Hoffman et al., 2000). In the United States, Figure 2-3 displays the most common cancers diagnosed in men and women in 2016. For men, the three most diagnosed cancers were prostate, lung & bronchus, and colon cancer.

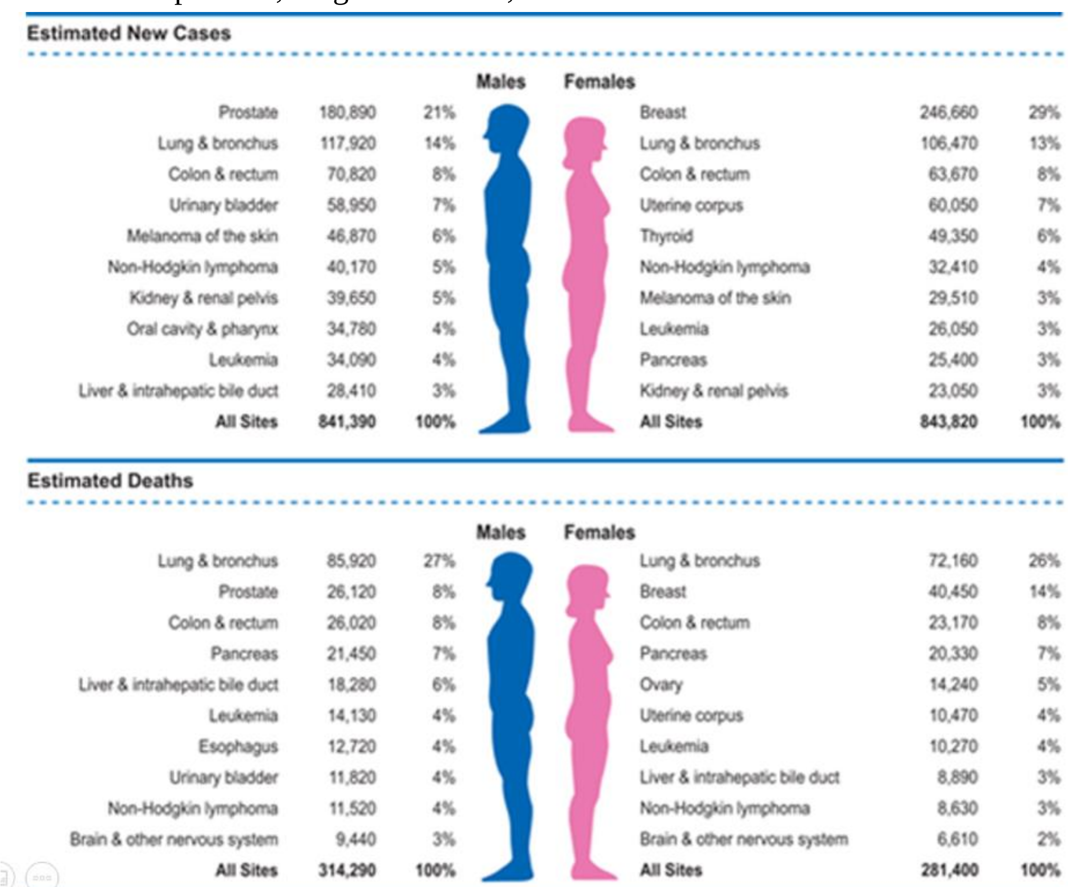


Figure 2-3: Leading cancer types for the estimated new cancer and deaths by gender, United States,2016. Source - (Siegel, Miller, & Jemal, 2016)

Lung and bronchus cancers account for 14% of all cases in men. For women, the three most diagnosed cancers are breast, lung, and bronchus cancers with colorectum, representing one-half of all cases. Lung cancer alone is expected to account for 13% of all new cancer diagnoses in women (Siegel et al., 2016).

In the United States, lung and bronchus cancer is the prominent cause of cancer-related death among men and women (Siegel, Miller, & Jemal, 2019). In 2019, an estimated 142,670 deaths were expected to occur, or about 23.5% of all cancer deaths. The lung cancer mortality rate among men is 51.6 per 100,000 and 34.4 per 100,000 for women. As a result of reductions in smoking, the lung cancer death rate declined 48% since 1990 in men and 23% decline since 2002 in women. From 2012 to 2016, the mortality rate dropped by about 4% per year in men and 3% per year in women (Schabath & Cote, 2019).

Geographically, lung cancer mortality follows a pattern parallel to incidence, with the highest rates observed in the South (Figure 2-4 and Figure 2-5). The lung cancer morality ratio among both males and females is higher in the Midwest, East, and South and lowest in most Mountain states and California.

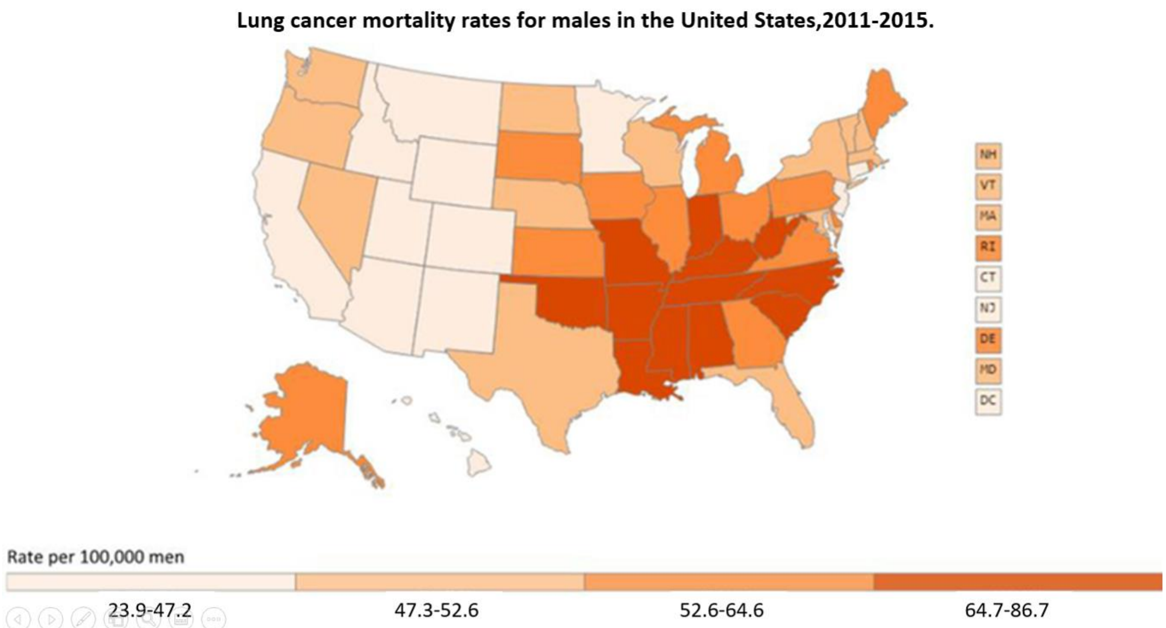


Figure 2-4: Lung cancer mortality rates for males in the United States, 2011-2015.

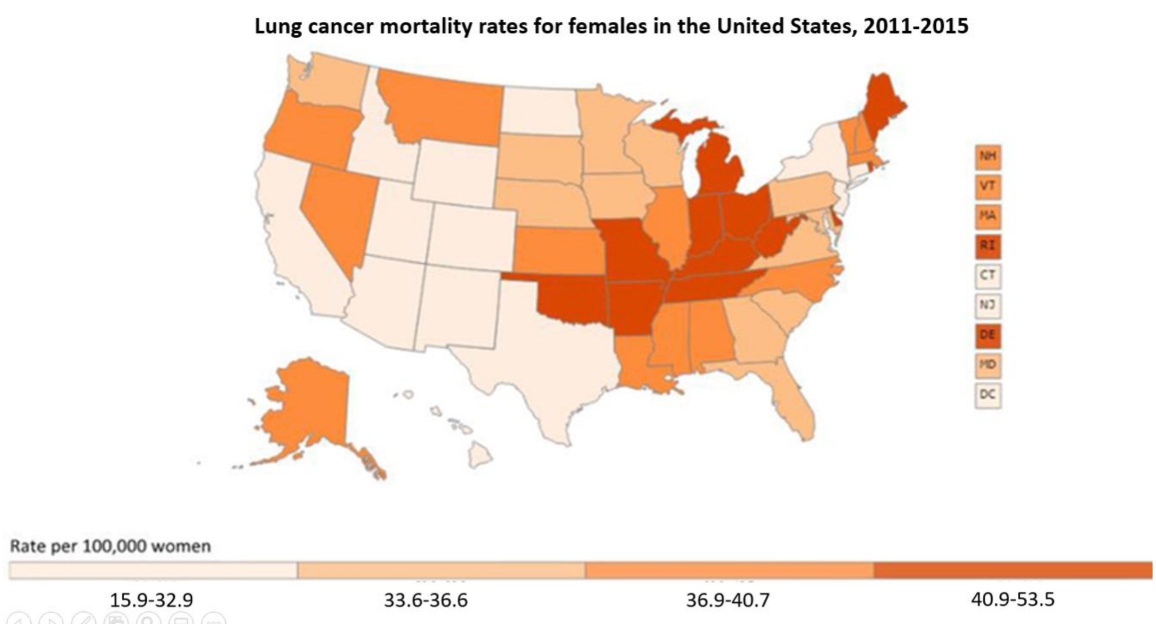


Figure 2-5: Lung cancer mortality rates for females in the United States, 2011-2015.

(Data source: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute (www.cdc.gov/cancer/dataviz))

Siegel et al (2016) recently estimated that the country could expect approximately 224 000 new cases and 158 000 deaths each year (Siegel et al., 2016). In 2018, the age-adjusted lung cancer mortality rate in the United States was 34.8 per 100,000 people. Twenty-one states had a higher lung cancer mortality rate than the national rate, 15 states and DC had lower death rates, and 14 states had rates that were not statistically different from the national rate. Most states with higher mortality rates were in the Midwest or Southeast (Figure 2-6). The five states with the highest lung cancer mortality ratio were Kentucky (53.5), West Virginia (50.8), Mississippi (49.6), Arkansas (47.4), and Oklahoma (46.8). Conversely, the five jurisdictions with the lowest lung cancer mortality rates belong to Utah (16.4), New Mexico (22.5), Colorado (23.0), DC (24.6), and California (25.0) (National Center for Health Statistics. National Vital Statistics System, 2018).

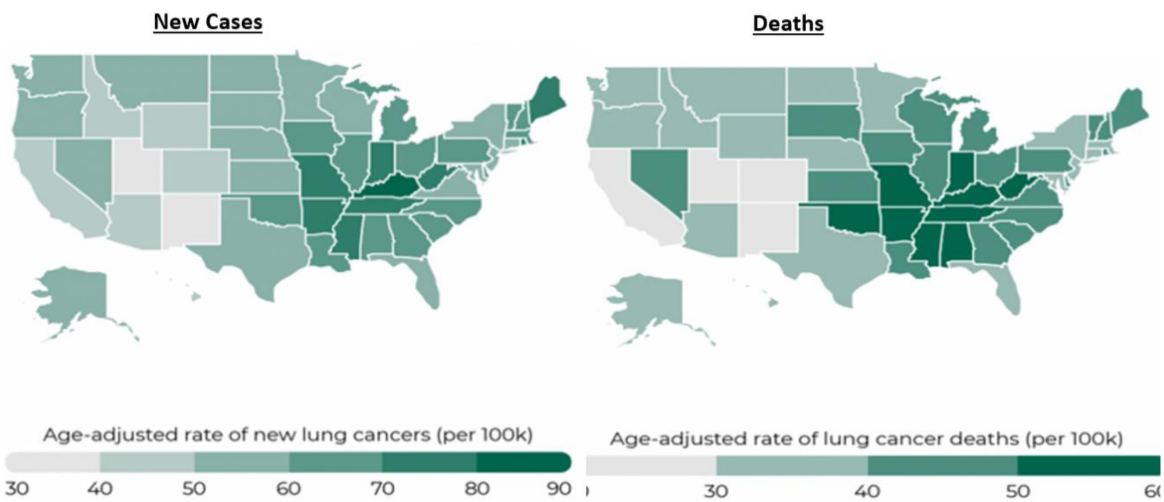


Figure 2-6: Rates of lung cancer incidence and mortality in the US counties.

(Source: center for disease and prevention U.S cancer statistic dataset and the U.S bureau 2009 American community survey)

States in the Southeast, particularly in the Appalachian region, lead the United States in new cases and mortality. For example, the top five states for new patients and lung cancer deaths are Kentucky, West Virginia, Arkansas, Mississippi, and Tennessee (Figure 2-6) (Kentucky Cancer Registry, 2019).

2.1.2 Lung Cancer Estimates in Kentucky

According to cancer statistics, Kentucky has the highest cancer incidence and mortality rates in the United States, and lung cancer is the prominent cause of cancer deaths in Kentucky (Knight, Williamson, Armstrong, & Westbrook, 2019). For example, in 2011–2015, the total lung cancer incidence was 94 per 100,000 population in Kentucky compared to 60.2 per 100,000 population in the United States (U.S. Cancer Statistics: Data Visualizations, November 2017). Between 2011–2015 the average number of overall lung cancer mortality in Kentucky was 3,460 per year. In 2011–2015, Kentucky's overall age-adjusted lung cancer mortality rate was 67.3 per 100,000 population compared to 43 per 100,000 population in the United States (Kentucky Cancer Registry, 2019).

During 2011-2015, the lung cancer incidence rate for males in Kentucky was 113.6 per 100,000 population and 71 per 100,000 population for males in the United States. However, females in Kentucky were 79.3 per 100,000 population and 52 per 100,000 population in the United States (Kentucky Cancer Registry, 2019). Lung cancer incidence rate is 1.43 times higher among males than females in Kentucky (Knight et al., 2019).

During the same period, the age-adjusted lung cancer mortality rate for males in Kentucky was 86.1 per 100,000 population and 54 per 100,000 population in the United States

(Kentucky Cancer Registry, 2019). The age-adjusted lung cancer mortality rate for females was 53.1 per 100,000 population in Kentucky and 35 per 100,000 population in the United States (Kentucky Cancer Registry, 2019; U.S. Cancer Statistics: Data Visualizations, November 2017). The lung cancer mortality rate is 1.62 times greater among males than females in Kentucky. There is a considerable need to address lung cancer disparities in both males and females in Kentucky (Knight et al., 2019).

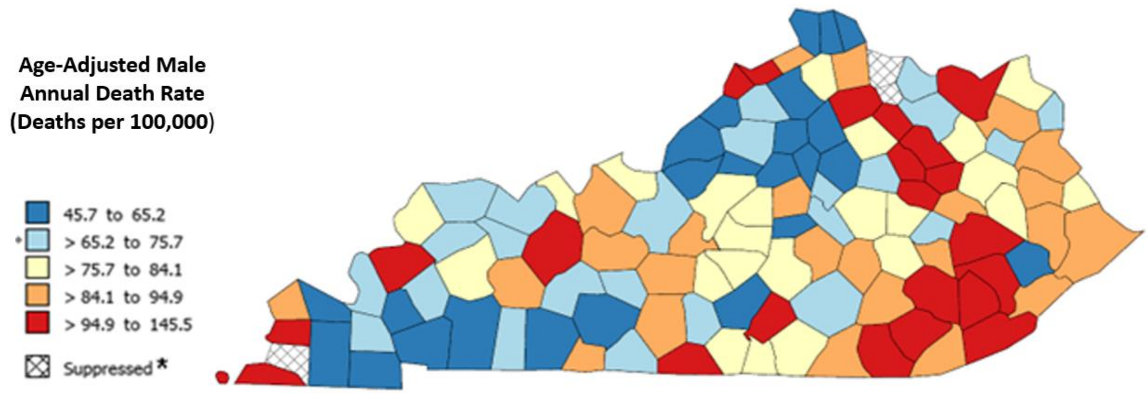


Figure 2-7: Male lung cancer mortality rates for Kentucky, 2001-2015. Source: Death data provided by National Vital Statistic System.

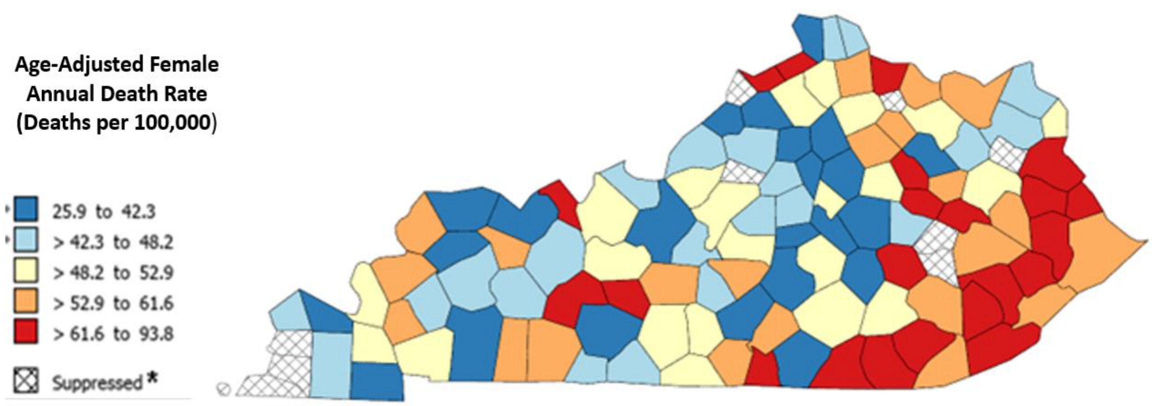


Figure 2-8: Female lung cancer mortality rates for Kentucky, 2001-2015. Source: Death data provided by National Vital Statistic System.

2.1.3 Appalachian Region and Non-Appalachia

A relationship between lung cancer incidence and mortality rates is observed. The highest rates of both are in the exact geographic location. Generally, Appalachia carries a higher cancer burden compared with non-Appalachia, particularly for tobacco-related cancers. In addition, for all cancer sites combined, Appalachia has higher rates regardless of gender, race, or region (R. J. Wilson, Ryerson, Singh, & King, 2016).

Appalachia comprises 420 counties in 13 states and spans 205,000 square miles, from southern New York to northern Mississippi. The Region's 25 million people live in parts of Alabama, Georgia, Kentucky, Maryland, Mississippi, New York, North Carolina, Ohio, Pennsylvania, South Carolina, Tennessee, Virginia, and West Virginia. (Figure 2-9)

The region of Kentucky designated as Appalachia is set by the Appalachian Regional Commission (ARC) and includes 54 counties in the state's Eastern region. Such as Adair, Bath, Bell, Boyd, Breathitt, Carter, Casey, Clark, Clay, Clinton, Cumberland, Edmonson, Elliott, Estill, Fleming, Floyd, Garrard, Green, Greenup, Harlan, Hart, Jackson, Johnson, Knott, Knox, Laurel, Lawrence, Lee, Leslie, Letcher, Lewis, Lincoln, McCreary, Madison, Magoffin, Martin, Menifee, Metcalfe, Monroe, Montgomery, Morgan, Nicholas, Owsley, Perry, Pike, Powell, Pulaski, Robertson, Rockcastle, Rowan, Russell, Wayne, Whitley, and Wolfe counties belong to Appalachian region (Figure 2-9) (Appalachian Regional Commission, 2021)

According to the Appalachian Regional Commission (APC), most of Kentucky's Appalachian counties are under significant economic hardship, which has been associated with overall poor health (Hosseinpoor et al., 2012). In addition, the Appalachian region has been identified as a medically underserved region due to the region's financial, geographic, and health system challenges (Denham, Meyer, Toborg, & Mande, 2004).

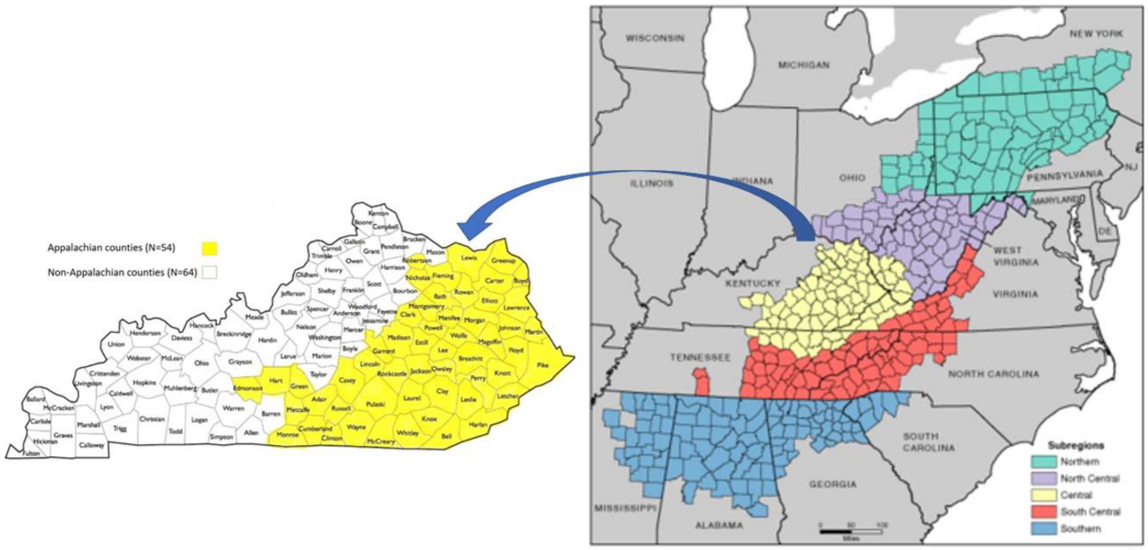


Figure 2-9: Appalachian region and non- Appalachian region. Data Source: The Appalachian Regional Commission (ARC).

The ARC reported in 2017 that Appalachian Kentucky's cancer mortality rate was 35% higher than the national rate and 18% higher than the rate in non-Appalachian Kentucky. As a result, factors related to social determinants of health have been researched to determine their impact on lung and bronchus cancer patients in Appalachian, Kentucky. Specifically, researchers have focused on lifestyle choices, environmental factors, and public policy to examine various reasons why incidence and mortality rates are historically more significant in Appalachian Kentucky as opposed to the rest of the state (Appalachian Regional Commission, 2017). There's a great need for understanding the correlation between these rates and the reasons behind the association. By understanding the causes that lead to high incidence and mortality rates of lung and bronchus cancers in this area of

Kentucky, efforts could be made through public policy to reduce these rates, which would be vital for the increased health of eastern Kentuckians.

2.2 The Impact of Social Environment, Physical Environment, and Health Behaviors on Health

Many factors affect how well and how long we live. Neighborhood physical environment, social environment, and health behaviors have been recognized as essential factors shaping health. Health Factors can be modified to improve the length and quality of life for residents. They are predictors of how healthy our communities can be in the future. No one factor influences the overall health of an individual or community. A combination of multiple modifiable factors, from clean air and water to stable and affordable housing, need to be considered to ensure community health for all.

Understanding the association between neighborhoods and health outcomes has been illustrated in several conceptual frameworks. First, this framework shows the relationship of social and built environmental characteristics on the cancer continuum (Figure 2-10). Then, the main inter-related components of social and built environments and their subcomponents are identified and explain how these neighborhood characteristics can impact the cancer continuum (Scarlett Lin Gomez et al., 2015).

Social and economic aspects, such as education, income, poverty, employment, community safety, and social supports, can significantly affect how well and how long we live. These factors affect our capability to make healthy choices, afford medical care and housing, manage stress, and more.

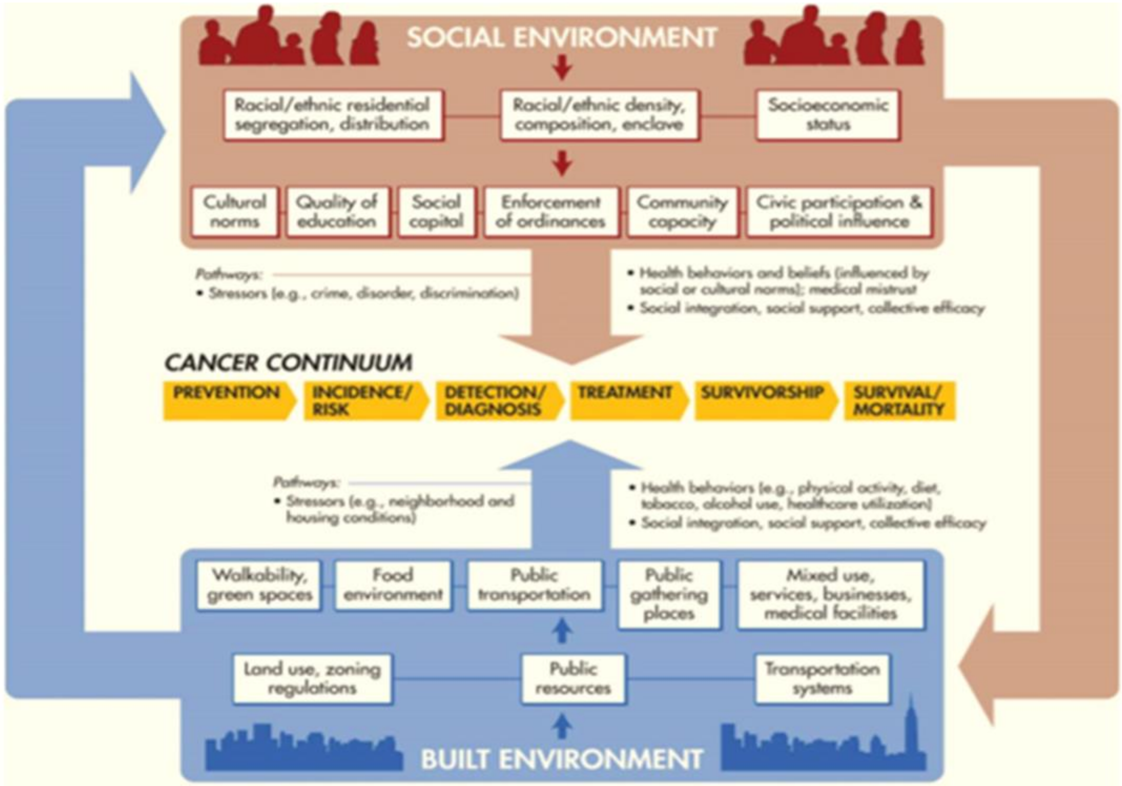


Figure 2-10: The impact of neighborhood social and built environment factors across cancer. Source: (S. L. Gomez et al., 2015).

An individual's social environment can negatively influence a person's health leading to obesity, cancer, mental health problems, and a higher risk of diseases. Typically, those lower on the social ladder are twice as likely to develop a health condition. A poor social environment can make a person feel anxious and stressed, leading to physical medical conditions in the long term. The work of Yen et al (1998) showed that aspects of the neighborhood environment contributed independently to overall mortality (Yen & Kaplan, 1998). A person's education, occupation, and income status are social environment factors that can weigh heavily on an individual. A person with a low-income occupation may not afford certain aspects that keep an individual healthy, such as clean housing, nutritious foods, and access to health care that are typically more costly.

The physical environment is where individuals stay, learn, work, and play. People interact with their physical environment through the air they breathe, the water they drink, houses they live in, and the transportation they use to travel to work and school. A poor physical environment can affect our families and neighbors' ability to live long and stay healthy (County Health Rankings & Roadmaps, 2022c).

Clean air and safe water are essential for good health. Air pollution and radon are associated with enhanced asthma rates and lung diseases, and an increase in the risk of premature death from lung or heart disease. Water contaminated with chemicals, pesticides, or other contaminants can lead to illness, infection, and increased cancer risks. Urban green and blue areas also provide opportunities for stress recovery and physical activity. Natural environments offer spaces for social interactions in the neighborhood and places for children's play. Stress, physical inactivity, and lack of social structure are three major risk factors for noncommunicable diseases, and therefore abundant urban greenery is an essential asset for health promotion (County Health Rankings & Roadmaps, 2022a).

Health behaviors are actions individuals take that influence their health. They include activities that lead to improved health, such as eating well and being physically active and behaviors that increase one's risk of diseases, such as smoking, excessive alcohol consumption, and unsafe sexual behavior (County Health Rankings & Roadmaps, 2022b).

In the United States, many of the prominent causes of death and disease are attributed to unhealthy behaviors. For example, poor nutrition and low physical activity levels are associated with a higher risk of cardiovascular disease, type 2 diabetes, and obesity. Tobacco use is associated with heart disease, lung cancer, and poor pregnancy outcomes if the mother smokes during pregnancy. Excessive alcohol use is correlated with injuries, certain types of cancers, and cirrhosis (County Health Rankings & Roadmaps, 2021a).

It is essential to consider that not everyone has the means and opportunity to make healthy decisions. In addition, policies and programs have marginalized some population groups and communities, keeping them from the supports and resources necessary to thrive. Therefore, addressing health behaviors requires strategies to encourage individuals to engage in healthy behaviors and ensure that they can access nutritious food, safe spaces to be physically active, and supports to make healthy choices.

Comprehensive policies, programs, systems, and environmental changes can make a change locally. Some interventions focus on individual behaviors, such as influencing dietary choices, exercise levels, or alcohol consumption. Other strategies try to tackle systems and structures, such as enhancing opportunities for education, stimulating economic development, and increasing neighborhood safety (County Health Rankings & Roadmaps, 2021a).

2.3 Social Environment and Inequalities in Health

Since the late nineteenth century, the socioeconomic position has been believed as an essential factor in cancer epidemiology (Yang Mao et al., 2001). SES reflects one’s place in societal hierarchies and is generally evaluated by the interdependent dimensions of education, occupation, and income. It is measured by how many years are spent in school (less than high school, high school, college, graduate school, etc.), yearly incomes, and whether they are employed or unemployed. For example, a person with a high SES may have a graduate school degree, a higher-than-average income, and a steady full-time job. In contrast, a person with a low SES may have less than a high school education, not have enough money to lead a comfortable life, and be unemployed or work in a low-paying job.

One study conducted in the United States (US) indicated that income inequality was associated with a lack of social trust and higher age-adjusted mortality rates from various chronic illnesses, including cancer (Kawachi & Kennedy, 1997). In addition, due to these socioeconomic inequalities, overall life expectancy and healthy life expectancy are considerably shorter among more socioeconomically deprived groups relative to more wealthy groups (Marmot, 2005).

This socioeconomic gradient indicates the social pattern of disease across all groups in society and the social strata. This relationship exists in lower- and middle-income countries (Bangal, Giri, Bangal, More, & Singh, 2014) and high and middle-income countries (Arnold et al., 2016). It also continues within and between countries (Mackenbach et al., 2008), suggesting that there is not an absolute level of poverty associated with poor health but a linear relationship - a “gradient” between socioeconomic circumstances and health (Watt, 2002).

SES is associated with health/disease through multiple interacting pathways in material and social resources, physical and psycho-social stressors, and health-related behaviors (P.

Braveman, Egerter, & Williams, 2011). Given this stepwise socioeconomic gradient's consistent and persistent nature, many diseases, including cancer (Marmot 2005), have a more significant incidence and mortality burden among lower socioeconomic groups than those of higher socioeconomic groups (Watt & Sheiham, 2012).

The relationship between SES and ill health is so well established that epidemiologists would almost always adjust by SES in the same way they change for age and gender when exploring the effect of other risk factors for disease. A person with a high SES is more likely to have insurance and sick leave through their employment. Therefore, they are more likely to access preventative services such as cancer screening and tobacco cessation services. Research has also found that people with a high SES are more likely to have higher survival rates because they are prone to early cancer diagnosis and treatment. On the other hand, people with a low SES may not get necessary cancer screenings and have cancer diagnosed at later stages, leading to lower cancer survival rates. People with a low SES may not go to the doctor for a variety of reasons. These may include not having access to transportation for a doctor visit, being worried about their screening tests, not being able to take off work to see a doctor, etc. (Pampel, Krueger, & Denney, 2010). While lung cancer patients are from disadvantaged populations, their survival from lung cancer is poor. Quality of life is considered an essential outcome in patients who develop lung cancer (Montazeri, Gillis, & McEwen, 1998).

2.3.1 Relationship Between Education and Lung Cancer

In recent years, awareness has been drawn to assess the education and association between low socioeconomic status (SES) and expanded risk of chronic lung diseases. Indeed, based on studies from Norway, South Korea, and different European cities, these diseases were more common in deprived communities and with people with low levels of education (Marí-Dell'Olmo et al., 2015; Strand et al., 2010). In addition, an analysis done using 16 European populations reported higher lung cancer mortality rates in groups with the lowest educational attainment (Van der Heyden et al., 2009).

In a study assessing the effect of education on smoking, researchers in Europe categorized education into high and low education. The high education group contained people who were college graduates or had professional degrees. The low education group included people with no education or people who never finished high school. In their analysis, the authors found that current male and female smokers in the low education group had odds ratios of 1.65 and 1.18, respectively, compared to the highly educated group. The result indicates a higher smoking prevalence among the low educated group (Cavelaars et al., 2000).

In a study performed in Finland, researchers found that smoking was widespread among participants with low education, low income, economic difficulties, and economic dissatisfaction. The prevalence of smoking across the college, high school, and less than high school levels was 23%, 26%, and 35% for men and 13%, 20%, and 30% for women. The odds ratio for smoking was 1.73 for men and 2.92 for women, with the lowest education level compared to those who had college degrees. The odds ratio for smoking amongst the lowest income level was 2.04 for men and 1.58 for women compared to the highest income level. Education level is an essential socioeconomic indicator because it reflects the skills and knowledge required to make healthy choices concerning smoking (Laaksonen, Rahkonen, Karvonen, & Lahelma, 2005). In a case-control study evaluating the risk factors for lung cancer in Iowa women, the authors discovered that women with a college education had 0.63 times the odds of having lung cancer as women without a college education (Neuberger, Mahnken, Mayo, & Field, 2006).

The study conducted in the Netherlands examined the effects of socioeconomic inequalities on smoking prevalence, initiation, and cessation. Researchers found that lower educated respondents were more likely to be smokers and have higher initiation ratios and lower quit ratios than higher educated study participants. For example, smoking prevalence was 29% among lower education participants compared to 20% among higher educated participants. In addition, for men, the odds ratio of smoking was 1.84, and for women, the odds ratio was 2.26 in the low education group compared to the high education group (Nagelhout et al., 2012).

2.3.2 Relationship Between Income and Lung Cancer

Income can come from employment, investments, government assistance packages, or retirement plans. Income allows families and individuals to obtain health insurance and medical care and provides healthy lifestyle choices. Unfortunately, low-income families and individuals probably live in unsafe neighborhoods, often with limited access to healthy foods, employment options, and quality schools (County Health Rankings & Roadmaps, 2021a). In addition, lower-income groups have less access to health care, which may cause them to be diagnosed at later stages of diseases and conditions.

In a national case-control study conducted in Canada, researchers found that the odds of having lung cancer among both males and females was significantly higher among people belonging to a low-income background (males 1.7 and females1.5). In addition, both male and female study participants who had more than 14 years of education had an odds ratio of 0.6 compared to those who had less than eight years of schooling. This study concluded that males who had unskilled jobs and belonged to a lower SES had substantially higher odds of having lung cancer when compared to males with a professional job and belonged to a higher SES (Mao, Hu, Ugnat, Semenciw, & Fincham, 2001).

A study conducted using death records from the National Center for Health Statistics in 2014 found that in 3135 US counties, cancer death rates varied significantly in counties with different income levels. For example, the mean cancer death rate per 100 000 person-years is 185.9 in high-income counties, 204.9 in medium-income counties, and 229.7 in low-income counties. The strongest possible facilitators were health risk behaviors, cost and quality of clinical care, and food insecurity (O'Connor, Sedghi, Dhodapkar, Kane, & Gross, 2018). Also, a study conducted using cancer patients diagnosed in 1973–2001 found that those with annual family incomes fewer than $12,500 had a lung cancer incidence ratio that was more than 1.7 times the lung cancer incidence ratio of those with incomes $50,000 or higher (Clegg et al., 2009).

The effect of socioeconomic differences on cancer survival has been examined for several cancer types showing lower cancer survival in patients from low-income groups. A study conducted using meta-analyses revealed a poorer diagnosis for patients with low individual income. Findings suggest a weak positive association between personal income and lung cancer survival (Finke, Behrens, Weisser, Brenner, & Jansen, 2018).

Communities can adopt and employ policies that help reduce and prevent poverty now and for future generations. The most significant health improvements may be made by increasing income at the lower levels, where small increases can have the most significant impacts (County Health Rankings & Roadmaps, 2021a).

2.3.3 Relationship Between Unemployment Rate and Lung Cancer

Unemployment has become an essential element among the socioeconomic determinants of health. According to the study done by Wilson & Walker (1993) unemployed men and their families have increased mortality experience, particularly from suicide and lung cancer. Unemployed men also experience reduced psychological well-being with a greater incidence of parasuicide, anxiety, and depression. Unemployed men are less likely to visit a general practitioner or hospital and receive more prescribed medicines. Smoking and alcohol consumption are frequently increased after the onset of unemployment. Women are less affected by enforced unemployment, but families with someone unemployed are at greater risk of physical illness, psychological stress, and breakdown. Maintaining financial security, providing proactive health care, and retraining for re-employment can all reduce the impact of unemployment on health (S. H. Wilson & Walker, 1993).

According to the study of Lynge (1997) unemployed men have excess cancer mortality of close to 25% compared with that of all men in the labor force. The available data from various countries indicate that this additional risk is found in periods when the unemployment rate is about 1% and in periods when it is about 10%. Furthermore, excess cancer mortality comes mainly from lung cancer, and the increased risk of lung cancer does not disappear when social class and the number of previous sick days are controlled. Also, the result reveals that unemployment does not increase smoking, but unemployed men have a slightly higher smoking prevalence before unemployment (Lynge, 1997).

2.3.4 Relationship Between Poverty and Lung Cancer

Poverty is associated with a massive array of human health problems and seriously undermines underprivileged populations' health. Limited financial resources in poor communities are frequently subjected to environmental risks due to the unavailability of suitable housing. As a result, they are less well-nourished, have less education, and have limited health care and appropriate insurance access. As a result, they consistently have a higher incidence of numerous illnesses (Heidary, Rahimi, & Gharebaghi, 2013).

In the last 50 years, lung cancer mortality has continued to increase in the lower socioeconomic groups but has decreased in more socioeconomically favored groups (Smith, Leon, Shipley, & Rose, 1991). As documented in the annual “Cancer Facts and Figures 2011” published by the American Cancer Society, poverty persists as one of the most potent carcinogens. These reports concluded that poverty is the initial contributing factor to cancer disparities among social groups and that racial differences in biological or inherited characteristics are less significant. The fact is that people living in poverty lack access to health care and subsequently endure more significant pain and illness (Heidary et al., 2013).

Clear evidence from industrialized and less developed societies demonstrates that cancer incidence and survival are related to socioeconomic circumstances. Lower social classes with high poverty rates tend to have a higher cancer incidence and poorer cancer survival overall rates than higher social classes. However, this pattern differs for specific cancers (Heidary et al., 2013).

2.3.5 Relationship Between Occupational Exposure and Lung Cancer

Some people are subjected to carcinogens (cancer-causing agents) such as arsenic, uranium, asbestos, and diesel discharge at their workplace. The relationships between occupational exposures to coal mine dust and mortality from coal workers' pneumoconiosis and chronic obstructive pulmonary disease have been established (Cohen & Velho, 2002). The mortality risk of lung cancer for coal miners has also been assessed in a series of epidemiological studies (Miller & MacCalman, 2010). Research conducted in China suggested that exposure to occupational dust might increase the mortality risk of lung cancer, especially for Asian populations in China (Li, Jiang, Li, & Zhou, 2021). Population-based ecological and cross-sectional studies have observed a high risk for several cancers in areas of Central Appalachia where mountaintop removal coal mines operate (W. J. Christian et al., 2020). Studies suggest that living near coal mining sites could increase the risk for lung cancer after adjusting for other relevant factors (W. J. Christian et al., 2020).

Occupational and environmental exposures might influence lung cancer patterns. For example, many residents in the Appalachian region rely on private wells for drinking water (Hopenhayn-Rich, Stump, & Browning, 2002). This issue puts them at risk of exposure to trace elements from natural or artificial sources (e.g., arsenic, nickel, and chromium), that are possible lung carcinogens. In addition, workers in the extensive mining industry are likely exposed to coal and silica dust, linked to various lung diseases (Ross & Murray, 2004).

Work-related exposure to such cancer-causing materials has reduced as the government and industry have taken steps to help protect workers. Still, we need to be careful to limit release whenever possible (American Cancer Society, 2020).

2.4 Health Behaviors and Inequalities in Health

Health behaviors are actions individuals take that affect their health, including activities that improve health, such as quitting smoking, health insurance, and being physically active. But some health behaviors may increase one’s risk of diseases, such as smoking, excessive alcohol intake, and risky sexual behavior. In the United States, many of the leading causes of death and illness are attributed to unhealthy behaviors. For example, poor nutrition and low physical activity levels are associated with a higher risk of cardiovascular disease, type 2 diabetes, and obesity. In addition, research evidence suggests that people with lower socioeconomic position (SEP) engage in fewer health-promoting behaviors (Beenackers, Oude Groeniger, van Lenthe, & Kamphuis, 2018).

Tobacco use is the prominent cause of preventable death in the United States. It affects those who use tobacco and people who live and work around tobacco (County Health Rankings & Roadmaps, 2021a). Each year, smoking kills 480,000 Americans, together with 41,000 from exposure to secondhand smoke. In addition, smoking causes cancer, heart disease, stroke, lung diseases, diabetes, and chronic obstructive pulmonary disease, including emphysema and chronic bronchitis. Usually, smokers die ten years earlier than nonsmokers (centers for Disease Control and Prevention, 2019).

A research study using 1,681 lung cancer patients suggests that patients without insurance are diagnosed at later stages. This late diagnosis is the primary driver of poor survival. Although underinsured or uninsured relates to a greater risk of death after diagnosis, adjusting for stage mitigates this effect. These findings encourage the need for equal access to early screening and proper health insurance (Mohamed, Herndon, Schmidt, & Manning, 2020).

2.4.1 Relationship Between Smoking and Lung Cancer

Most lung cancers are associated with lifestyle choices like smoking. Cigarette smoking is the number one lung cancer risk factor. In the United States, cigarette smoking is associated with about 80% to 90% of lung cancer fatalities. Using different types of tobacco products such as cigars or pipes also increases the risk of lung cancer. Tobacco smoke is a toxic mixture that contains more than 7,000 chemicals. Many are harmful poisons. At least 70 are known to affect cancer in people or animals (Division of Cancer Prevention and Control, 2020). Not all people who have lung cancer smoke, but 20% of people die from lung cancer in the United States. However, lung cancer in people who have never smoked is one of the fatal cancers in the United States (American Cancer Society, 2020).

There is a higher concentration of smokers among people in lower SES. Smoking prevalence increases with decreasing SES (Singh, Williams, Siahpush, & Mulhollen, 2011). In a study conducted in Rhode Island, researchers found that the influence of SES on persistent smoking accumulates over the individual's lifespan. The results showed that lower SES was associated with increased odds of first cigarette use. In addition, lower adult SES increased the probability of becoming a regular smoker (Gilman, Abrams, & Buka, 2003).

In a study conducted in Tennessee examining the association between SES and smoking, researchers found that individuals who had some college or more education were 0.60 times more likely to smoke when compared to individuals who had a high school degree or less. The study also found that participants belonging to neighborhoods with higher education levels were less likely to smoke. (Scarinci, Robinson, Alfano, Zbikowski, & Klesges, 2002).

Smoking in the United States has dropped in recent decades. From 2005 to 2019, cigarette smoking among US adults dropped from 21% to 14% (Cardarelli et al., 2021). However, the decline in cigarette smoking has not been experienced uniformly across US communities; instead, smoking rates have declined more rapidly in urban compared with rural areas (Doogan et al., 2017). In rural regions, 28.5% of adults say they smoke cigarettes, compared with 25.1% of urban adults (Vander Weg, Cunningham, Howren, & Cai, 2011). Rural residents in the US are more prone to smoke than non-rural residents

(Doogan et al., 2017). Furthermore, rural residents demonstrate greater cigarette smoking intensity than urban residents (Roberts et al., 2016).

Rural areas have higher smoking levels than urban areas, most likely caused by the demographic and psychosocial factors typically associated with rural areas, such as lower income and education levels and higher unemployment (Buettner-Schmidt, Miller, & Maack, 2019). Additionally, Doogan et al (2017) found that tobacco control policies and other regulatory aspects promote urban regions more than rural ones (Doogan et al., 2017). Furthermore, tobacco crops are the primary income resource for many rural areas; thus, tobacco is more normalized into the culture (Buettner-Schmidt et al., 2019).

In the Kentucky Central Appalachia region, smoking rates have remained high over the last several decades (Appalachian Regional Commission, 2019). In 2017, 24.6% of adults smoked in Kentucky. Nationally, the rate was 17.1% (CDC, 2017a). In 2017, 14.3% of high school students in Kentucky smoked cigarettes on at least one day in the past 30 days. Nationally, the rate was 8.8% (CDC, 2017c). In 2017, 6.1% of adults used e-cigarettes, and 7.6% used smokeless tobacco (CDC, 2017b).

Implementation of tobacco control policies and programs can motivate users to quit, help people choose not to start, and improve the air quality we all breathe.

2.4.2 Relationship Between Health Insurance and Lung Cancer

Lung cancer is the main reason for cancer death in the US. Considerable improvements in survival have occurred with better medications. Payer status has been recognized as an obstacle to medication access across multiple cancer types, including lung cancer.

Differences in insurance status may provide directly to different cancer outcomes. Furthermore, differences in race, ethnicity, income, education, and other factors related to insurance status may also affect processes and results of care (Kinsey, Jemal, Liff, Ward, & Thun, 2008; Slatore, Au, & Gould, 2010). In the United States, 20% of adults under 65 are uninsured (Slatore et al., 2010).

A study done by Mohamed et al (2020) found that rates per 10 patients diagnosed with lung cancers were that 3.5 had commercial insurance, 3.8 had Medicare, 3.3 had Medicaid, and 5.4 had uninsured patients. Of those uninsured patients, 56.7% presented stage IV cancer compared to full coverage (41.4%). Also, 40.7% of those without insurance or underinsured were current tobacco product users compared to 25.1% with full coverage. Their risk of death is 1.34 times greater for underinsured patients than those with full coverage (Mohamed et al., 2020). This study suggests patients without insurance are diagnosed at later stages of the disease. Late diagnosis is the primary driver of poor survival rates.

A study was conducted using a systematic review of the existing literature to examine the correlation between insurance status and lung cancer practices and outcomes. The result shows that patients with Medicaid or no insurance had poorer lung cancer outcomes, including higher incidence rates, later diagnosis, and lesser survival. Overall, patients with Medicaid or no insurance were less likely to undergo corrective procedures, but patients without insurance were more likely to obtain guideline-concordant care (Slatore et al., 2010).

2.4.3 Relationship Between Physically Unhealthy Days and Lung Cancer

During the late 1980s, the Centers for Disease Control and Prevention (CDC) developed a survey instrument to capture the health-related quality of life measures in a short questionnaire. “Healthy Days” consists of 4 questions asking people how they perceive their recent health. The first question is, would you say that your health is excellent, very good, good, fair, or poor in general? Second, about your physical health, which includes physical illness and injury, how many days during the past 30 days were your physical health not good? Third, how many days during the past 30 days has your mental health not been good? Fourth, during the past 30 days, for about how many days did poor physical or mental health care from doing your usual activities, such as self-care, work, or recreation? (centers for Disease Control and Prevention, 2018).

The study investigated factors associated with patient-reported health-related quality of life with the Healthy Days tool for a patient with Medicare Advantage undergoing treatment for metastatic breast, lung, and colorectal cancer (Casebeer et al., 2019). According to 1567 respondents, the mean number of unhealthy days was 14.0, with 46.2% experiencing frequent sick days. On average, patients reported 10.5 physically and 6.7 mentally unhealthy days.

Also, patients with pain had 83% more unhealthy days than patients without pain; patients with fatigue had 104% more unhealthy days than patients without fatigue. Diarrhea/constipation and shortness of breath also were associated with more unhealthy days. Cancer-related symptoms, most notably pain and fatigue, were associated with worse health-related quality of life for patients with metastatic cancer (Casebeer et al., 2019).

2.5 Physical Environment and Inequalities in Health

The physical environment is where individuals live, learn, work, and play. The factors in the physical environment are essential to health, such as air pollution or vicinity to toxic sites, access to numerous health-related resources (e.g., healthy, or unhealthy foods, recreational resources, medical care), and community design and the "built environment" (e.g., land use mix, street connectivity, transportation systems).

The environment can impact health through physical exposures, such as air pollution. A large body of work has recorded the effects of exposure to particulate matter (solid particles and liquid droplets found in the air) on cardiovascular and lung cancer mortality (Mustafic et al., 2012). The impacts of particulate matter on mortality seem to be consistent across countries. For example, a recent review of studies from the late 1990s to mid-2000s found a reliable opposite relationship between airborne particulate matter and birth weight in Australia, Brazil, Canada, France, Italy, South Korea, Netherlands, the United Kingdom, and the United States (Parker et al., 2011).

A study conducted using meta-analysis of 28 case-control studies, which included 13,748 lung cancer cases and 23,112 controls, indicates that residential radon is a risk factor in all histological types of all lung cancer. Furthermore, this study found strong associations with lung cancer. Therefore, residential radon exposure remains a primary concern worldwide, and applicable measures should be undertaken to decrease radon exposure to ensure the health of environmental conditions and residents (C. Li et al., 2020).

2.5.1 Relationship Between PM 2.5 and Lung Cancer

Many etiologic factors for lung cancer have been identified, such as smoking and exposure to air pollution, cooking fumes, and asbestos. Atmospheric pollution has become increasingly heavy in recent years. Accordingly, more significant numbers of people are paying interest in the air quality around them. PM2.5 (particulate matter with a diameter of 2.5 micrometers or less), one of the most significant indicators for measuring air quality, can penetrate and be retained in lung tissue. Inhalable airborne fragments (PM2.5, PM10) have a statistical correlation with lung cancer (Raaschou-Nielsen et al., 2013), and each 10 $\mu g/m^3$ increase in PM2.5 concentration is correlated with a 15–27% growth in lung cancer mortality (Turner et al., 2011). Thus, it is believed that PM2.5 may represent a new type of etiological factor for lung cancer (Shu et al., 2016). A study based on China results indicated a significant positive correlation between PM2.5 concentration and lung cancer mortality (Cao, Rui, & Liang, 2018).

A study was conducted using meta-analyses for examining the relationship of exposure to $PM_{2.5}$ and PM_{10} with lung cancer incidence and mortality, found that the meta-relative risk for lung cancer associated with $PM_{2.5}$ was 1.09. The meta-relative risk of lung cancer connected with PM_{10} was similar but less precise: 1.08 (Hamra et al., 2014). Analyses done by smoking status showed that lung cancer risk associated with $PM_{2.5}$ was greatest for former smokers (1.44), followed by never-smokers (1.18) and then-current smokers (1.06).

In addition, meta-estimates for adenocarcinoma correlated with $PM_{2.5}$ and PM_{10} were 1.40 and 1.29, respectively (Hamra et al., 2014).

Research conducted in Italy to analyze the association between exposure to outdoor particulate matter with lung cancer found a positive association between PM10 exposure and lung cancer risk (Consonni et al., 2018).

2.5.2 Relationship Between Radon and Lung Cancer

Radon is a colorless, odorless, radioactive gas. It is created naturally from the decay of radioactive elements, such as uranium, and is located in different quantities in soil and rock throughout the world. Radon gas in the ground and rock can move into the air, underground water, and surface water (American Cancer Society, 2022). Therefore, radon exposure comes from being indoors in homes, offices, schools, and other buildings. The radon levels in homes and other buildings depend on the rock and soil characteristics in the area. As a result, radon levels vary significantly in different parts of the United States, sometimes even within neighborhoods (American Cancer Society, 2022). Radon deteriorates quickly, giving off tiny radioactive particles. When inhaled, these radioactive particles can damage the cells that line the lung. Long-term radon contact can lead to lung cancer (National Cancer Institute, 2011).

According to the study done by Lubin et al (1995) in miners, about 40% of all lung cancer deaths may be due to radon exposure, 70% of lung cancer deaths in never-smokers, and 39% of lung cancer deaths in smokers. In the United States, 10% of all lung cancer deaths might be due to indoor radon exposure, 11% of lung cancer deaths in smokers, and 30% of lung cancer deaths in never-smokers (Lubin et al., 1995). Therefore, the study suggests reducing radon in all homes exceeding the U. S. Environmental Protection Agency's recommended action level may reduce lung cancer deaths by about 2%-4% (Lubin et al., 1995).

Research done by using meta-analysis of 28 case-control studies, which included 13,748 lung cancer cases and 23,112 controls, suggests that residential radon is a risk factor in all histological types of lung cancer. With increasing residential radon quantities per 100 Bq/m^3, the risk of lung cancer, small-cell lung carcinoma, and adenocarcinoma improved by 11%, 19%, and 13%, respectively (C. Li et al., 2020).

2.6 Interactive Effect of Lung Cancer and Risk Factors.

Several studies have been conducted to explore the geographical distribution and the impact of risk factors and lung cancer mortality. Research shows that lung cancer mortality is dependent on numerous factors, such as smoking, radon, education, unemployment, and income. Apart from the independent effects of multiple risk factors on lung cancer mortality, complex interactive products might exist between different risk factors. Earlier research analyzed the independent influence of a single or a set of contextual factors on lung cancer mortality rate; however, the study on the interactive effects of two or more risk factors is very lacking.

A study conducted in Shanghai, China, from 2009 to 2013 identified certain built environmental factors correlated with lung cancer distribution patterns, including the percentage of industrial land (which explains 28% of the cases), location factors (11%), and the portions of cultivated land and green space (6% and 5%, respectively) (L. Wang, Sun, Zhou, Zhang, & Bao, 2019).

Other research studies have shown that smoking status, family history, and other factors (e.g., indoor air pollution, radon exposure) have significant interactive effects on lung cancer (He et al., 2013; Ridge, McErlean, & Ginsberg, 2013).

There is sufficient evidence to suggest that radon exposure increases lung cancer risk significantly higher for smokers than nonsmokers. In addition, there is an interactive effect of radon and smoking on lung cancer mortality (Lantz, Mendez, & Philbert, 2013).

Research was conducted in China to analyze the relationships between lung cancer incidence of males and females from 207 counties in 2013 with annual concentrations of PM2.5, PM10, SO_2, NO_2, CO, and O_3 were analyzed. GeoDetector q statistic was used for assessing the non-linear spatial association between outdoor air pollution and the incidence of lung cancer. This study found a spatial association between outdoor air pollution and lung cancer incidence. In north China, the contact between SO_2 and PM2.5 is the predominant interaction. The dominant collaborative factors are between SO2 and O3 in males and between SO2 and CO in females in the south. Also, they found that smoking is a substantial contributor to lung cancer among men, either in South or North China, and the interaction between smoking and air pollutants increases this risk (Xing et al., 2019).

2.7 Disparities in Lung Cancer

Although the term disparities are often interpreted to mean racial or ethnic disparities, many dimensions of disparity occur in the United States, especially in health. If a health outcome is seen to a larger or lesser extent between populations, there is disparity. Race, age, ethnicity, gender, disability, socioeconomic status, and geographic location contribute to achieving good health. Therefore, it is essential to recognize the impact that social determinants have on the health outcomes of particular people. Healthy People tries to improve the health of all groups (Healthy people.gov, 2022)

During the past two decades, one of Healthy People's overarching goals has concentrated on disparities. In Healthy People 2000, it was to reduce health disparities among Americans. In Healthy People 2010, it was to eliminate, not just reduce, health disparities. In Healthy People 2020, that goal was expanded even further: to achieve health equity, eliminate disparities, and improve the health of all groups (Paula Braveman, 2014).

Compared with all other racial and ethnic groups in the United States, African Americans are disproportionally impacted by lung cancer, both in terms of incidence and survival (Siegel, Miller, & Jemal, 2017). These differences were first formally noted in 1972 (Burbank & Fraumeni, 1972) and have been continuously observed by Surveillance Epidemiology and End Results (SEER). For example, the age-adjusted lung cancer incidence rate is ~32% higher in African Americans than European Americans, with disparities most predominant among men. In addition, on average, African Americans are diagnosed with lung cancer three years earlier than European Americans (Robbins et al., 2015).

About educational disparities, studies reveal that there are transparent gradients in both sexes. For example, men and women with a high academic level have a 26 % and 33%

lower incidence risk than individuals with a low educational level (Tetzlaff, Epping, Tetzlaff, Golpon, & Geyer, 2021).

Also, there is concern that disparities in the implementation of and access to lung cancer screening will further widen existing gaps in lung cancer care and mortality among racial and ethnic minorities, individuals of low socioeconomic level, and uninsured or underinsured people. Studies reveal that healthcare disparity in lung cancer screening occurs when two people at equal lung cancer risk and who have an equal harm-to-benefit ratio from lung cancer screening are not managed equitably. Therefore, it is critical to address disparities in eligibility, referral, healthcare access, and appropriate follow-up for lung cancer screening and propose strategies by which they may be minimized (Rivera et al., 2020).

2.8 Summary

This literature review has traced the methodological history of relationship with social environment, physical environment, and health behavior on health. While different types of social and economic risk factors, environmental factors, and health behaviors have been investigated. Various socio-economic and lung cancer impacts constantly persist in public health and social concern. Although previous research studies have made significant steps towards understanding the result and relationship of the imbalances in the geographic distribution of lung cancer mortality, they have been limited methodologically in different ways.

First, previous research was based on models of incidence or death rates within one time, believing that rates increase or decrease with time at a steady pace. Also, the interest of male and female lung cancer trends had inadequate attention in previous research.

Second, lung cancer mortality and its relationship with socioeconomic variables have been studied by different researchers. But there has been a lack of cumulative source assessment in other geographic areas with risk factors that had limited attention. Additionally, identifying regions of the state where the contribution of the lung cancer risk factors could produce measurable effects on lung cancer mortality. Primarily few research projects focus on the Appalachian region and non-Appalachian region.

Third, there has been no research on the combined effects of various risk factors on lung cancer; complex collaborative outcomes might exist between different risk factors. Previous studies analyzed the independent influence of a single or a set of contextual factors on lung cancer incidence or mortality but not a mutual relationship.

Forth, most previous research has focused on disparities between groups (e.g., Black/White, poor/rich) and used rate ratios to quantify the difference. However, current research lacks disparities across multiple subgroups and how those may change over time and by gender. These four gaps in public health methodology present an opportunity for further research and improvement.

The following chapter outlines the data sources and methodology used in a case study that evaluates Kentucky's geographic distribution of estimated health risks of lung cancer mortality.

Chapter 3

Study Area and Methodology

This section describes the study area, data sources used, and the methodology followed to assess the relationship of various known risk factors of lung cancer mortality in Kentucky. First, the study area is introduced, and the source of the data and the process used to derive the key variables are outlined. Next, the variables used in the case study are defined and described, along with their data sources. Finally, the methods that were chosen to address the research problem are explained.

3.2 Study Area

Lung and bronchus are the primary cause of cancer-related deaths in both the United States and in Kentucky. Furthermore, Kentucky has the highest age-adjusted rate of cancer deaths for lung and bronchus cancers compared to all states. Kentucky represents the highest lung cancer incidence rate at 91.4 per 100,000 compared to the national ratio of 58.3 per 100,000 (Cardarelli, Madabhushi, Bledsoe, & Weaver, 2019). The geographic scope of this analysis includes the entire state of Kentucky. The area of Kentucky designated as Appalachia is set by the Appalachian Regional Commission (ARC) and includes 54 counties in the state's Eastern region (Figure 3-1) (Appalachian Regional Commission, 2021).

Generally, Appalachia carries a higher cancer burden compared with non-Appalachia, particularly for tobacco-related cancers. Appalachia has higher rates for all cancer sites combined regardless of gender, race, or region (R. J. Wilson et al., 2016).

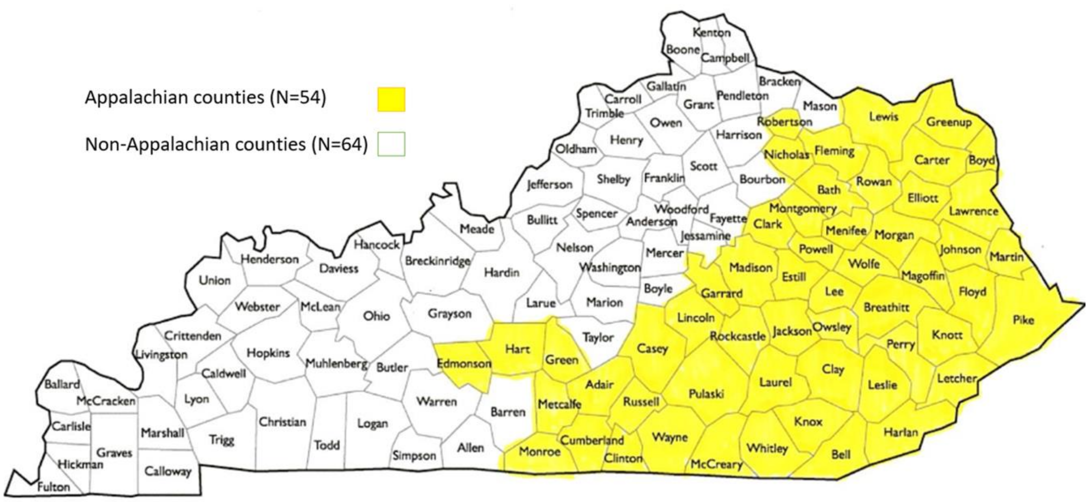


Figure 3-1: A breakdown of Kentucky counties. Source - Appalachian Regional Commission (https://www.arc.gov/).

According to the Appalachian Regional Commission, most of Kentucky’s Appalachian counties are under considerable economic distress which correlates with overall poor health (Hosseinpoor et al., 2012). In addition, the Appalachian region has been identified as a medically underserved region due to the economic, geographic, and health system challenges in the region (Denham et al., 2004).

Lung cancer figures vary widely across the 120 counties in Kentucky; however, counties in the southeastern portion of the state generally have higher lung cancer incidence and mortality rates. Most counties are part of Central Appalachia, a subregion of Appalachia noted for its high poverty and low educational attainment (W. Jay Christian et al., 2011).

Specifically, researchers have focused on lifestyle choices, environmental factors, and public policy to examine various reasons why incidence and mortality rates are historically more significant in Appalachian Kentucky as opposed to the rest of the state. This study will examine three determinants of health: social environment, physical environment, and health behaviors. By understanding the reasons that lead to high incidence and mortality rates of lung and bronchus cancers in this area of Kentucky, efforts could be made through public policy to reduce these rates, which would be vital for the increased health of Kentuckians.

3.2 Variables and Data Sources

3.2.1 Data of Explanatory Variables

According to the literature, lung cancer mortality is determined by diverse and complex factors. Research has discovered several risk factors that may increase the chances of getting lung cancer. Such risk factors belong to the social environment, physical environment, and health behavior factors.

Social and Economic Environment characteristics are important factors that could increase getting lung cancer. Therefore, this study considered the influence of high school graduation rate, median household income, unemployment rate, poverty, and coal mine employment variables.

Radon level and PM 2.5 levels are used as the physical environment risk factors because evidence suggests that increased exposure to particulate matter PM 2.5 and radon was proven to increase the lung cancer death rate (Cao et al., 2018; C. Li et al., 2020).

Health behaviors could affect the increase in lung cancer mortality. For example, previous studies revealed that the risk of getting lung cancer is highest for people who smoke (Islami, Torre, & Jemal, 2015). Also, insurance status and being physically unhealthy increase lung cancer mortality (Casebeer et al., 2019; Mohamed et al., 2020). So, the study obtained the adult smoking rate, uninsured rate, and physically unhealthy days as health behavior factors.

Overall, the study collected a variety of 10 types of lung cancer risk factors, including social and economic environment, physical environment, and health behavior factors that may affect lung cancer mortality rates.

3.2.2 Dependent Variables

County-level lung cancer mortality rate was used as the dependent variable. Lung cancer mortality data was gathered from 2002 to 2019 using the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) database. The SEER*Stat statistical software provides an accessible, intuitive mechanism for analyzing SEER and other cancer-related databases. SEER data do not include individual-level measures of socioeconomic position therefore, data linked each case's county to attributes from 2012 – 2016. The calculation of each socio-economic variable for each county was based on county of residence in the 2000 US Census. All counties in the SEER database (n = 120) were ranked according to the 2000 US Census. The following table (Table 3.1) describes the Dependent variables used in this study, together with descriptions and data sources.

Table 3.1: Dependent variables and data sources

Data type	Year	Description	Source
Age-adjusted Lung cancer mortality rates	2002 - 2019	The weighted average of the age-specific lung cancer mortality rates	SEER*Stat Database Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Mortality - All COD, Aggregated with County, Total U.S. (1969-2016) <Katrina/Rita Population Adjustment> - Linked to County Attributes - Total U.S., 1969-2017 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released December 2018. Underlying mortality data provided by NCHS (www.cdc.gov/nchs).

3.2.2 Explanatory Variables

The explanatory variables are obtained from the county level. All explanatory variables used in this study are described in Table 3.2, along with their variable source.

Table 3.2: Explanatory variables and data sources.

Data type	Year	Description	Source
High school graduation rate	2016	Percentage of the ninth-grade cohort that graduates in four years	County health ranking- 2019 https://www.countyhealthrankings.org/explore-health-rankings
Median household income	2017	Income level earned by a given household where half of the homes in the area earn more and half earn less	County health ranking – 2019 https://www.countyhealthrankings.org/explore-health-rankings
Poverty Rate	2016	The percent of persons below the poverty level	Census 2012-2016 ACS table https://www.census.gov/acs/www/data/data-tables-and-tools/data-profiles/2016/
Unemployment	2017	Number of people ages 16+ unemployed and looking for work	Bureau of Labor Statistics https://www.bls.gov/
Coal Mine Employment	2009-2015	Number of people employed in the coal mining industry	Kentucky Energy and Environment Cabinet Department for Energy Development and Independence. https://eec.ky.gov/Pages/index.aspx
P.M 2.5	2014	The average daily amount of fine particulate matter in micrograms per cubic meter	U.S. Environmental Protection Agency https://www.epa.gov/outdoor-air-quality-data/download-daily-data

Table continued

Data type	Year	Description	Source
Radon	2016	Counties with predicted average indoor radon screening levels	U.S. Environmental Protection Agency https://www.epa.gov/sites/default/files/2014-08/documents/kentucky.pdf
Adult smoking rate	2016	Percentage of adults that reported currently smoking	Behavioral Risk Factor Surveillance System from county health ranking (county health ranking) https://www.countyhealthrankings.org/
Physically unhealthy days	2016	The average number of reported physically unhealthy days per month	National Center for Chronic Disease Prevention and Health Promotion, Division of Population Health. https://www.americashealthrankings.org/explore/annual/measure/Overall_a/state/ALL
Uninsured rate	2016	Percentage of population under age 65 without health insurance	County health ranking – 2019 https://www.countyhealthrankings.org/explore-health-rankings

3.2.3 Data Used to Measure Disparity

The fourth phase of this study focuses on measuring Kentucky lung cancer disparities. When calculating disparity in lung cancer mortality, an index of socioeconomic variables was analyzed using five aspects of the social and economic environment (e.g., education, unemployment, income, poverty, geographic region).

According to the requirement of health disparity measures, data should be classified into quantiles. Therefore, health disparity measures for all counties in the State of Kentucky were classified into four quantiles of an equal number of counties of socioeconomic variables based on each variable's minimum and maximum value.

All counties in the SEER database (n=120) were ranked according to the ratio of the population ages 25 and over with at least a high school degree, assessed from the Census 2012-2016 American Community Survey (ACS 2012-2016). Male females with less than high school graduation attainment ranged from 7.22 % in Oldham County to 36.3% in Clay County. To determine quintile cut points for a user-defined variable based on the percentage of the county population that had less than high school education, case listing results from county attributes data were used. Based on that percentage of less than high school metrics divided into four quantiles, such as, first quantile (7.22% - 14.20%), second quantile (14.21% - 19.67%), third quantile (19.68% - 24.03%) and fourth quantile (24.03% - 36.33%).

Four quantiles were determined for median household income (in ten thousand) as first quantile ($01897 – $03298), second quantile ($03299 – $03975), third quantile ($03976 – $04503), and fourth quantile ($04504 -$08632).

The percent of people ages 16 and over who are unemployed is calculated using the Census 2012-2016 ACS data (ACS 2012-2016). According to the percentage of the county population, four unemployed quantiles were defined as, first quantile (4.10% - 6.74%) second quantile (6.75% - 7.90%) third quantile (7.91% - 10.17%) and fourth quantile (10.18% - 18.60%).

The percent of persons whose incomes are below the poverty level are calculated using tables from the Census 2012-2016 ACS data (ACS 2012-2016). Percentage of persons below poverty was separated into four quantiles such as, first quantile (5.98 % - 16.98%) second quantile (16.99% - 20.39%) third quantile (20.40% - 25.37%) and fourth quantile (25.38% - 42.50%).

Also, this study obtained male and female mortality estimates for two geographic regions, such as Appalachian and Not Appalachian regions. Fifty-four counties belong to the Appalachian region in Kentucky, and sixty-six counties belong to the non-Appalachian area (Appalachian Regional Commission, 2021).

3.3 Methods

This dissertation implements four stages of statistical analysis to assess the relationship of lung cancer mortality with explanatory variables in Kentucky. First, the Joinpoint regression method was used to analyze lung cancer trends. Then, linear association between the dependent and each explanatory variable was measured using the ordinary least square (OLS) method. Also, the Geodetector method was used to investigate the Spatial Stratified Heterogeneity (SSH) of lung cancer mortality and suspected risk factors. Finally, Health Disparity Calculator was used to measure the disparity in lung cancer mortality.

3.3.1 Joinpoint Regression Method

This study used Joinpoint regression analysis to identify points where a statistically significant change across time in the linear slope of the trend occurred (Kim, Fay, Feuer, & Midthune, 2000). In Joinpoint analysis, best-fit points where the rate changes significantly (increase or decrease) were chosen. The study starts with the minimum number of join points and tests whether one or more join points are statistically significant and should be added to the model (up to four join points). In the final model, each join point indicates a statistically significant change in trend. An annual percentage change (APC) is computed for each movement using generalized linear models assuming a Poisson distribution. Notable changes include direction or the rate of increase or decrease. Joinpoint analyses were performed using the 'Joinpoint' software (Version 4.5.0.1) from the US National Cancer Institute (National Cancer Insitute, 2022).

For this study, the Joinpoint regression assessment involves measuring a sequence of joined straight lines on a log scale to the tendencies in the yearly age-adjusted lung cancer incidence and mortality rates. Line portions are joined at points called joinpoints. Every joinpoint denotes a statistically significant (P = .05) change in trend (National Cancer Insitute, 2022). The number of join points is found using a permutation test via Monte Carlo resampling.

The percent change (PC) in rates over a particular time period is calculated by taking the difference between the initial rate and the end rate. The rates can either be a single year rate or a two-year average. The difference is then divided by the initial rate and multiplied by 100 to convert it to a percent. When n = number of years, r = rates, $y = Ln(r)$, x = calendar year, y = $mx + b$.

$$Initial\ Rate = Rate_x \text{ or } \left(\frac{Rate_x + Rate_{x+1}}{2} \right) \qquad PC_{x-y} = \left(\frac{\left(End\ Rate - Initial\ Rate \right)}{Initial\ Rate} \right) \times 100$$

$$End\ Rate = Rate_y \text{ or } \left(\frac{Rate_y + Rate_{y+1}}{2} \right)$$

To identify the year(s) when a trend change is created, it calculates the annual percentage change (APC) in rates between trend-change points, and it also assesses the average annual percentage change (AAPC) in the whole period studied. To calculate the APC, the following model is used:

$$\boldsymbol{APC = 100 \times (e^m - 1)}$$

In Equation, e^m is the slope coefficient of each section. For example, if the APC is 1%, and the rate is 50 per 100,000 in 1990, the rate is 50 x 1.01 = 50.5 in 1991 and 50.5 x 1.01 = 51.005 in 1992. Rates that adjust at a continuous percentage every year change linearly on a log scale.

The AAPC over any fixed interval is calculated using a weighted average of the slope coefficients of the underlying Joinpoint regression line with the weights equal to the length of each segment over the interval. The final step of the calculation transforms the weighted average of slope coefficients to an annual percent change. When b_i denotes as the slope coefficients for each segment in the desired range of years, and the w_i as the length of each segment in the range of years, then:

$$AAPC = \left\{ exp\left(\frac{\sum w_i b_i}{\sum w_i}\right) - 1 \right\} \times \ 100$$

For example, in 50- to 54-year-old men, join point regression recognizes two join points in 2005 and 2009, so the entire period is segmented in three periods: 1999–2005, 2005–2009, and 2009–2015, with APC equivalent to – 0.014, − 0.032, and – 0.012, respectively, and segment sizes comparable to 6, 4 and 6 years, respectively. So then, AAPC is estimated as:

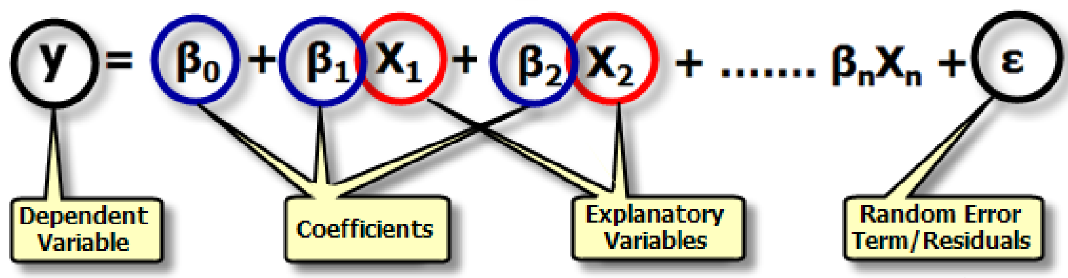
$$AAPC = \left(e^{\frac{-6\times0.014-4\times0.032-6\times0.012}{6+4+6}} - 1 \right) \times 100 = -1.8\%$$

3.3.2 Ordinary Least Squares Regression

Regression analysis is a statistical method used for assessing the relationships among variables. It contains many techniques for modeling and analyzing several variables when focusing on the correlation between dependent and independent variables. Regression analysis helps to understand how dependent variable changes independent variables is varied. In contrast, the other independent variables are held fixed. This study utilized the ordinary least square regression model to study the relationship between lung cancer mortality from socio-economic risk factors.

The dependent variable is considered "y" in a regression equation and is always " x " for independent or explanatory variables. The Independent variable is associated with a regression coefficient describing the strength and the sign of that variable's relationship to the dependent variable.

The relationship was examined at the county-level using an Ordinary Least Squares (OLS) regression. The model is:



$$CR_I = \beta_O + \quad \beta_3 income + \ + \beta_6 education + \ \beta_7 smoking$$

$$+ \ \beta_6 coalmine\ employement$$

- **Dependent variable** (*y*): This represents variable the study is trying to predict or understand (e.g., lung cancer mortality rate).

- **Independent/Explanatory variables** (*x*) are the variables used to model or predict the dependent variable values in the regression equation, often referred to as *explanatory* variables. Also, the dependent variable is a function of the explanatory variables. In this study, independent variables include four sociodemographic factors: adult smoking, household median income, high school education, and coal mine employment.

- **Regression coefficients** (β): coefficients are calculated by the regression tool. They are values, one for each independent variable, representing the explanatory variable's strength and type of relationship to the dependent variable. When the connection is positive, the sign for the associated coefficient is also positive.

Coefficients for negative interactions have negative indications. When the relationship is strong, the coefficient is significant. Weak interactions are linked with coefficients close to zero.

- β_0 is the regression *intercept*. It represents the expected value for the dependent variable if all the independent variables are zero.

However, it should be stated that in public health studies for large areas, i.e., the entire US, regression models could be spatially non-stationary, meaning that the coefficients of the regression model are spatially variable (A. S. Fotheringham, Charlton, & Brunsdon, 1998). In this case, local regression models such as the Geographically Weighted Regression (A. Fotheringham, Brunsdon, & Charlton, 2002) were used to avoid the 'ecological fallacy' problem and explain the variability of cancer mortality (Holt, Steel, Tranmer, & Wrigley, 2010).

3.3.2.1 Moran's I

Moran's I is used to measure the overall spatial autocorrelation of the data set. In other words, it measures how one object is similar to others bordering it. If objects are attracted by each other, it implies that the observations are not independent. This violates a fundamental assumption of statistics, the independence of data. In other words, the presence of autocorrelation renders most statistical tests invalid, so it is essential to test for it. Moran's I is used as a one way to test for autocorrelation. The equation is,

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ji}} \times \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Where n is the total number of counties in the area, i and j represent different counties, xi is the residual of i, and $\bar{X}$ is the mean of residuals. Wij is a measure of spatial proximity pairs of i and j (Wong & Lee, 2005). The values of Moran's I would be between −1 and +1. Negative autocorrelation values mean nearby locations tend to have dissimilar values; positive autocorrelation values mean that similar values are likely to occur in adjacent areas. Along with the index, Z-scores are usually reported for the statistical significance test. If Z is out of ±1.96, the null hypothesis of the randomness test is rejected at the 95% confidence level, which means the pattern is spatially auto correlated. Otherwise, the spatial arrangement would be completely random (Lin & Wen, 2011b). If values are:

- -1, it shows a perfect clustering of dissimilar values (perfect dispersion).
- 0, it shows that there is no autocorrelation (perfect randomness.)
- +1, it shows a perfect clustering of similar values (it's the opposite of dispersion).

Figure 3-2: Interpretation of spatial autocorrelation. Source: (GIS/Data Center Guides, 2021)

3.3.2.2 Geographically Weighted Regression (GWR)

Geographically Weighted Regression (GWR) is one of the various spatial regression techniques increasingly utilized in geography and other disciplines. GWR offers a local model of the variable or procedure trying to understand or forecast by fitting a regression equation to every feature in the dataset. This method is robust and produces reliable statistics for examining/estimating linear relationships (Esri; Matthews & Yang, 2012).

GWR was initially developed to analyze spatial point data and allows for the interpolation of values not included in the data set. It is applied under the assumption that the strength and direction of the relationship between a dependent variable and its predictors may be modified by contextual factors (Columbia Public Health, 2019). GWR has high utility in epidemiology, particularly for infectious disease research and evaluations of health policies or programs. Limitations of GWR include problems of multicollinearity and the approaches to calculating goodness of fit statistics (Columbia Public Health, 2019)

GWR is considered a localized regression model that allows the parameters of a regression estimation to vary over the spatial domain (Lin & Wen, 2011a). Therefore, the model can be expressed as:

$$LCR_I = \beta_{Oi} + \beta_{1i}\, adult\ smoking + \beta_{2i}\, High\ school\ graduation\ rate \\ + \beta_{3i}\, M\ houshold\ income + \beta_{4i}\, coal\ mine\ employement$$

Where βni is the estimated regression coefficient at county i. The spatial variability of an estimated local regression coefficient was examined to determine whether the underlying process exhibited spatial heterogeneity (A. Fotheringham, Brunsdon, & Charlton, 2000). The optimal solution of the regression equation in GWR is constrained by a geographically weighted matrix Wi (A. Fotheringham et al., 2000).

$$\beta_i = \left(X^T W_i X\right)^{-1} X W_i Y$$

where Wi is defined by the spatial relationships between neighboring points:

$$W_i = \begin{pmatrix} w_{i1} & 0 & 0 & \cdots & 0 \\ 0 & w_{i2} & 0 & \ldots & 0 \\ 0 & 0 & w_{i3} & \ldots & 0 \\ . & . & . & \ldots & . \\ 0 & 0 & 0 & \ldots & w_{in} \end{pmatrix}$$

Where wij is the strength of association between location i and location j and (i and j =1…n) are defined by their distance and a kernel function. The kernel function is a distance decay function usually defined as Gaussian with a user specified band width or spatially adaptive band widths.

This research will use the calibration method that minimizes Akaike Information Criterion (AIC) of regression models to obtain the spatially adaptive band width values. The analyses will use ESRI's ArcGIS 10.1 and GWR4 software packages.

These values are mapped using GIS software, thus providing a way to visually interpret the geographic distribution of the nature and strength of the relationships between explanatory and dependent variables.

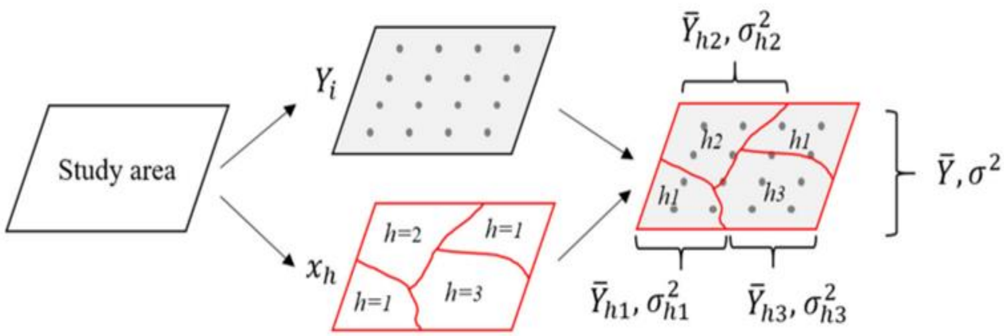
3.3.3 Geographical Detector

The Geographical Detector technique is designed to estimate the associations between a geographical phenomenon and relevant risk factors. It is based on spatial variance analysis. The underlying principle is to evaluate the consistencies between the spatial distribution patterns of the studied geographical event (e.g., lung cancer mortality rate) and potential risk factors (e.g., adult smoking rates). The study assumes that a geographical event would show a similar spatial distribution pattern with a risk factor if the risk factor's element significantly influences the geographical event's occurrence (J. F. Wang et al., 2010).

Specifically, the Geographical Detector method first requires division of the spatial distributions of the geographic issue and risk factors into subregions corresponding to their stratified spatial heterogeneity. Spatial heterogeneity is a significant feature of the geographic phenomenon, and it indicates the irregular distributions of events across a region or, basically, the spatial variation of attributes (Anselin, 2010). A stratification of heterogeneity is, fundamentally, a segmentation of a research region, where observations are homogeneous within each stratum but not between strata (J.-F. Wang, Zhang, & Fu, 2016). The spatial heterogeneity between areas (each area contains one or more units) is commonly called stratified spatial heterogeneity, that is a common phenomenon, such as climate or ecological zones, spatial variability of soil types, and land-use patterns (J.-F. Wang et al., 2016). If the attributes within the strata are consistent or the variances within the strata are zeros, the stratified heterogeneity is primarily significant; on the contrary, the stratification of heterogeneity will disappear if there is no difference between the strata. Commonly, a stratification of heterogeneity separates a target population by minimizing the within-strata variance and increasing the between-strata variance of an attribute. Technically, the stratification of heterogeneity can be achieved by either prior experience or classification methods (J. F. Wang et al., 2010). Therefore, the uniformity between the spatial stratified heterogeneities of a pair of geographic events suggests the possibility that there is a statistical correlation between these phenomena (J.-F. Wang et al., 2016).

The Geographical Detector model aims to assess whether stratified spatial heterogeneity exists for a geographic phenomenon and investigate the interpretation of a geographic phenomenon by comparing the spatial correlation of its strata against the strata of believed determinants. Geographical Detector model used four detectors (i.e., a factor detector, risk detector, ecological detector, and interactive detector) to determine the main and interactive effects of possible factors on the examined geographic event. First, a factor detector is used to determine which factors are responsible for the incidence of the studied geographic event. Second, a risk detector is used to recognize the geographical area with a high probability of the event's occurrence. Third, an ecological detector is used to evaluate whether the influences of various factors on the studied geographic event change remarkably from each other. Finally, an interactive detector is applied to determine whether multiple factors independently or dependently affect the occurrence of the studied event (J. F. Wang et al., 2010).

The Geographical Detector technique is unique as it extracts the underlying mutual correlations between a studied geographic event and suspected causes, without any restrictions on the response and explanatory variables (J. F. Wang et al., 2010). Additionally, the Geographical Detector method is appropriate for quantitative and qualitative data, while traditional regression models may experience problems when the nominal data has too many categories (Yue & Hu, 2021).



.

Figure 3-3: The principle of the Geographical detector

This study assumes that a potential risk factor X stratifies the study area into subregions ($x_1 x_2 x_3$) in the geographical space (Figure 3-3). The risk factor layer overlaps the spatial distribution of the lung cancer mortality rate.

The averages and variances of mortality rates in each subregion and the whole study area are, respectively, represented as $y_{h1,}\, y_{h2}, y_{h3}, y$ and $\sigma_{h1^2}\ \sigma_{h2^2}\ , \sigma_{h3^2},\ \sigma^2$. If the lung cancer mortality rate is wholly determined by factor D, the rates will be identical everywhere in each of the subregions ($x_1 x_2 x_3$) and σ_{h1^2} σ_{h2^2} , σ_{h3^2} will be zeros. On the contrary, if the lung cancer mortality rate is entirely independent of X, the accumulated variance within the subregions will be stable with the whole study area's pooled variance. This method is measured by the Power of Determinant (PD) (J.-F. Wang et al., 2016).

$$PD = 1 - \frac{\sum_{h=1}^{L} \sum_{i=1}^{N_h} \left(Y_{hi} - Y_h\right)^2}{\sum_{i=1}^{N} \left(Y_i - Y\right)^2} = 1 - \frac{\sum_{h=1}^{L} N_h {\sigma_h}^2}{N\sigma^2}$$

where a study area contains of N units and is stratified into L stratums by a factor, signified as h = 1, 2, . . . , L, respectively; stratum h is comprised of $N_h$units; $Y_i$and Y_{hi} respectively, represent the value of unit *i* and stratum h; Y_h= $(1/N_h)$ $\sum_{i=1}^{N_h} Y_{hi}$ represents the stratum mean; σ_{h^2} = $(1/N_h)$ $\sum_{i}^{N_h}$ $(Y_{hi}$ - $Y_h)^2$ represents the stratum variance; Y_h= (1/N) $\sum_{i=1}^{N} Y_i$ represents the population mean; and σ^2 = (1/N) $\sum_{i}^{N}$ (Y_i -Y)^2 represents the population variance. (PD) quantifies how a risk factor explains lung cancer mortality, and its value lies between 0 and 1. The more significant the amount of PD, the greater the impact of the factor. If a factor completely controls the lung cancer mortality rate, PD = 1; if it is entirely unrelated to the lung cancer mortality rate, PD = 0. The Geographical Detector method is built on the PD, which generates the following four detectors: risk detector, factor detector, ecological detector and interactive detector (J. F. Wang et al., 2010).

3.3.3.1 Risk Detector

The risk detector aims to clarify whether the mean mortality rates in each subregion are statistically different from each other when a possible risk factor X stratifies the study area. This is achieved by the t-value test (J. F. Wang et al., 2010).

$$t_{Y_{h1}-Y_{h2}} = \frac{Y_{h1} - Y_{h2}}{\sqrt{\frac{\sigma_{h1}^2}{N_{h1}} + \frac{\sigma_{h2}^2}{N_{h2}}}}$$

where $y_{h1,}\sigma_{h1^2}$ and N_{hi} denotes the mean mortality rate, the variance of fatality rate, and sample size in subregion h_i, respectively. The t-value follows nearly a student's t distribution, with the degree of freedom (df) as:

$$df = \frac{\frac{\sigma_{h1}^2}{N_{h1}} + \frac{\sigma_{h2}^2}{N_{h2}}}{\frac{1}{N_{h1}-1}\left[\frac{\sigma_{h1}^2}{N_{h1}}\right]^2 + \frac{1}{N_{h2}-1}\left[\frac{\sigma_{h2}^2}{N_{h2}}\right]^2}$$

If the null hypothesis H_0 : Y_{hi} = Y_{h2} is denied at the confidence level α (usually 5%), there is a considerable difference between mortality rates of two subregions.

3.3.3.2 Factor Detector

The factor detector calculates to which extent a factor explains the dependent variable's spatial variance, estimated by the PD, shown as (J. F. Wang et al., 2010).

$$PD = 1 - \frac{\sum_{h=1}^{L} \sum_{i=1}^{N_h} \left(Y_{hi} - Y_h\right)^2}{\sum_{i=1}^{N} \left(Y_i - Y\right)^2} = 1 - \frac{\sum_{h=1}^{L} N_h {\sigma_h}^2}{N\sigma^2}$$

3.3.3.3 Ecological Detector

The ecological detector is used to investigate the impacts of two factors X_1, X_2. The dependent variable has a significant difference, determined by the F statistics (J. F. Wang et al., 2010).

$$F = \frac{N_{X1}(N_{X1} - 1)SSW_{X1}}{N_{X2}(N_{X2} - 1)SSW_{X2}}$$

$$SSW_{X1} = \sum_{h=1}^{L1} N_h {\sigma_h}^2, SSW_{X2} = \sum_{h=1}^{L2} N_h {\sigma_h}^2$$

where N_{x1} and N_{X2}, respectively, represent the sample size within the coverage of X_1 and.X_2, SSW_{x1} and SSW_{x2} represent the sum total of the within strata variances $X_1$and X_2 respectively, form the strata. $L_1$and L_2 respectively, represent the strata number of $X_1$and X_2. If the null hypothesis H_0: SSSW_{x1}= $SSW_{x2}$is denied at the confidence level α (usually 5%), the influences of $X_1$and X_2on the dependent variable are statistically significant.

3.3.3.4 Interactive Detector

The interactive detector is used to calculate the interaction effect between the impacts of two factors, e.g., $X_1$and X_2 on the dependent variable (J. F. Wang et al., 2010). This is accomplished by firstly overlapping the geographical layers of $X_1$and $X_2$to form a new layer X_3 and obtaining the attributes of layer X_3 by combining layer $X_1$and X_2 . Then, by comparing the PD of layer X_3 with those of $X_1$and X_2, the interactive detector could verify whether two factors, $X_1$and X_2 , when taken together, have stronger or weaker influences on the lung cancer mortality rate than they do independently. Interaction effects can classify into factors as:

Enhance: if PD ($X_1 \cap X_2$) > PD(X_1) or PD(X_2);

Enhance, bivariate: if PD ($X_1 \cap X_2$) > PD(X_1) and PD(X_2);

Enhance, nonlinear: if PD ($X_1 \cap X_2$) > PD(X_1) + PD(X_2);

Weaken: if PD ($X_1 \cap X_2$) < PD(X_1) + PD(X_2);

Weaken, univariate: if PD($X_1 \cap X_2$) < PD(X_1) or PD(X_2);

Weaken, nonlinear: if PD($X_1 \cap X_2$) < PD(X_1) and PD(X_2);

Independent: if PD($X_1 \cap X_2$) = PD(X_1) + PD(X_2);

3.3.4 Heath Disparity Calculator

This study used the choice of measures of health disparity guided by the national cancer institute (Harper & Lynch, 2004). According to the literature, there are at least three essential issues to consider in choosing a summary measure of disparity. The first is whether the disparity is measured on the absolute scale (i.e., differences in rates) or relative scale (i.e., ratios of rates). The second is whether all individuals in the population are weighted equally (i.e., population-weighted measures of disparity) or weighted inversely to the population size of their social group (i.e., unweighted measures of disparity). The third issue is choosing the reference point from which differences among subgroups are measured (e.g., population average, the best-observed rate, a target rate).

A brief review of several potential summary measures of health disparity that differ across these dimensions is summarized in Table 3.3, separated by whether they measure disparity on the relative or absolute scale. However, computing multiple summary measures is complex, and it seemed unlikely that researchers would perform this series of tasks. To make the process more manageable, 11 summary measures were programmed into a new online tool.

Table 3.3: Summary table of characteristics of potential health disparity measures.

Disparity Measure	Absolute or Relative	Reference Group	All Social Groups	Reflect SES Gradient	Social Group Weighting	Inequality Aversion Parameter	Graphical Analog
Absolute difference	Absolute	Best	No	Yes	No	No	Yes
Relative difference	Relative	Best	No	Yes	No	No	Yes
Slope index of inequality	Absolute	Average	Yes	Yes	Yes	No	Yes
Relative index of inequality	Relative	Average	Yes	Yes	Yes	No	Yes
Index of disparity	Relative	Best	Yes	No	No	No	No
Relative concentration index	Relative	Average	Yes	Yes	Yes	Yes	Yes
Absolute concentration index	Absolute	Average	Yes	Yes	Yes	Yes	Yes
Between group variance	Absolute	Average	Yes	No	Yes	Yes	No
Theil index	Relative	Average	Yes	No	Yes	Yes	No
Mean log deviation	Relative	Average	Yes	No	Yes	Yes	No
Kunst—Mackenbach index	Relative	Best	Yes	Yes	Yes	Yes	Yes

SES = Socioeconomic Status. Source: (Harper & Lynch, 2004)

The Health Disparities Calculator (HD*Calc) is a statistical software created to extend the National Cancer Institute monographs. It can be used with any population-based data. HD*Calc generates multiple indices to evaluate and monitor health disparities (Health Disparities Calculator, September 12, 2019), enabling users to examine trends in disparity among numerous groups. The tool calculates results for 11 existing summary measures for any inequality factor in a data set, such as race, ethnicity, income, education, or geographic region. As shown in Table 3.3, these existing measures vary in their characteristics. HD*Calc has been used to explore cancer control outcomes, including screening incidence, survival, and mortality, but HD*Calc is not limited to use with cancer data (Breen, Scott, Percy-Laurry, Lewis, & Glasgow, 2014).

HD*Calc is a powerful tool for public health research, policy, planning, program development, and evaluation. It is being used for monitoring and planning in public health departments and federal agencies and is available for use in research and public health practice settings free of charge. HD*Calc is also a teaching tool for use in schools of public health, departments of health policy and health services research, and training. HD*Calc will advance understanding and ultimately support the Healthy People goal to eliminate health disparities (Breen et al., 2014).

Here, is a brief outline of the absolute and relative measures used in this study.

3.3.4.1 Rate Difference (RD)

The absolute disparity between two health status values is the simple arithmetic difference. It is calculated as:

$$\text{RD} = r_1 - r_2$$

where r_1 and r_2 are indicators of health status in two social groups. In this case r_2 works as the reference population, and the RD is expressed in the same units as r_1 and r_2. A typical disparity measure that utilizes the absolute difference between two rates for an entire population is the range, in which case r_1 relates to the least healthy group and r_2 the healthiest group. In measuring health disparities, RD is frequently used to evaluate the health of less-advantaged social groups to more-advantaged (Harper & Lynch, 2004).

3.3.4.2 Between-group Variance (BGV)

The variance summarizes all squared deviations from a population average. In grouped data, this is the between-group variance (BGV), and it is calculated by squaring the differences in group rates from the population average and weighting by population size:

$$BGV = \sum_{j=1}^{j} P_j \ (y_1 - \mu)^2$$

where p_j is group j's population size, y_j is group j's average health status, and μ is the average health condition of the population. If there is no disparity, the between-group variance is 0. One way to clarify the between-group variance is the variance that would occur in the population if each individual had the mean health of their social group (i.e., if there were no within-social group variation) (Harper et al., 2008a). The between-group variance may be a helpful indicator of absolute disparity for unordered social groups because it weights by population group size and is responsive to the magnitude of more significant deviations from the population average because it uses a squared deviation (Harper & Lynch, 2004).

3.3.4.3 Absolute Concentration Index (ACI)

The Absolute Concentration Index (ACI) measures how health or illness is concentrated among certain social groups on the absolute scale. It may only be applied to social groups with a natural ordering, such as income or education. It is a quantity for the covariance between social rank and health. It is developed by plotting the cumulative share of the population, ordered by social status, against the cumulative amount of ill health (i.e., the cumulative contribution of each subgroup to the mean level of health in the population). The absolute form of the concentration index is analyzed by multiplying the relative concentration index (RCI) – described below - by the mean rate of the health variable:

$$ACI = \mu RCI$$

where RCI is the Relative Concentration Index defined below, and μ is the mean level of health in the population (Harper & Lynch, 2004).

3.3.4.4 Slope Index of Inequality (SII)

The SII, introduced by Preston, Haines and Pamuk (1981) may be obtained via regression of the mean health variable on the mean relative rank variable (Preston, Haines, & Pamuk, 1981). To calculate relative rank, the social groups are first ordered from lowest to highest.

The population of each social group classification covers a range in the cumulative distribution of the people (Harper & Lynch, 2004). It is given a score based on the midpoint of their range in the cumulative distribution in the population. The regression equation is listed as follows: where j indexes social group, y is the average health status, and Rj is the average relative ranking

$$\overline{y_J} = \beta_0 + \beta_1 \overline{R_J}$$

of social group j in the cumulative distribution of the population, β_0 is the estimated health status of a hypothetical person at the bottom of the social group order (i.e., a person whose relative rank Rj in the social group distribution is zero), and β_1 is the difference in average health status between the hypothetical person at the bottom of the social group distribution and the hypothetical person at the top (i.e. Rj =0 vs. Rj =1). Because the relative rank variable is based on the cumulative magnitudes of the population (from 0 to 1), a "one-unit" change in relative rank is equivalent to moving from the bottom to the top of the social group distribution. Because this regression is run on grouped data (as opposed to individual data), it is estimated via weighted least squares, with the weights equal to the population share pj of group j. The coefficient β_1 is the SII, which is interpreted as the absolute difference in health status between the bottom and top of the social group distribution (Harper & Lynch, 2004).

3.3.4.5 Rate Ratio (RR)

The RR is virtually equal to the RD described above but is calculated by dividing r_1 by r_2 rather than subtracting:

$$RR = r_1 / r_2$$

where again, r_2 is considered as the reference population. While in the background of social group comparisons, the RR is typically based on comparison. For example, RR is the ratio from the least advantaged group (e.g., the lowest socioeconomic group) to the highest group (Harper & Lynch, 2004).

3.3.4.6 Index of Disparity (IDisp)

The Index of Disparity analyzes the difference between several groups and a reference rate and expresses the added differences as a proportion of the reference rate. This measure was introduced by Pearcy and Keppel (2002) and is calculated as:

$$ID_{isp} = \left(\sum_{j=1}^{J-1} \left| r_j - r_{ref} \right| / J \right) / r_{ref} \text{ x } 100,$$

where r_j indicates the measure of health status in the j group, the *ref* is the health condition indicator in the reference population, and J is the number of groups compared. While in principle any reference group may be chosen, the authors suggest the best group rate as the comparison since that denotes the rate desirable for all groups to achieve. In this case, it is not necessary to take the absolute value of the rate differences since they will all be positive (Harper & Lynch, 2004).

3.3.4.7 Relative Concentration Index (RCI)

The Relative Concentration Index (RCI) measures how health or illness is focused among particular social groups. The RCI may only be used with social groups with an inherent ranking, such as income or education. The general formula for the RCI is,

$$RCI = \frac{2}{\mu}\left(\sum_{j=1}^{J} p_j \mu_j R_j\right) - 1$$

where p_j is the group's population share, μ_j is the group's mean health, and R_j is the relative rank of the j the socioeconomic group, which is defined as:

$$R_j = \sum_{j=1}^{J} p_\gamma - \frac{1}{2} P_j$$

where p_y is the cumulative share of the population up to and including group j and $p_j$is the share of the population in group j. R_j indicates the cumulative split of the population up to the midpoint of each group interval, similar to the categorization used for the Slope Index of Inequality above (Harper & Lynch, 2004).

3.3.4.8 Relative Index of Inequality

The SII examined above is a measure of absolute disparity. Nevertheless, dividing this estimated slope by the mean population health gives a relative disparity measure, the Relative Index of Inequality or RII.

$$RII = SII / \mu = \beta_1 / \mu$$

where μ is mean population health and the SII is the estimate of β_1 from the regression that generates the SII. Its explanation is similar to the SII, but it now measures the proportionate (regarding the average population level) rather than absolute increase or decline in health between the highest and lowest socioeconomic group (Harper & Lynch, 2004).

3.3.4.9 Theil Index (T) and Mean Log Deviation (MLD)

The Theil Index and Mean Log Deviation are measures of general disproportionality created by the economist Henri Theil (Theil, 1967). They are both summaries of the difference between the natural logarithm of shares of health and the population. The equation is written as follows:

$$T = \sum_{j=1}^{J} p_j r_j \ln r_j$$
$$MLD = \sum_{j=1}^{J} p_j \left(- \ln r_j \right)$$

where p_j is the proportion of the population in group j and r_j is the ratio of the prevalence or rate of health in group j relative to the total rate, i.e., $r_j = y_j$ / μ where y_j is the prevalence of the outcomes in group j, and μ is the total prevalence. Both measures are population-weighted, are more responsive to health differences further from the average rate, and may be used for both ordered social groups (education) and unordered groups (gender, race) (Harper & Lynch, 2004).

Chapter 4

A Joinpoint Regression Analysis of Trends in Lung Cancer Incidence and Mortality Rates From 2000 – 2016 in Kentucky

4.1 Introduction

Identifying changes in the recent trend is an essential concern in analyzing cancer mortality and incidence data. The recently developed Joinpoint regression model helps identify and explain changes in different periods throughout trends in data (Kim et al., 2000). Therefore, this chapter aims to illustrate male and female recent changes in lung cancer incidence and mortality trends in Kentucky from 2000 to 2016 using Joinpoint regression models.

4.2 Lung cancer incidence trends

This study used the Joinpoint regression model to identify the recent trends in lung cancer incidence and mortality. Therefore, the following figure and tables provide up-to-date information to examine Kentucky's current lung cancer incidence trends during 2000 – 2016.

The data for this analysis was obtained from the SEER database, “Incidence – SEER 18 Regs Research Data, Nov 2018 Sub (2000-2016) Linked to County Attributes - The total U.S., 1969-2017 Counties. All lung cancer incidence in the Kentucky cancer registry in the SEER database obtained data from age recode 00 years to 85+ years (unknown excluded), by male and female, and year of diagnosis between 2000- 2016.

Lung cancer incident results for males and females are summarized in Table 4.1, Figure 4-1. Table 4.1 demonstrates the results of the Joinpoint regression analysis and the annual percentage change (AAPC) for each trend in males and females. Figure 4-1 illustrates the standardized yearly percentage changes (APC) in lung cancer incidence for males and females.

4.2.1 Lung Cancer Incidence Trend: Male and Female

According to the Joinpoint analysis findings, annual average percentage change (AAPC) in Kentucky lung cancer incidence rates for males and females declined by 1.0 % per year for the 16 years of 2000 through 2016 (Table 4.1).

Table 4.1: Lung cancer incidence trend by gender, Kentucky, 2000-2016.

Incidence		Trend 1		Trend 2		Trend 3		Trend 4	
Gender	AAPC 2000-2016	Years	APC	Years	APC	Years	APC	Years	APC
Male & Female	-1.0*	2000 - 2008	-0.1	2008 - 2011	-1.6	2011 - 2014	-0.9	2014-2016	-3.6
Male	-2.0*	2000 - 2008	-1.5*	2008 - 2011	-2.6	2011 - 2014	-1.2	2014-2016	-4.5
Female	0.1	2000 - 2005	1.6*	2005 - 2010	0.6	2010 - 2014	-1.0	2014-2016	-2.4

According to Figure 4-1, the incidence of lung cancer for male and female age-adjusted rates was relatively stable for 2000 through 2008 (APC decreased by 0.1%). But 2008 – 2011 rates decreased by 1.6% per year, and between 2011 – 2014 continued to decline by 0.9%. But during 2014 and 2016, it dramatically reduced by 3.6% per year.

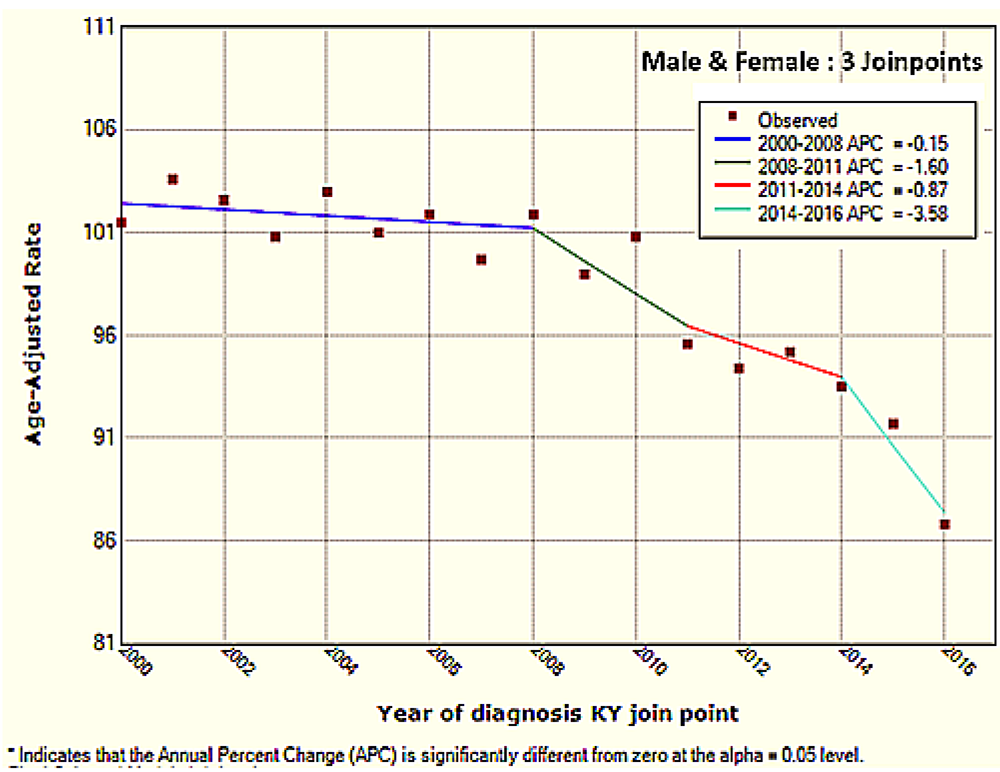


Figure 4-1: Male & female age-adjusted cancer incidence rates 2000-2016.

AAPC- Average annual percentage change, APC- Annual percentage change

4.2.2 Lung Cancer Incidence Trend: Male

According to Table 4.1, male lung cancer incidence rates have decreased by 2.0 % per year for the 16 years of 2000 through 2016.

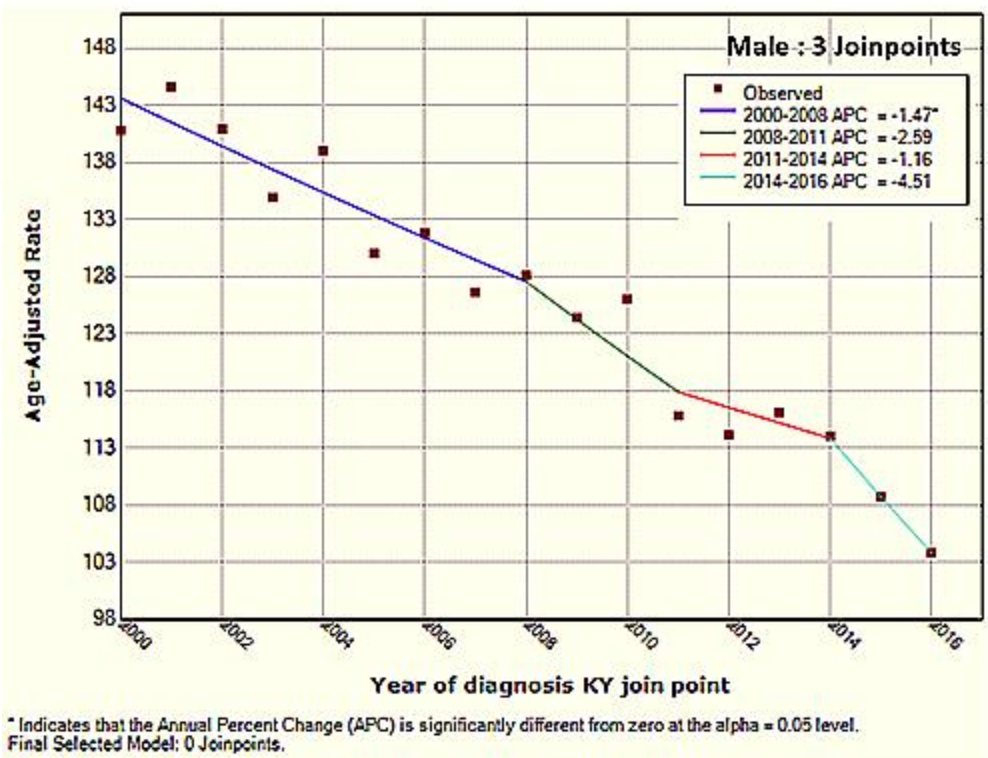


Figure 4-2: Male age-adjusted lung cancer incidence rates 2000-2016.

Figure 4-2 represents the trend for male age-adjusted lung cancer incidence rates. Lung cancer rates for males decreased by 1.5% per year from 2000-2008. Between 2008 – 2011 rates decreased by 2.6% per year. But during 2011 through 2014, it was only a 1.1% decrease per year. But again, rates significantly reduced during 2014 – 2016 by 4.5% per year.

4.2.3 Lung Cancer Incidence Trend: Female

According to Table 4.1, female lung cancer incidence rates have increased by 0.1% per year for the 16 years of 2000 through 2016. But for males, it declined by 2.0% per year.

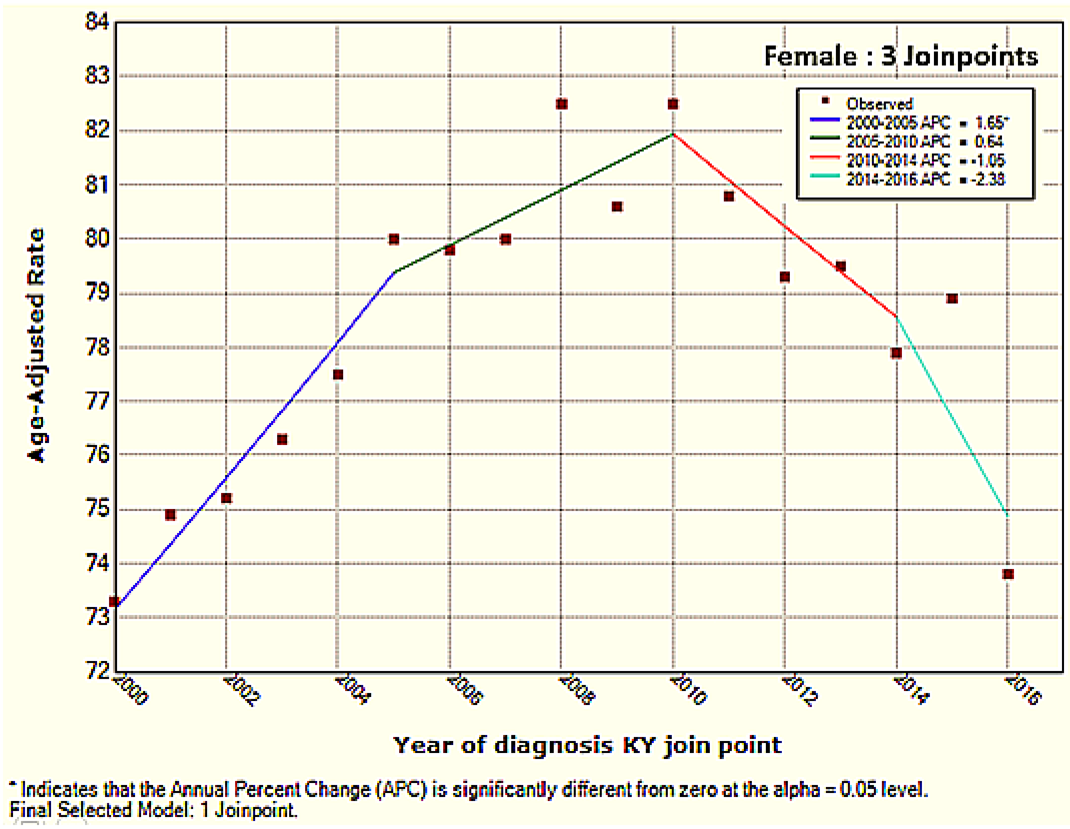


Figure 4-3: Female age-adjusted lung cancer incidence rates 2000-2016.

According to Figure 4-3, female age-adjusted lung cancer incidence rates increased by 1.6% per year for 2000-2005. Between 2005 – 2010 rates continue to increase by 0.6 % per year. But during 2010 through 2014, incidence rates started to decrease by 1% per year. But rates significantly reduced during 2014 – 2016 by 2.4 % per year.

4.3 Lung Cancer Mortality Trends

The following figures and tables provide up-to-date information to examine Kentucky's recent lung cancer mortality trends from 2000 through 2016.

Data on population and lung cancer mortality in Kentucky during 2000 - 2016 were obtained from Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database. Lung cancer mortality data were collected for ages one to 84+ years. For each gender group, age group-specific rates and standardized rates were calculated using 2000 US standard population.

Lung cancer incident results for males and females are summarized in Table 4.2 and Figure 4-2. Table 4.2 demonstrates the results of the Joinpoint regression analysis and the average annual percentage change (AAPC) for women and men. Figure 4-1 illustrates the average percentage changes (APC) in lung cancer incidence for males and females.

4.3.1 Lung Cancer Mortality Trend: Male and Female

According to Table 4.2, Kentucky age-adjusted lung cancer mortality rates for males and females decreased by 1.6 % per year for the 16 years of 2000 through 2016 (Table 4.2).

Table 4.2: Lung cancer mortality trends by gender, Kentucky, 2000-2016.

Mortality		Trend 1		Trend 2		Trend 3		Trend 4	
Gender	AAPC 2000-2016	Years	APC	Years	APC	Years	APC	Years	APC
Male & Female	-1.6*	2000 - 2002	0.1	2002 - 2009	-1.7*	2009 - 2013	-0.8	2013-2016	-3.8*
Male	-2.6*	2000 - 2002	-1.4	2002 - 2009	-2.9*	2009 - 2013	-1.2	2013-2016	-4.4*
Female	-0.6	2000 - 2002	3.0	2002- 2010	-0.2	2010 - 2014	-1.0	2014-2016	-4.4

According to Figure 4-2, male and female lung cancer mortality average percentage change increased (APC) by 0.1% per year for 2000 through 2002. But APC starts to decline from 2002 – 2009 by 1.7% per year. Between 2009 – 2013 continued to decrease by 0.8%. During 2013 and 2016, lung cancer mortality dramatically reduced by 3.8% per year.

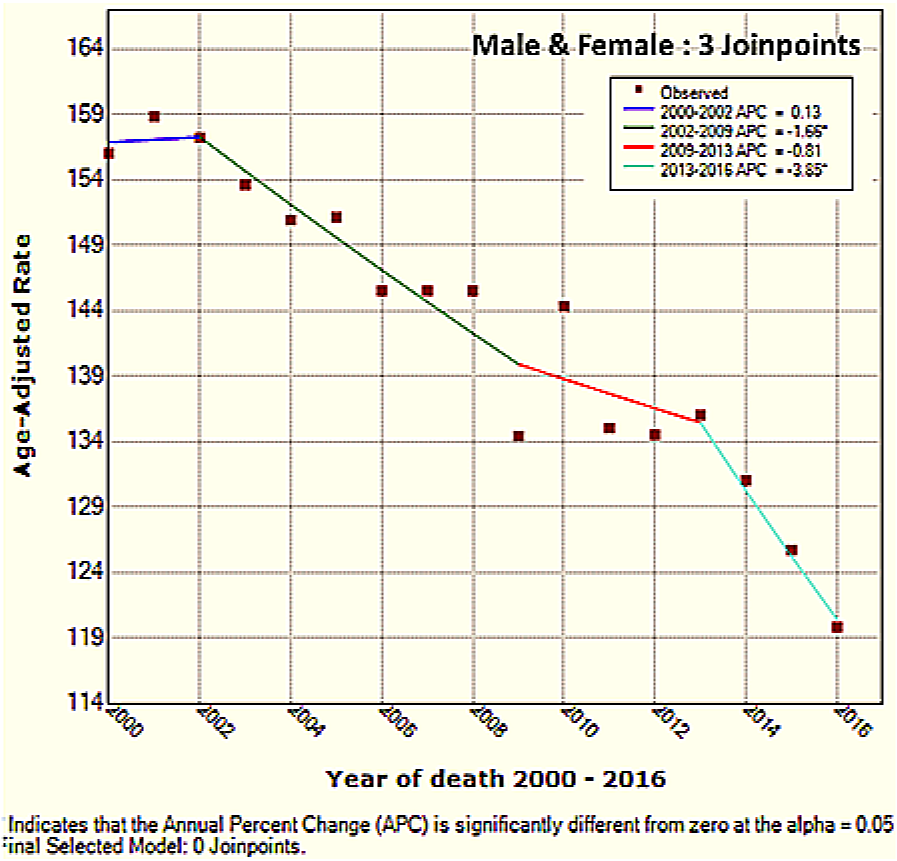


Figure 4-4: Male and female lung cancer mortality rates 2000-2016.

4.3.2 Lung Cancer Mortality Trend: Male

According to Table 4-2, the average annual percentage change (AAPC) for male lung cancer mortality rates has decreased by 2.6 % per year for the 16 years of 2000 through 2016.

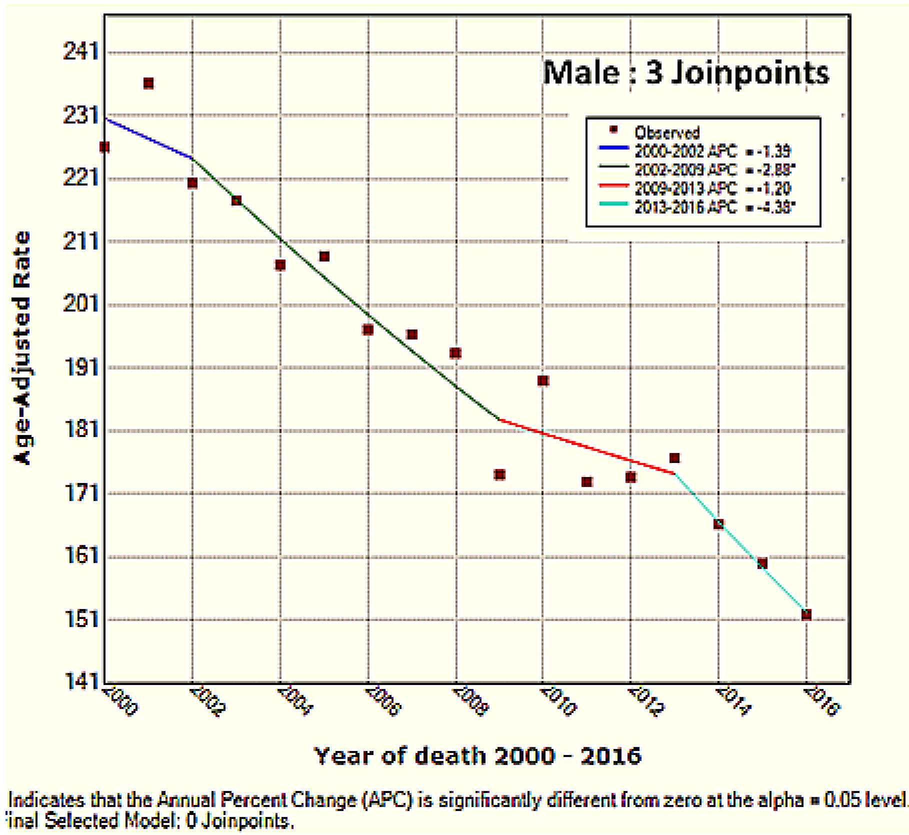


Figure 4-5: Male age-adjusted lung cancer mortality rates 2000-2016.

Figure 4-5 represents the trend for male age-adjusted lung cancer mortality rates. Lung cancer mortality rates for males decreased by 1.4% per year from 2000-2002. Between 2002 – 2009 rates decreased by 2.9% per year. During 2009 through 2013, it was only a 1.2% decrease per year. But again, rates significantly reduced during 2013 – 2016 by 4.4% per year.

4.3.3 Lung Cancer Mortality Trend: Female

According to Table 4.2, the average annual percentage change (AAPC) for female lung cancer mortality rates declined by 0.6 % per year for the 16 years of 2000 through 2016. For males, it was around a 2.6% decline.

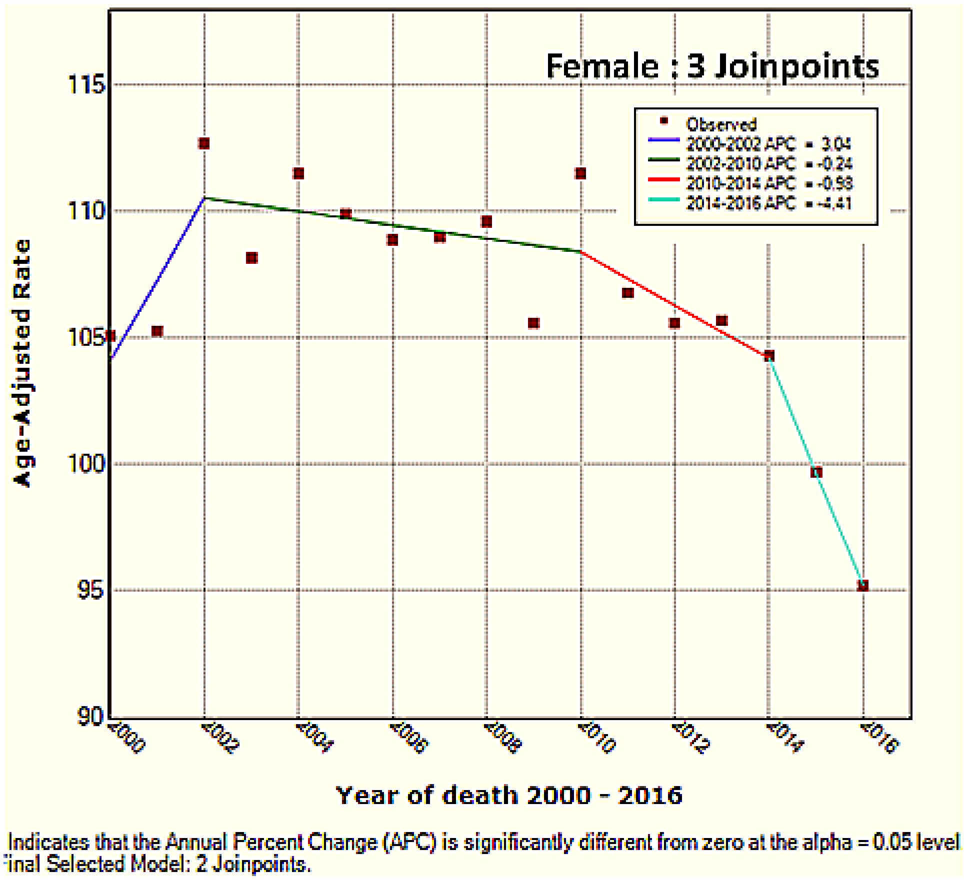


Figure 4-6: Female age-adjusted lung cancer mortality rates 2000-2016.

According to Figure 4-6, female age-adjusted lung cancer mortality rates increased by 3.04% per year for 2000-2002. However, between 2002 – 2010 rates continue to decrease by 0.2 % per year. From 2010 through 2014, rates decreased by 1.0% per year. But rates significantly reduced during 2014 – 2016 by 4.4 % per year.

4.4 Comparison of Lung Cancer Incidences and Mortality Trends

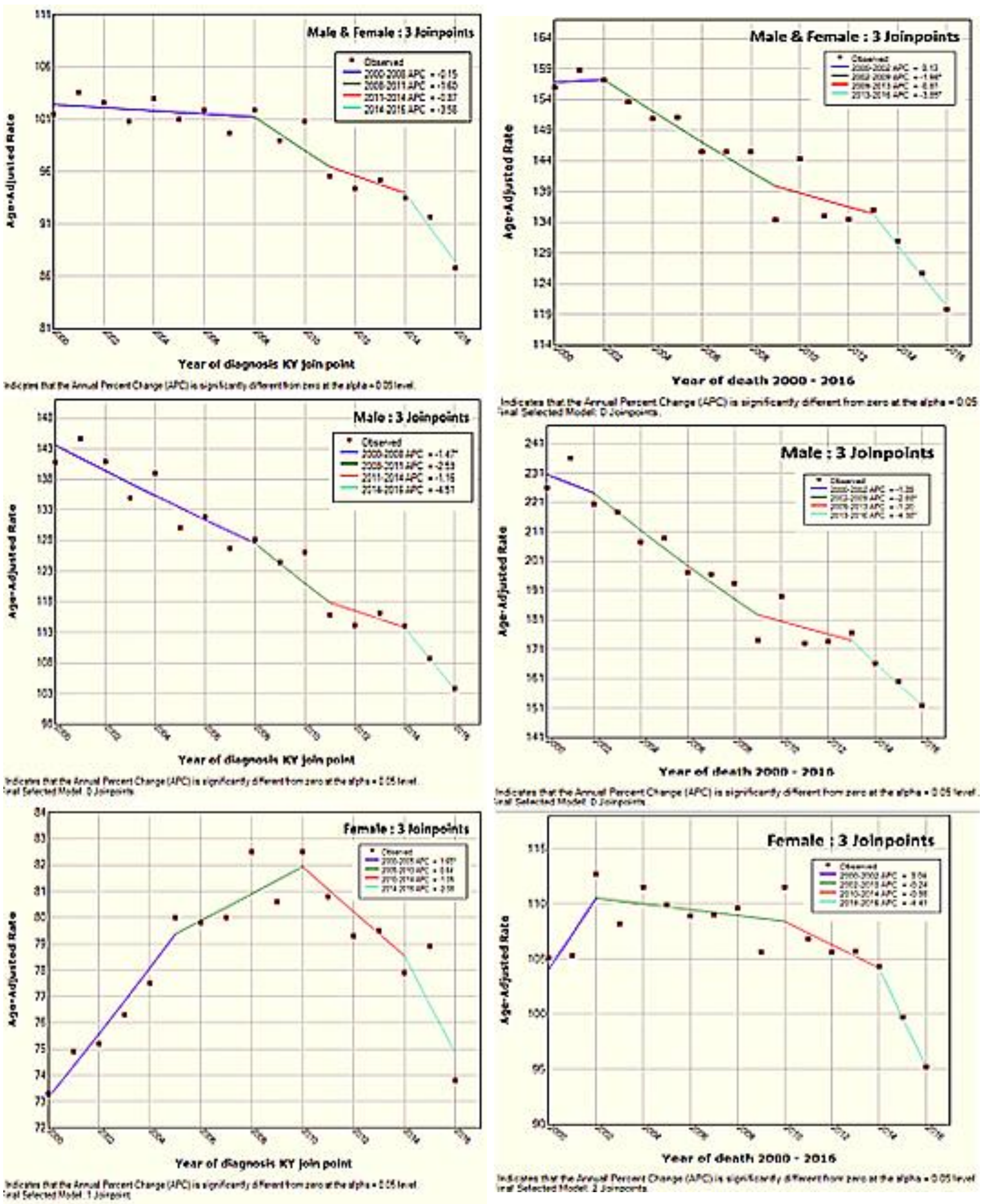


Figure 4-7: Comparison of lung cancer incidence and mortality rates 2000-2016

The Joinpoint analysis of the trends in the lung cancer incidence & mortality rates allows the user to interpret changes more accurately over time and, more importantly, determine if those changes are statistically significant. Also, help to graphically display the results of the Joinpoint analysis are shown in Figure 4-7 by different gender groups.

Overall, Kentucky lung cancer incidence trends show that progress is being made to reduce the lung cancer burden among residents of Kentucky. The age-adjusted lung cancer incidence rates have shown a significant decline by 1.0% between 2000-2016 for both men and women. Age-adjusted lung cancer mortality rates also showed a substantial decrease by 1.6% between 2000 – 2016 for men and women. Medical advances and the growth in cancer knowledge, technology, and resources have contributed to this progress. Overall, lung cancer incidence rates and mortality rates are considerably lower for women than for men.

A declining trend is observed in lung cancer incidence and mortality in men. However, incidence and mortality follow similar trend characteristics. For example, between 2011-2014, incidence and mortality declined by 1.2%, and in 2014 – 2016 it was around a 4.5% decline. The result explained that men's highest lung cancer incidence and mortality rate appeared in 2000.

But the age-adjusted lung cancer incidence rates among women have risen significantly between 2000 – 2005 (1.6%) and 2005 – 2010 (0.6%), but it started to decline after 2010. Results indicate that lung cancer mortality in women in Kentucky peaked in 2010 and began to decrease until the year 2016. A noticeable trend is observed for lung cancer mortality rates for women. Between 2000 – 2003, the mortality trend considerably increased by 3.0%, but after 2003 it continued to decline. However, age-adjusted lung cancer mortality rates have been declining among men since 2000.

4.5 Conclusion

Mortality rates are a better indicator of prevention measures against cancer than incidence or survival rates because they are less impacted by biases resulting from changes in detection practices(Siegel, Miller, & Jemal, 2020). However, the cancer mortality rate rose during most of the 20th century, mainly because of the tobacco epidemic's rapid increase in lung cancer deaths among men(Siegel, Miller, Fuchs, & Jemal, 2021). This increasing trend appears at the beginning of a lung cancer incidence and mortality during 2000 -2002 in Kentucky for men and women. For example, the highest lung cancer incidence and mortality for men occurred during 2000 -2002.

According to the literature, the lung cancer epidemic is associated with tobacco use because of the continued decline in the prevalence of smokers in recent decades. However, declines in smoking and improvements in early detection and treatment have resulted in a continuous reduction around 2014 – 2016 in the cancer incidence and death rate (Siegel et al., 2020). This declining trend appears in Kentucky's lung cancer incidence and mortality during 2014-2016 for men and women. For example, between 2014-2016, lung cancer mortality and incidence declined around 4.5% for men in Kentucky.

A previous study evaluated lung cancer incidence and survival according to cancer subtype, sex, and trends in incidence-based mortality. Results reveal that Among men, incidence-based mortality from Non-small cell lung cancer (NSCLC) decreased 6.3% annually from

2013 through 2016, whereas the incidence decreased 3.1% annually from 2008 through 2016 (Howlader et al., 2020). This declining trend also appears in Kentucky, suggesting the highest 4.5% mortality decline during 2014-2016 among men.

Although in the U.S., smoking rates have historically been lower among women than men, they have not declined as quickly for women as for men. For example, since 2005, smoking rates among women have reduced to 25.4 % compared with a 26.8 % decline among men. Additionally, smoking rates among women have dropped by about 59% since 1965, compared with a 66 % drop among men (truth initiative, 2019). This trend can be found in the women lung cancer incidence and mortality in Kentucky.

Also, population-level mortality from NSCLC in the United States fell sharply from 2013 to 2016, and survival after diagnosis improved substantially (Howlader et al., 2020). Our analysis suggests that a reduction in smoking and treatment advances, particularly approvals for and use of targeted therapies, is likely to explain the decline in mortality observed during this period.

However, our study also has some limitations. First, the underlying relationship cannot be established because Joinpoint regression consists of trend analysis in incidence and mortality. So, study results require further confirmation with individual-level data. As such, the effect only hypothesizes about associations highlighted by our data and has strayed from making connection claims.

Chapter 5

Exploring Geographic Location and Social Determinants of Health on Lung Cancer Mortality

5.1 Introduction

This chapter reveals the State of Kentucky relationship between socioeconomic variables and lung cancer mortality rate at the county scale using the ordinary least squares method; Moran's I, and Geographically Weighted Regression (GWR) method. The independent variables include four socioeconomic factors: adult smoking rate, median household income, high school graduation rate, and the number of coal mine employment.

Result demonstrates the relationship between lung cancer mortality rate in 54 Appalachian counties and 66 non-Appalachian counties in Kentucky.

5.2 Ordinary Least Square Regression (OLS)

The regression takes the age-adjusted lung cancer mortality rate as the dependent variable. The independent variables are adult smoking rate, median household income, high school graduation rate, and the number of coal mine employment. The relationship was examined on a county-wide basis using Ordinary Least Square (OLS) regression. The purpose was to test the significance of the variables and potential multicollinearity problems among the variables. The model is:

$$\begin{aligned} LCM_I = \ & \beta_O + \ \beta_1\, Smoking\ Rate + \ \beta_2\, Median\ Houshold\ Income \\ & + \ \beta_3\, High\ School\ Graduation\ Rate \\ & + \ \beta_4\ No\ Coal\ Mine\ Employement + \ e \end{aligned}$$

Where LMC stands for lung cancer mortality rates; β0, β1, β2, β3, and β4 are the regression coefficients; and *e* is the random error in the two models.

Table 5.1 illustrates the minimum, maximum, and mean values of dependent and independent variables and the statistical differences. According to table 5.1, there are 120 total observations in Kentucky. Lung cancer mortality range between 30.1 – 122.1. The mean mortality rate is 73.82 per 100,000 people.

The mean adult smoking rate is 22.64%, and the adult smoking range is between 15.88% – 30.75%. The median household income range is \$25,344 – \$97,960 per year, and the mean household income is \$43,483.67 per year.

The mean high school graduation rate is 94.58%, and the minimum and maximum are 84%, 100% respectively. The maximum number of coal mine employment 19804 people and mean is 912.46 people.

Table 5.1: Minimum, maximum, and mean values of dependent and independent variables and statistical difference.

Variables	Min	Max	Mean	SD	No. observation
Dependent Variable					
Lung Cancer Mortality Rate	30.1	122.1	73.823	16.07	120
Independent Variables					
Adult Smoking Rate	15.88	30.75	22.64	2.88	120
Median Household Income	25344	97960	43483.67	11662.18	120
High School Graduation Rate	84.07	100.00	94.58	3.33	120
Number of Coal Mine Employment	0	19804	912.46	2714.93	120

Figure 5-1 illustrates the distribution of dependent and independent variables used in this study. According to figure 5-1(a) the highest lung cancer mortality rates are clustered in the east region of Kentucky. The highest adult smoking rate (b) and low median household income (d) also clustered in the eastern part of the state. The highest number of coal mine employment (e) is also located in the same region.

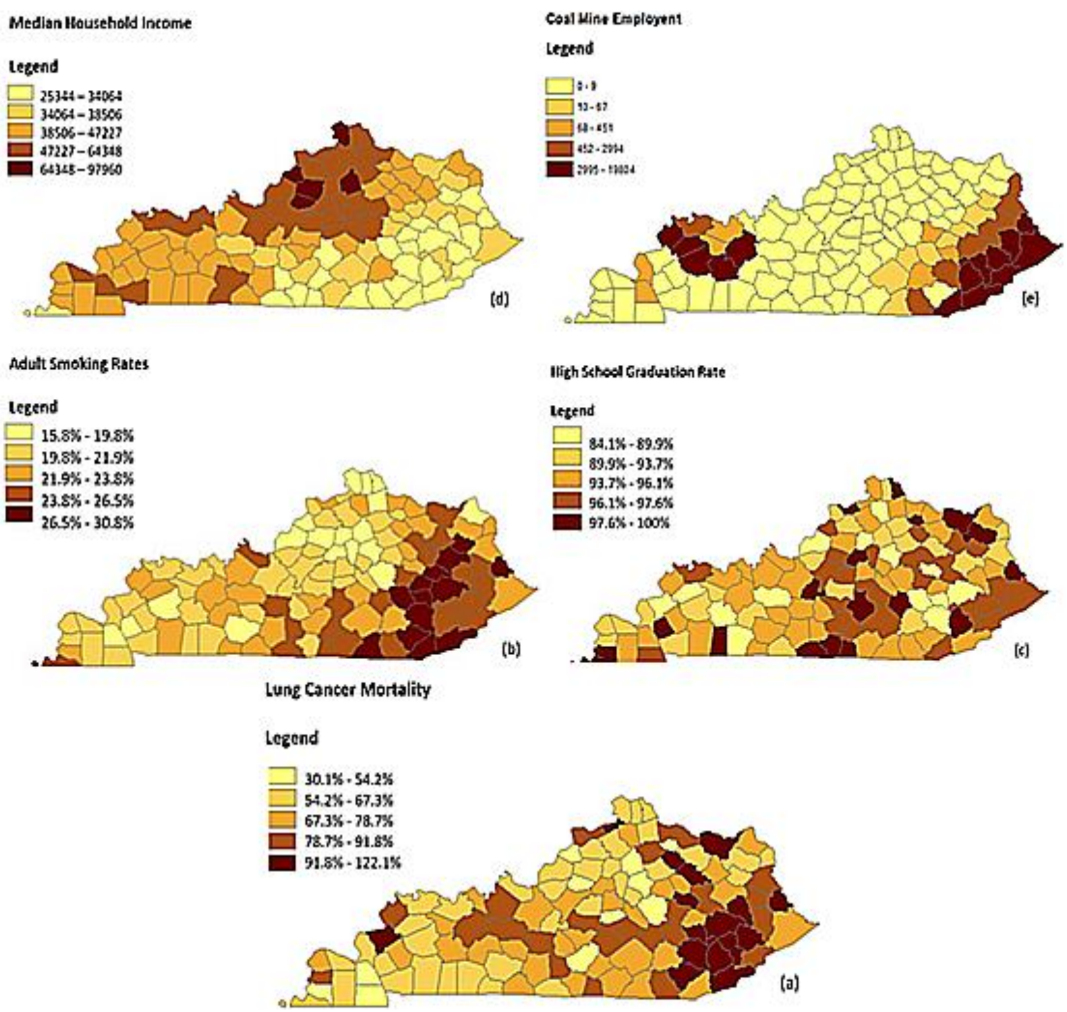


Figure 5-1: Distribution patterns of socio-economic variables.

Distribution patterns of variables: lung cancer mortality rates (a) adult smoking rate (b), high school graduation rate (c), median household income (d) and number of coal mine employment (e).

Table 5.2: Result of ordinary least square regression.

Variable	Coefficient	SE	p - Value	VIF
All States KY				
Intercept	127.9986071	37.8601795	0.00098791	
Adult smoking Rate	2.009289325	0.6444644	0.002302551	2.6776
HS Graduation Rate	-0.90595723	0.34157884	0.009124621	1.0065
Median Household Income	-0.000340764	0.00016036	0.035731557	2.7167
Coal Mine Employment	0.000917453	0.00043322	0.036352007	1.0745
Moran's, I index	-0.029186			
Adjusted R2	0.407280469			
AIC	951.99134			
Not Appalachian counties				
Intercept	77.84445481	48.7636294	0.115575924	
Adult smoking Rate	2.859771003	1.0701747	0.009650383	1.8267
HS Graduation Rate	-0.673354305	0.44551243	0.135846184	1.0461
Median Household Income	-0.000138012	0.00017832	0.441947963	1.8519
Coal Mine Employment	0.000103851	0.00093209	0.911651224	1.0581
Moran's, I index	0.093051			
Adjusted R2	0.227209441			
AIC	513.46365			
Appalachian counties				
Intercept	241.4057286	67.962229	0.000856295	
Adult smoking Rate	0.349516427	1.09067095	0.749982291	2.4459
HS Graduation Rate	-1.42558749	0.57399242	0.016473893	1.0981
Median Household Income	-0.000973681	0.00047061	0.043844938	2.5953
Coal Mine Employment	0.001003062	0.00054464	0.071574098	1.1181
Moran's, I Index	-0.015525			
Adjusted R2	0.336552893			
AIC	443.56292			

Table 5.2 illustrates the result of OLS regression for all the geographic regions, such as all counties in Kentucky, the Appalachian region, and the non-Appalachian region. The coefficient, standard error (SE), P-value, and variance inflation factor (VIF) values are provided.

5.1.1 OLS Regression for All Counties in KY

The OLS regression for all counties in KY results reveal the significant variables for lung cancer rates: adult smoking rate, high school graduation rate, median household income, and the number of coal mine employment (Table 5.2).

The adult smoking variable and number of coal mine employment has a positive coefficient (2.009, 0.0009), indicating that the relationship is positive. In other words, the lung cancer mortality rate is higher in areas with high adult smoking rates and high coal mine employment.

In addition, the negative sign of the high school graduation rate (-0.905) and median household income (-0.0003) variables indicates that lung cancer mortality rate is higher with low high school graduation rate and low median household income.

The Variance Inflation Factor (VIF) values for all counties in Kentucky in Table 5.2 do not suggest any high multicollinearity among the independent variables. The coefficient of determination R2 for lung cancer mortality rate is 0.40; a significant amount of variance is unexplained. The residual maps show some spatial autocorrelation in the residuals. The Moran's I of the residuals is -0.025 ($p < 0.01$). The spatial autocorrelation in the residuals suggests there is some negative auto correlation with dissimilar values.

5.1.2 OLS Regression for Non-Appalachian Counties

The OLS regression for non-Appalachian counties in KY shows that the significant variables for lung cancer mortality rates are adult smoking rates (Table 5.2). In addition, the adult smoking variable has a positive coefficient (2.85), indicating that the relationship with lung cancer is positive. In other words, in the non-Appalachian region, the lung cancer mortality rate is higher in areas with high adult smoking rates.

The Variance Inflation Factor (VIF) values in non-Appalachian counties (Table 5.2) do not suggest any high multicollinearity among the independent variables. The coefficient of determination R2 for lung cancer mortality rate is 0.22, a significant amount of variance unexplained. The residual maps show some spatial autocorrelation in the residuals. The Moran's I of the residuals is 0.09 ($p < 0.01$). The spatial autocorrelation in the residuals suggests some positive autocorrelation that is unexplained by the global OLS model.

5.1.3 OLS Regression for Appalachian Counties

The OLS regression for Appalachian counties in KY result shows that the significant variables for lung cancer mortality rates are high school graduation rate and median household income (Table 5.2). High school graduation rate and median household income have negative coefficients (-1.42, -0.0009). The negative sign of the high school graduation rate and median household income variables suggests that it is more common for higher lung cancer mortality rates with low graduation rates and low median household income.

The Variance Inflation Factor (VIF) values in Table 5.2 do not suggest any high multicollinearity among the independent variables. The coefficient of determination R2 for lung cancer mortality rate is 0.33; a significant amount of variance is unexplained. The residual maps show some spatial autocorrelation in the residuals. The Moran's I of the residuals is -0.015 ($p < 0.01$). The spatial autocorrelation in the residuals suggests there is some negative autocorrelation with dissimilar values.

Spatially correlated variability is unexplained by the global OLS model. Therefore, we used the local regression model instead of the global model, which allows the regression coefficients to vary over the spatial domain.

5.2 Geographically weighted regression (GWR)

Table 5.3 Result of geographically weighted regression. (Global and Local Parameter Estimates of the Model)

Variables	Minimum	Lower Quartile	Median	OLS Coefi	P - Value	Upper Quartile	Maximum
Intercept	36.424415	113.99314	156.3247	127.99860	0.00098791	193.9094	296.43892
Adult Smoking Rate	-1.54005	0.750617	1.438707	2.009289	0.00230255	1.935344	4.267447
HS Graduation Rate	-1.441087	-1.193958	-0.997896	-0.905957	0.00912462	-0.734065	-0.414901
M.Household Income	-0.001432	-0.00072	-0.000339	-0.000341	0.03573156	-0.000284	-0.000059
CoalMine Employemt	-0.079728	0.000395	0.000893	0.000917	0.03635201	0.001842	0.004883
R2			0.447934	0.40728			
AIC			950.75232	951.991346			
F Statistic				21.442373 *(0.00)			

Geographically weighted regression (GWR) analyses were conducted using ESRI's ArcGIS 10.1 and GWR 4 software packages. Table 5.3 presents the descriptive statistics of the parameter estimates from OLS and GWR models. In Table 5.3, R2 denotes the coefficient of determination, and AIC indicates Akaike Information criterion.

The global model OLS estimates, presented in Table 5.1, indicate that increases in adult smoking and coal mine employment positively affect lung cancer mortality. Also, high school graduation rate and median household income show a negative effect on lung cancer mortality. The negative sign of the high school graduation rate (-0.905) and median household income (-0.0003) variables indicates that lung cancer mortality rate is higher with low high school graduation rate and low median household income.

The model selection criterion (AIC) indicates the selection of the GWR model. Additionally, the F statistics reported at the bottom of Table 5.3 show the rejection of the null hypothesis (P-value 0.00 for the partial F-test), suggesting that the GWR model significantly improves model fit over the OLS model. Therefore, the GWR model shows significant improvement over the OLS model (Table 5.3). It returns an overall R2 of 0.44, much better than the OLS model (R2 = 0.40). Figure 5-2 shows the spatial pattern of the locally weighted R2. The spatial distribution of parameters of the local model is presented in Figures 5-2 and 5-3.

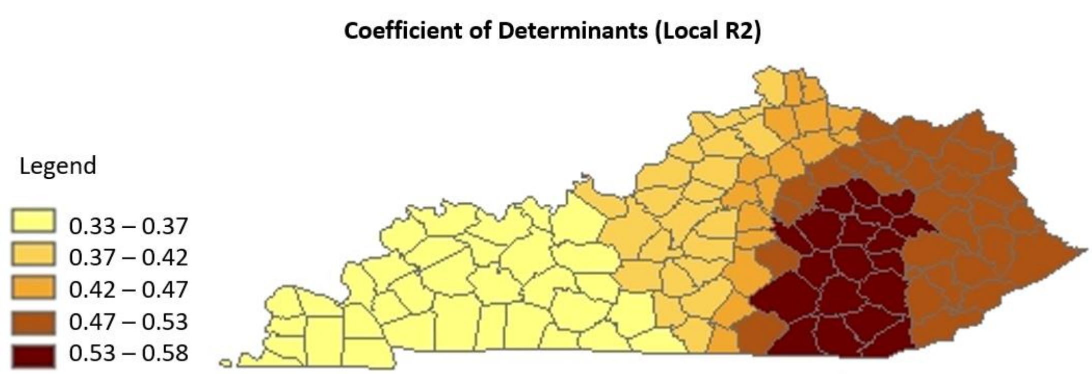


Figure 5-2: Coefficient of determinants (R2) map of the GWR model.

The spatial distribution of R2 is not even over the study area (Figure 5-2). Some counties have a high R2 (of up to 0.58), and some are very low. Generally, the counties in the Appalachian region (most areas of the eastern and central areas) have better regression results than others. Not surprisingly, there are many areas where significant associations were found and can be targeted to reduce lung cancer incidence and mortality.

Figures 5-3 illustrates coefficients of the intercept, adult smoking rate, high school graduation rate, median household income, and the number of coal mine employment. Figure 5-3 (A) shows the spatial patterns of the GWR model coefficients. The intercept is lower in the western, central, and north-central counties, indicating a generally lower lung cancer mortality rate in those areas, and counties in the eastern area indicate a higher lung cancer mortality rate.

Figure 5-3 (B) shows that high school graduation rates are strongly associated with lung cancer mortality in central and north-central counties and low in western and eastern counties. This suggests that the light color areas are in the study area where the high school graduation rate variable predicts lung cancer mortality. The dark areas are locations where the high school graduation rate is less critical. But the consistency in the high school graduation coefficients leads to the conclusion that the high school graduation rate is the crucial factor behind the lung cancer problem in Kentucky. The global OLS model also showed high school graduation rate is a significant factor in explaining the variability of lung cancer mortality in Kentucky.

Figure 5-3 (C) shows that median household income is strongly negatively associated with lung cancer mortality in central and north-central counties and lowest in eastern counties.

This suggests that the light color areas are where the median household income variable is a predictor of lung cancer mortality. The dark areas are locations where the median household income is less critical than in light-colored counties. But the consistency in the median household income coefficients indicates that the median household income is the crucial factor behind the lung cancer problem in Kentucky. The global OLS model also showed median household income is a significant factor in explaining the variability of lung cancer mortality in Kentucky.

The relationship between adult smoking and lung cancer mortality rate is strongly positively associated in some counties in the north-central area of Kentucky and negatively associated in the eastern counties. (Figures 5-3 (D)). This suggests that the dark color areas are where the adult smoking variable is a strong predictor of lung cancer mortality. The light color areas are locations where the adult smoking rate is less important.

The consistency in the positive adult smoking coefficients in the non-Appalachian area (north-central area) leads to the conclusion that adult smoking is the crucial factor behind the lung cancer problem in this region. But the relationship between adult smoking and lung cancer mortality is negative in most counties in the Appalachian region. This suggests that an increase in adult smoking may not increase lung cancer mortality in these areas. The global OLS model also showed adult smoking is a significant factor in explaining the variability of lung cancer mortality in non-Appalachian counties in Kentucky. But adult smoking is not a significant factor in the Appalachian region.

According to GWR coefficients, the relationship between adult smoking and lung cancer is negative in most counties in the Appalachian region, with outliers of slightly positive values in the north-central area. The outliers are mainly in high population density areas. According to the Appalachian regional commission population in the Appalachian region was 1,159,828 and non-Appalachian 3,307,845 in 2018. Further investigation of counties with negative coefficients might be interesting, but this is beyond the scope of the research reported here.

The number of coal mine employment is strongly and positively correlated with the lung cancer mortality rate in only three counties of Kentucky (Figure 5-3(E)). This suggests that the dark color areas where the number of coal-mine employment variables is a strong predictor of lung cancer mortality. Conversely, the light color areas are locations where the number of coal-mine employment is less critical.

The consistency in the coal-mine employment coefficients in the central region indicates that the number of coal-mine employment is the essential factor behind the lung cancer mortality in Kentucky. In addition, the global OLS model also showed the number of coal-mine employment is a significant factor in explaining the variability of lung cancer mortality in all counties in Kentucky.

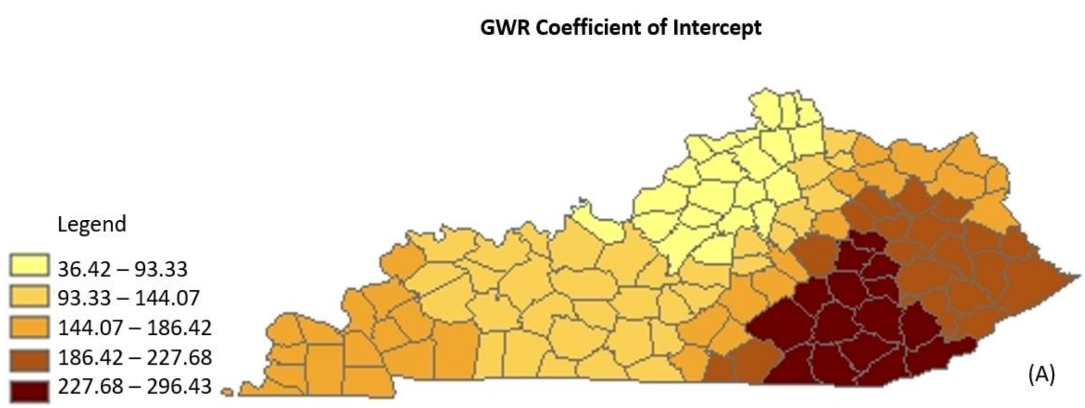


Figure 5-3: (A) GWR Coefficient of intercept

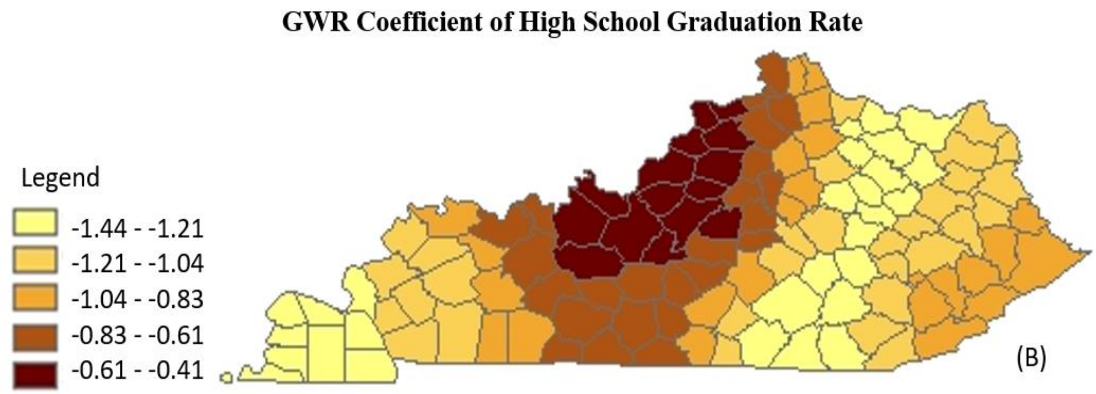


Figure 5-3:(B) GWR coefficient of high school graduation rate.

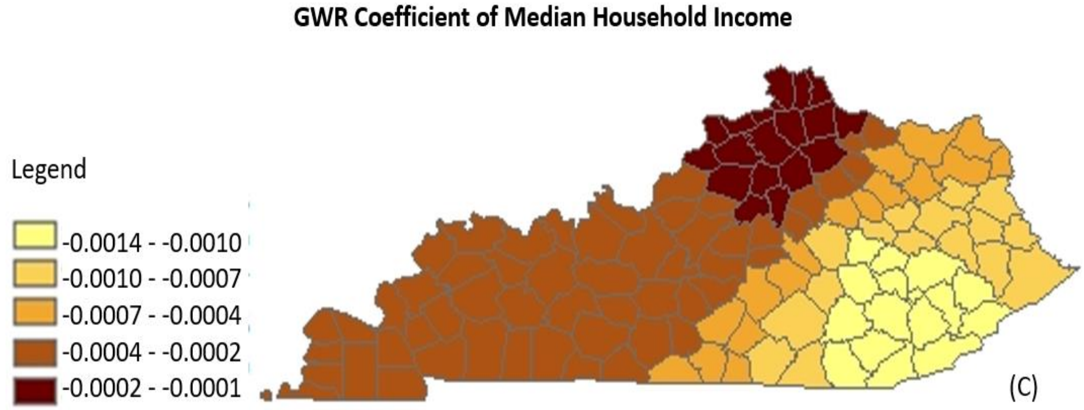


Figure 5-3:(C) GWR coefficient of median household income

Figure 5-3:(D) GWR coefficient of adult smoking

Figure 5-3:(E) GWR coefficient of coal- mine employment

Figure 5-3 Spatial patterns of the GWR model coefficients. (A) coefficients of the Intercept, (B) high school graduation rate, (C) median household income, (D) adult smoking rate, and the (E) number of coal mine employment.

5.3 Conclusion

This study examined the statewide relationship between socioeconomic variables and lung cancer mortality at the county scale by using the ordinary least squares method and the Geographically Weighted Regression method. The independent variables include four socioeconomic factors, such as adult smoking rates, median household income, high school graduation rate, and the number of coal-mine employment.

Appalachian counties showed the strongest statistical association between lung cancer mortality rates with median household income and high school graduation rate, which may explain higher lung cancer mortality in this region. High school graduation rate and median household income have negative coefficients. The negative sign of the high school graduation rate and median household income variables suggests that it is more common for the highest lung cancer mortality rates with low graduation rates and low median household income. The global OLS model and the local regression model have shown that median household income and high school graduation rate are significant factors in explaining the variability of lung cancer mortality in the Appalachian counties.

Adult smoking rates showed the strongest association with non-Appalachian counties in Kentucky. However, some outliers in the local regression model were observed in the adult smoking rate. For example, some counties in the north-central region of Kentucky have positive coefficients (higher adult smoking is related to higher lung cancer mortality). On the other hand, the Appalachian region has negative coefficients (lower adult smoking is related to higher lung cancer mortality). An inconsistency may cause such outliers in the population data used in the study. Another possible reason for this relationship might be the use of centroids of census tracts to approximate the population centers in the algorithm. Further refinement of the adult smoking rates might help improve the use of the findings in public health studies.

In addition, the global OLS regression model showed statewide adult smoking rate, high school graduation rate, median household income, and the number of coal mine employment have a strong relationship with lung cancer mortality rates.

The local regression model (GWR) has its strength in finding geographic heterogeneity among counties by clustering their coefficients. The spatial patterns of coefficients are more valuable than the regression itself to geographical analysis. General statistic methods used in Human Geography have been criticized for generalizing human objects and neglecting the spatial structure of society (Xu & Wang, 2014). The use of the localized regression model compensates for the weakness of statistical models that ignore spatial heterogeneity. The GWR is more powerful in explaining the variability in lung cancer mortality than the OLS model when adult smoking rate, high school graduation rate, median household income, and the number of coal mine employment classifications are used.

The spatial pattern of their coefficients is more interesting to Human Geographers than the regression itself. In each coefficient map, one can visually identify clusters of counties that are significantly different from other areas. Therefore, public health policies cannot depend on a global model. For example, the global model identifies the smoking rate as significantly contributing to lung cancer mortality in the non-Appalachian region. But from the local model, identify specific counties with positive coefficients. This means that the global model's conclusion regarding the smoking rate does not apply equally across the region. Therefore, public policies should be flexible and consider the unique characteristics of each region.

Education level can influence occupation, income, adherence to healthy behaviors, and participation in health promotion and screening programs. This finding indicates that Kentucky's emphasis on improving graduation rates may reduce lung cancer mortality and increase personal income and may be able to address health disparities between Appalachian counties. However, Kentucky has a long way to address these significant health issues. The first steps could be raising policies and programs to reduce tobacco use and implementing strict smoke-free laws in north-central regions.

Chapter 6

Geographical Detector-Based Assessment of the Lung Cancer Mortality Rate in Kentucky.

6.1 Introduction

Mutual interaction of physical environment, social environment, and health behaviors can be significant causes of human diseases. These disease determinants have individual spatial distributions across geographical units so that their satisfactory study involves the investigation of the associated geographical strata. Examining the spatial distribution patterns of disease and suspected determinants could help to understand health risks. This chapter illustrates the study's result investigating the potential risk factors associated with lung cancer mortality in the State of Kentucky.

Lung cancer mortality data were collected from SEER*Stat database for 120 counties from 2012 to 2016. Nine potential risk factors are incorporated in this study. Such as socioeconomic variables (high school graduation rate, unemployment rate, median household income, number of coal mine employment), health behaviors factors (adult smoking rate, uninsured rate, physically unhealthy days) environment factors (radon and P.M 2.5) were gathered and considered as potential risk factors.

This chapter describes the result of four geographical detectors based on spatial variation analysis of the geographical strata to assess the socio-economic and environmental risks of lung cancer mortality: the risk detector result indicates where the risk areas are; the factor detector identifies factors that are responsible for the risk; the ecological detector reveals relative importance between the factors, and the interaction detector discloses whether the risk factors interrelate or lead to disease independently.

6.1 Classifications of the Explanatory Variables

To analyze the influence of risk determinants of lung cancer mortality rates, the Geographical Detector method first needs to create discrete values for these risk factors and then turn original data into continuous data layers. Therefore, multiple sorting techniques were used in this study, such as the Jenks Natural Breaks classification method and the Geometric interval method.

Diverse and complex risk factors determine lung cancer mortality. Research literature has found various risk factors that may increase the chances of getting lung cancer. Figure 6-1 illustrates the classifications of the explanatory variables used in this study.

Figure 6-1(a) displays the map of lung cancer mortality rate in Kentucky. There is a high lung cancer mortality cluster in the east region of Kentucky. And the western part of the state has a lower lung cancer mortality rate.

The following maps display the socio-economic risk factors of lung cancer mortality, such as Figure 6-1(b) demonstrate the map of median household income in Kentucky. The eastern region of Kentucky has lower household income, and the central area has higher household income levels.

Figure 6-1(c) provides the map of the high school graduation rate in Kentucky. This map does not reveal any significant trend in high school graduation levels.

Unemployment rates are illustrated in Figure 6-1 (d). According to the map, the highest unemployment rate is displayed in the eastern part of Kentucky.

Figure 6-1 (e) provides a map of the coal mine employment rate in Kentucky. The map reveals that the eastern region has higher coal mine employment and few counties in the western part of the state have coal mine employment. The rest of the counties in Kentucky have the lowest level of coal mine employment.

Health behavior risk factors are displayed in the following figures. Figure 6-1(f) is the map of smoking rates in Kentucky. The smoking map reveals the highest rate of smoking is in the eastern part of Kentucky.

Figure 6-1(g) explains the uninsured rate in Kentucky. Again, the East region of Kentucky displays the highest rate of uninsured people.

Physically unhealthy days are illustrated by Figure 6-1(h). Again, counties in the east region of Kentucky show the highest rate of physically unhealthy days compared to the other counties in Kentucky.

Figures (i) and (j) demonstrate the environmental risk factors of lung cancer mortality, such as the radon level of Kentucky revealed by Figure 6-1(i). According to the radon map, counties with high radon levels are in the western region of Kentucky.

Figure 6-1 (h) displays the PM 2.5 level in Kentucky. High PM 2.5 levels can be seen in the counties in the central and western regions.

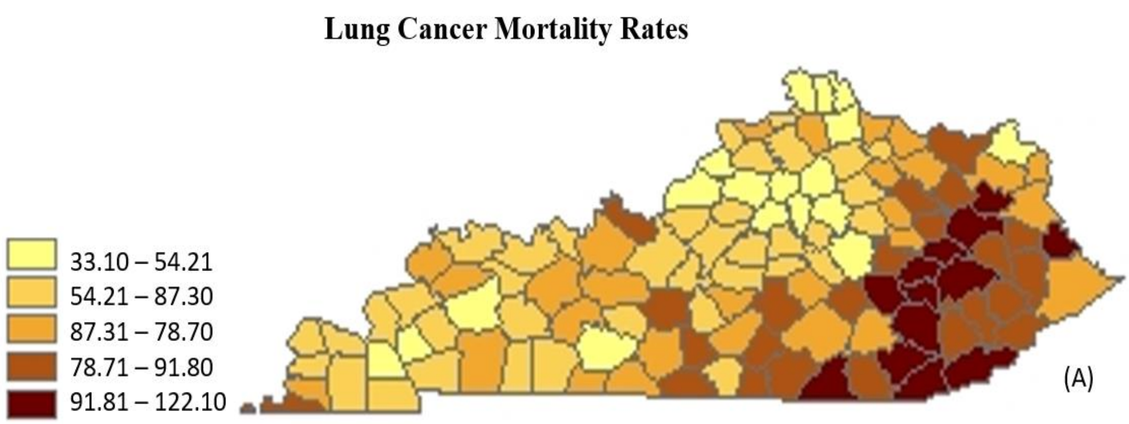


Figure 6-1:(a) Lung cancer mortality rates

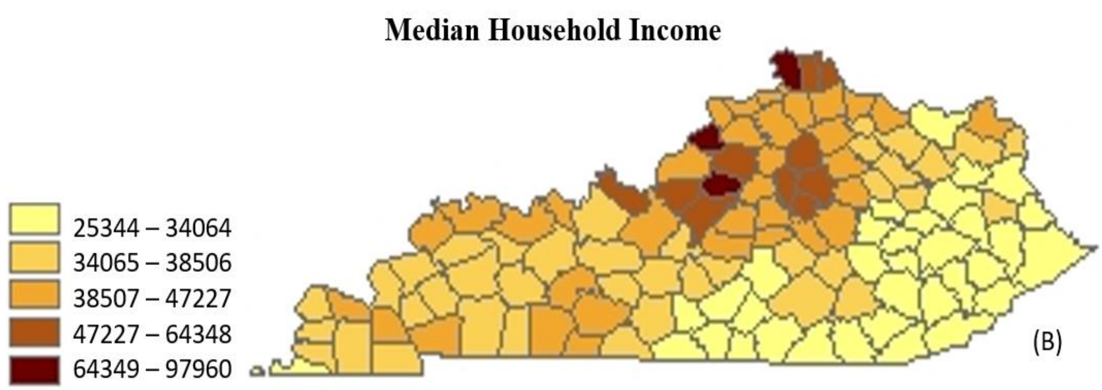


Figure 6-1:(b) Median household income

High School Graduation Rate

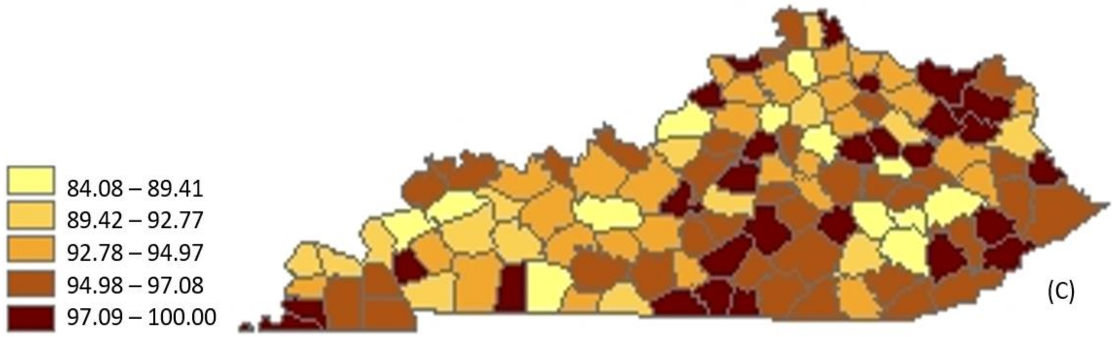


Figure 6-1:(c) High school graduation rate

Unemployment Rate

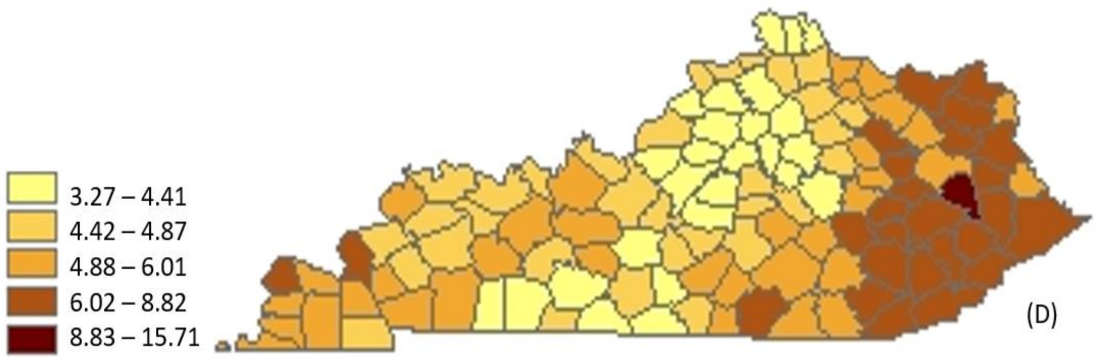


Figure 6-1:(d) Unemployment rate

Number of Coal-mine Employment

Figure 6-1:(e) Number of coal-mine employment

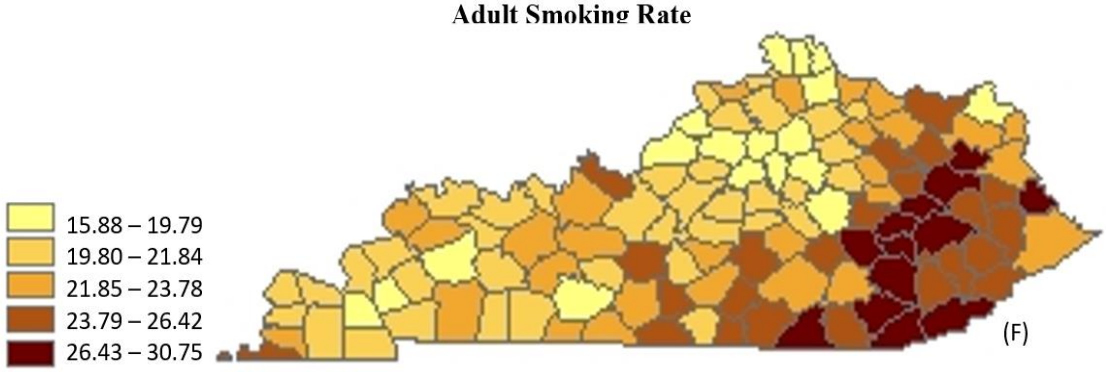


Figure 6-1:(f) Adult smoking rate

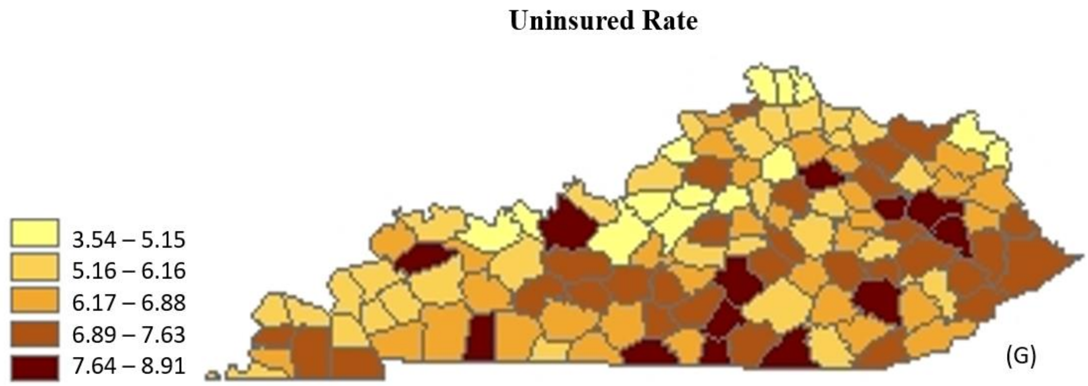


Figure 6-1:(g) Uninsured rate

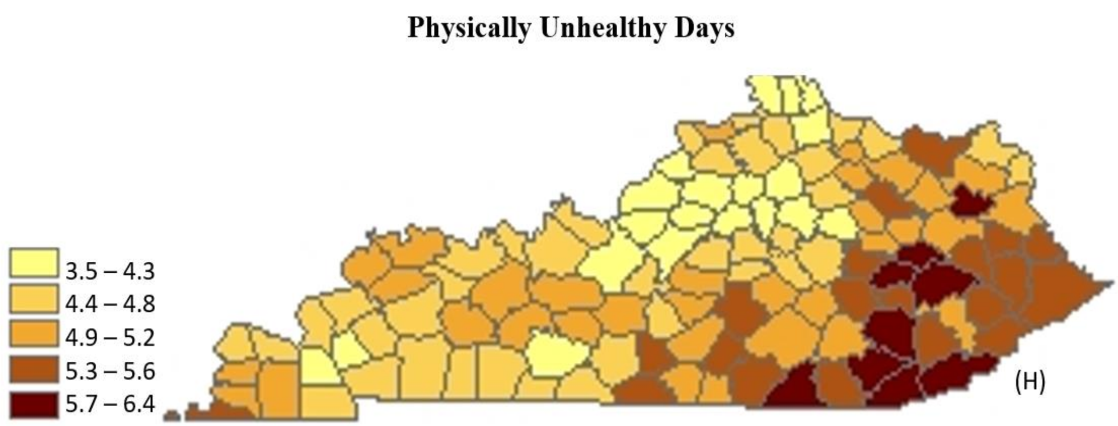


Figure 6-1:(h) Physically unhealthy days

Figure 6-1:(i) Radon zones

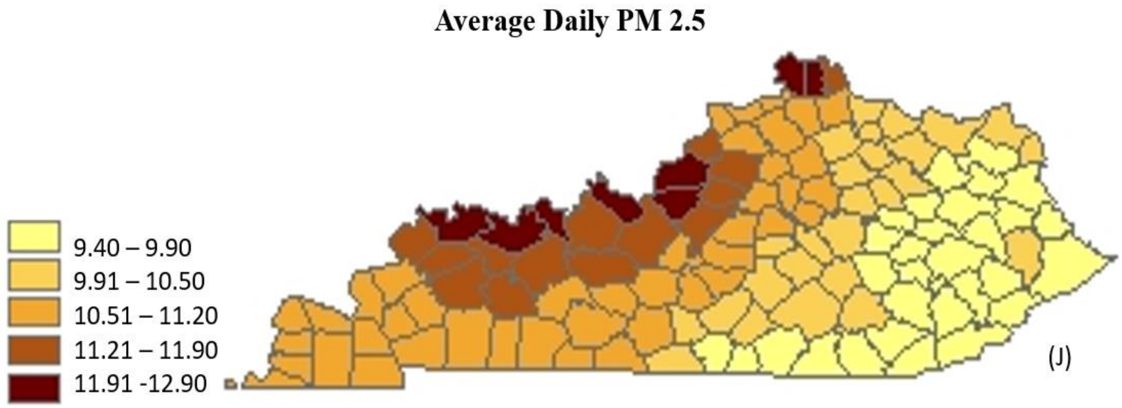


Figure 6-1:(j) Average daily PM 2.5

Figure 6-1: Maps of explanatory variables of lung cancer morality. (a) lung cancer mortality, (b) median household income, (c) high school graduation rate, (d) unemployment rate, (e) number of coal mine employment, (f) adult smoking rate, (g) uninsured rate, (h) physically unhealthy days, (i) radon zones, (j) daily PM 2.5 levels.

The results of the Geographical Detector are recorded in Table 6.1 (risk detector), Table 6.2 (ecological detector), and Table 6.3 (interactive detector).

6.2 Result of Risk Detector

The risk detector examines the influence of various factors on the lung cancer mortality rate. Table 6.1 presents each subregion's average mortality rate when the study area is stratified by a corresponding explanatory variable. According to the risk detector, when the adult smoking rate is high, the lung cancer mortality rate is also high, especially when the adult smoking rate is higher than 26.42%. The mean lung cancer mortality rate is 88.99% (rate per 100,000 people). This finding means that there is a correlation between the adult smoking rate and the lung cancer mortality rate.

The lung cancer mortality rate also becomes smaller with the increase in median household income. When medium household income is high, the lung cancer mortality rate is low. For example, median household income is higher than $64348.28, the mean lung cancer mortality rate is only 53.38 (rate per 100,000 people). Conversely, when median household income is low ($25344 – $34064) mean lung cancer mortality rate is as high as 87.3%.

Also, when the high school graduation rate increased, the lung cancer mortality rate declined. For example, when the high school graduation rate is around 84 %– 89%, mean lung cancer mortality rate is 81.5%. But when the high school graduation rate increased up to 97% - 100%, lung cancer mortality is only 72.5%.

When the unemployment rate increases, lung cancer mortality rates increase. The unemployment rate is higher than 8.83% mean lung cancer mortality rate is 87.51% (rate per 100,000 people). This indicates that there is a correlation between unemployment and lung cancer mortality.

Table 6.1: Result of risk detector

Variables	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Stratum 5
Adult smoking rate Values in Each Stratum	15.88 - 19.79	19.80 - 21.84	21.85 - 23.78	23.79 - 26.42	26.43 - 30.75
Average Mortality Rate in Each Stratum	58.5	68.30789	75.10714	84.7913	88.99231
Median household income Values in Each Stratum	25344.00 - 34064.74	34064.75 - 38506.80	38506.81 - 47227.54	47227.55 - 64348.27	64348.28 - 97960.00
Average Mortality Rate in Each Stratum	87.3	80.5	70.87895	66.87714	53.38
Unemployment Values in Each Stratum	3.27 - 4.41	4.42 - 4.87	4.88 - 6.01	6.02 - 8.82	8.83 - 15.71
Average Mortality Rate in Each Stratum	62.74231	70.415	74.22143	79.82162	87.51111
Coal Mine Employment Values in Each Stratum	0.00 - 8.80	8.81 - 66.96	66.97 - 451.50	451.51 - 2993.97	2993.98 - 19804.00
Average Mortality Rate in Each Stratum	70.22892	75.7	76.7	84.075	86.76429
Physically unhealthy days Values in Each Stratum	3.57 - 4.35	4.36 - 4.81	4.82 - 5.23	5.24 - 5.68	5.69 - 6.43
Average Mortality Rate in Each Stratum	59.285	69.51053	75.82727	88.38421	85.01
Uninsured rate Values in Each Stratum	3.54 - 5.15	5.16 - 6.16	6.17 - 6.88	6.89 - 7.63	7.64 - 8.91
Average Mortality Rate in Each Stratum	61.92143	69.84194	76.22258	79.40645	77.1
High school graduation rate Values in Each Stratum	84.08 - 89.41	89.42 - 92.77	92.78 - 94.97	94.98 - 97.08	97.09 - 100.00
Average Mortality Rate in Each Stratum	81.5	70.76	73.78621	73.32	72.56786
P.M 2.5 Values in Each Stratum	9.40 - 9.90	9.91 - 10.50	10.51 - 11.20	11.21 - 11.90	11.91 - 12.90
Average Mortality Rate in Each Stratum	87.9	72.51923	68.69286	68.25	61.9625
Radon Values in Each Stratum	1	1.01 - 2.00	2.01 - 3.00		
Average Mortality Rate in Each Stratum	66.97667	77.37561	63.0875		

Note: average of the explained variable (lung cancer mortality rate) according to the stratums of each explanatory variable.

If coal mine employment is higher than 2993.98, mean lung cancer mortality is 89.76%. But the lower rate of coal mine employment reveals low mean lung cancer mortality.

Overall, the average lung cancer mortality rate rises gradually with the increase in physically unhealthy days. For example, when physically unhealthy days increased 5.69 – 6.43, mean lung cancer mortality increased 85.01%. But physically unhealthy days between 3.57 – 4.35 show only a 59.2% lung cancer mortality rate.

The uninsured rate also follows the same trend. For example, when the uninsured rate increased 7.64-8.91 mean lung cancer rate increased by 77.1%. But lowest uninsured rate displays (3.54 – 5.15) only 61.9% lung cancer mortality.

The correlations between the average lung cancer mortality rate and the rest of the factors can be examined in the same way based on the results of the risk detector.

6.3 Result of Factor Detector

According to the power determinant (PD) estimate, the factor detector discovers the extent to which a factor explains the variation of the lung cancer mortality rate. The factor detector in the Geo-detector model ranks the influence of affecting factors on the lung cancer mortality rate by PD value as follows: adult smoking rate (0.36) > physically unhealthy days (0.33) > medium household income (0.31) > PM 2.5 (0.28) > unemployment rate (0.20) > number of coal mine employment (0.14) > uninsured rate (0.12) > radon level (0.10) > high school graduation rate (0.03).

Factor detector shows that adult smoking rate explains the spatial variability of the lung cancer mortality rate to the maximum extent. Followed by physically unhealthy days, medium household income, PM 2.5, unemployment rate, number of coal mine employment, uninsured rate, and radon level, while high school graduation rate has a minor influence.

6.4 Result of Ecological Detector

Table 6.2: Result of the ecological detector.

	PUnhealthyD	Uninsured	Unemployment	P.M2.5	ASmoking	HSGraduation	MHIncome	Coalmine	Radon
PUnhealthyD									
Uninsured	N								
Unemployment	N	N							
P.M2.5	N	N	N						
ASmoking	N	y	N	N					
HSGraduation	N	N	N	N	N				
MHIncome	N	y	N	N	N	y			
Coalmine	N	N	N	N	N	N	N		
Radon	N	N	N	N	N	N	N	N	

Note: Y means the variation between the influences of two factors on the lung cancer mortality rate is statistically significant with 95% confidence, and N indicates that there is not.

The ecological detector recognizes the difference between the values of two PDs: in other words, the difference between the impacts of two factors on the explained variable. For example, Table 6.2 shows that the differences between the high school graduation rate and household medium income factors are statistically significant. Also, PD values of uninsured rate with adult smoking and medium household income are statistically significant. However, the differences between the rest of the factors are not statistically significant.

However, the differences between any one of the first four factors and any one of the rest of the factors are statistically significant. With the factor detector and the ecological detector, the analysis shows that adult smoking rate and medium household income substantially affect lung cancer mortality. In contrast, the remaining factors have a weak effect.

6.5 Result of Interactive Detector

Table 6.3: Result of the interactive detector.

	PUnhealthyD	Uninsured	Unemployed	P.M2.5	Adultsmoking	HSGraduation	MHIncome	Coalmine	RadonZone
PUnhealthyD	0.336379543								
Uninsured	0.385600488	0.1222088							
Unemployed	0.412904476	0.314425	0.209594893						
P.M2.5	0.422528351	0.3775129	0.381857931	0.2813773					
Adultsmoking	0.424398081	0.425428	0.425671726	0.4406919	0.3636995				
HSGraduation	0.435098887	0.2596314	0.355654833	0.3714603	0.4600664	0.0312083			
MHIncome	0.374233649	0.3572699	0.388843686	0.3715595	0.4212932	0.4248596	0.310267		
Coalmine	0.414988575	0.361096	0.320987509	0.3779412	0.495112	0.3394222	0.432201	0.1412241	
RadonZone	0.422025446	0.2658424	0.382545809	0.3346643	0.4220103	0.2161175	0.409824	0.202697	0.10934057

The interactive detector defines the interaction effects between pairs of PDs. The results demonstrated in Table 6.3 are impressive, as the interaction impacts are either “enhance, bivariate” or “enhance, nonlinear.” This implies that the joint effects of two factors on the lung cancer mortality rate measured by the PD are more significant than the effects of two different factors. For example, the interactive PD of physically unhealthy days and the high school graduation rate is 0.435, which is higher than PDs of two sole factors, physically unhealthy days (0.33) and high school graduation rate (0.03) (0.33+0.03 = 0.36<0.435). Furthermore, the interactive effect is stronger than the sum of two individual results, so the interactive effect between physically unhealthy days and the high school graduation rate is "enhanced, nonlinear."

Also, the interactive PD of uninsured rate and high school graduation rate is 0.259, which is higher than PDs of two individual factors, uninsured rate (0.12) and high school graduation rate (0.03) (0.12+0.03 = 0.15<0.259). Hence, the interactive effect between uninsured and high school graduation rates is "enhanced, nonlinear."

Interactive PD of uninsured rate and the number of coal mine employment rate is 0.361, which is higher than PDs of two individual factors, uninsured rate (0.12) and the number of coal mine employment (0.14) (0.12+0.14 = 0.26<0.361) so, the interactive effect between uninsured rate and coal mine employment rate is "enhance, nonlinear." This means when uninsured people work in the coal mine industry, the death rate of lung cancer increases.

The interactive effects between uninsured rate and radon zone increased lung cancer mortality. Because interactive PD of uninsured rate and radon zone is 0.265, which is higher than PDs of two sole factors, uninsured rate (0.12) and radon zone (0.10) (0.12+0.10 = 0.22<0.265) so, the interactive effect between uninsured rate and radon zone is "enhance, nonlinear."

PD values of unemployment and high school graduation rate, unemployment, and radon zone are "enhance, nonlinear," which has the most potent effect on lung cancer mortality.

The interactive effects between the unemployment rate and radon zone increased lung cancer mortality. Because interactive PD of uninsured rate and radon zone is 0.382, which is higher than PDs of two sole factors, unemployment rate (0.20) and radon zone (0.10) (0.12+0.10) = 0.30<0.382 so, the interactive effect between the unemployment rate and radon zone is “enhance, nonlinear.”

Also, the interactive effects between high school graduation rate and unemployment increased lung cancer mortality. Because interactive PD of high school graduation rate and unemployment is 0.355, which is higher than PDs of two sole factors, high school graduation rate (0.03) and unemployment (0.20) (0.03+0.20 = 0.23<0.355) so, the interactive effect between high school graduation rate and unemployment is “enhanced, nonlinear.”

The interactive effect of the following risk determinants shows the most substantial effect on lung cancer mortality: PM 2.5 and high school graduation, adult smoking and high school graduation, high school graduation with medium household income, number of coal mine employment and radon zone.

The interactive PD of uninsured and medium household income is 0.357, which is higher than PDs of two sole factors, uninsured (0.12) household income (0.31) (0.12+0.31 = 0.43 >0.35). The interactive impact is weaker than the sum of two individual effects, so the interactive effect between uninsured rate and household income is “enhance, bivariate.” Therefore, the uninsured rate and medium household income could reinforce each other’s influence on the lung cancer mortality rate.

6.6 Conclusion

The Geographical Detector method is new as it extracts the associations between the observed process and possible influencing factors by the consistency of their spatial distribution patterns. It is an efficient tool and easy to implement. This study applied four geographical detectors to analyze the effects of the physical environment, social environment, and health behaviors on lung cancer mortality rate. The study goal was to determine the differences of the degrees to which factors influence the spatial distribution of the lung cancer mortality rate and the interaction effects between different factors.

Firstly, the study focused on which factors play more critical roles in the lung cancer mortality rate. According to the results of four geographical detectors, adult smoking and median household income were the first two most important factors responsible for the lung cancer mortality rate. The higher the adult smoking is, the higher the mortality rate is valid for median household income. In a national case-control study conducted in Canada, researchers found that the likelihood of developing lung cancer for men and women was significantly higher in people with low incomes, so study outcomes were reliable with previous studies (Yang Mao et al., 2001).

In general, counties with higher adult smoking have higher lung cancer mortality. Therefore, adult smoking was positively associated with the lung cancer mortality rate in Kentucky. Previous research demonstrated that counties with higher smoking percentages had more lung cancer diagnoses and deaths; this result is consistent with this study (Hopenhayn, Jenkins, & Petrik, 2003). This may be why people in Kentucky have a higher prevalence of secondhand smoke, which increases the risk of lung cancer. Previous research reveal that 27% of Kentucky's blue-collar workers are exposed to secondhand smoke at work (Hahn E, 2008). Compared with physically unhealthy days, P.M 2.5, unemployment rate, number of coal mine employment, uninsured rate, radon level, high school graduation rate had relatively small impacts on the lung cancer mortality rate.

The factor detector ranked the influence of risk factors according to their PD values: adult smoking, physically unhealthy days, medium household income, PM 2.5, unemployment rate, number of coal mine employment, uninsured rate, radon level, and high school graduation rate. Physically unhealthy days were 0.33. Cancer-related side effects, primarily pain, and exhaustion were altogether connected with physically unhealthy days. Researchers found that patients with pain had 83% more undesirable days than patients without pain; patients with fatigue had 104% more unhealthy days than patients without fatigue (Casebeer et al., 2019). PM2.5 has been demonstrated as an essential factor of lung cancer. Previous epidemiological studies have indicated that ambient PM2.5 may increase the morbidity and mortality rates associated with lung cancer, and PM2.5 has been suggested to decrease the survival time of patients with lung cancer (J. Li, Li, Bai, & Song, 2017). A study conducted in Poland reveals a significant positive correlation between the unemployment rate and lung cancer incidence rates in the male population was recognized (Chawińska, Tukiendorf, & Miszczyk, 2013). Working in coal mines has been associated with an elevated lung cancer risk. PD values of the study were 0.14. Existing research studies reveal that the high mortality risk of respiratory disease is associated with residents living in Virginia coal-mining counties (Shi et al., 2019).

The ecological detector recognizes the difference between the values of two PDs: in other words, the difference between the impacts of two factors on the risk factors. For example, the study result shows that the differences between the high school graduation rate and household medium income factors statistically significantly influence lung cancer mortality. Sidorchuk et al (2009) found that lung cancer incidence was associated with low educational, occupational, and income (Sidorchuk et al., 2009). Also, PD values of uninsured rate with adult smoking and medium household income are statistically significant. However, the differences between the rest of the factors are not statistically significant.

Through the interactive detector, the study examined the interaction effects between pairs of factors. The results demonstrated that these interactive effects were either "enhance, bivariate" or "enhance, nonlinear." Therefore, for any two factors considered in this study, they had a more substantial influence on the lung cancer mortality rate when they were taken together than when taken independently.

For example, the interaction of physically unhealthy days and the high school graduation rate nonlinearly enhanced lung cancer mortality. Also, the interactive effect between uninsured and high school graduation rates is nonlinear. This finding may be due to the relationship between low rate of graduation rate with low income. People without good income cannot afford an insurance plan, or they work in low-paying, blue-collar jobs. These conditions may increase exposure to a toxic environment and finally lead to lung cancer.

Additionally, adult smoking and high school graduation, high school graduation with medium household income, high school graduation rate, and unemployment increased the interactive effect of lung cancer mortality. Previous research reveals that people with a high school education smoke cigarettes for a duration of more than twice as many years as people with at least a bachelor's degree (Siahpush, Singh, Jones, & Timsina, 2010). Also, people in the most socioeconomically deprived groups, such as low income and educational attainment, have higher lung cancer risk than those in the most affluent groups (Clegg et al., 2009).

Chapter 7

An Overview of Methods for Monitoring Health Disparities.

Lung Cancer Mortality Trend Analysis by Socioeconomic Quantile and Geographic Region 2002 – 2019.

This chapter summarizes the outcome of socioeconomic and geographic disparities in lung cancer mortality. Six measures of relative disparity and four measures of absolute disparity were used in this study. In addition, lung cancer mortality results were explained using four socioeconomic factors with male-female disparity differences. Also, Appalachian, and non-Appalachian disparity outcomes for males and females were discussed.

7.1 Lung Cancer Mortality Trends

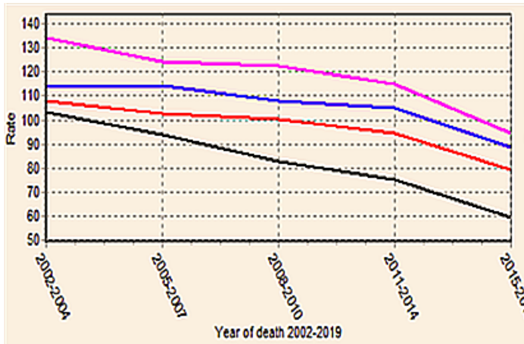
Table 7.1 describes the age-adjusted lung cancer mortality rates, percentage change between 2002 -2019, and population distribution for the first and last year observation period by gender, socioeconomic variables, and geographic region.

To give a complete picture of changes in lung cancer mortality, Figure 7-1 illustrates the trends in the age-adjusted mortality of lung cancer, by gender, for all socioeconomic variables characterized by socioeconomic quantiles and different geographic regions in Kentucky.

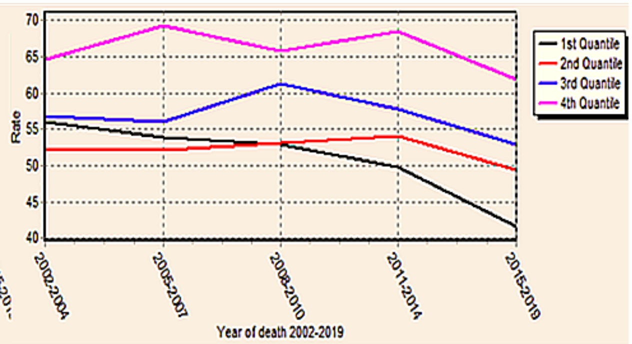
Table 7.1 and Figure 7-1 shows that lung cancer morality generally declined among males and females for all socioeconomic variables and geographic regions. Still, the picture was more mixed among females, with rates increasing among some quantile groups. For example, the second quantile of median household income, those living in counties with 3299 – 3975 median household income mortality increased between 2005 – 2008. In addition, there is considerably more variation in socioeconomic variables than men's age-adjusted mortality rate. Therefore, it is essential to individually analyze the male and female variations to reveal the importance of each socioeconomic factor.

Table 7.1: lung cancer mortality and population distribution according to socioeconomic variables and geographic region by gender, Kentucky, 2002 – 2019

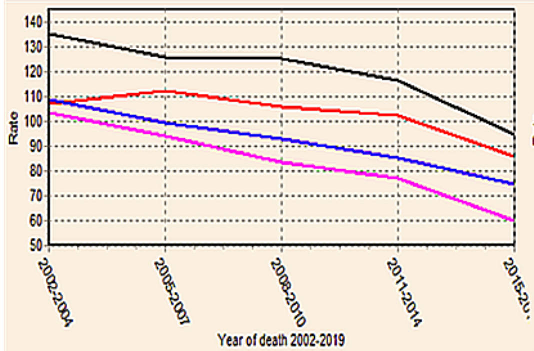
Education - Male	2002-2004	2005-2007	2008-2010	2011-2014	2015-2019	% Change 2002-2004	2002 - 2004	2015 - 2019
1st Quantile (7.22% - 14.20%)	103	94	82.9	75.3	59.4	-42.33	57.051	59.587
2nd Quantile (14.21% - 19.67%)	108.2	102.7	100.3	94.7	79.5	-26.525	16.455	16.426
3rd Quantile (19.68% - 24.03%)	114	114.1	107.9	104.8	88.6	-22.281	13.082	12.369
4th Quantile (24.04% - 36.33%)	134.2	124.4	122.8	114.8	94.5	-29.583	13.412	11.619
Education - Female	2002-2004	2005-2007	2008-2010	2011-2014	2015-2019	% Change 2002-2004	2002 - 2004	2015 - 2019
1st Quantile (7.22% - 14.20%)	56	53.8	52.9	49.9	41.7	-25.536	57.381	59.913
2nd Quantile (14.21% - 19.67%)	52.3	52.1	53.2	54	49.4	-5.545	16.533	16.565
3rd Quantile (19.68% - 24.03%)	56.8	56.1	61.3	57.8	53	-6.69	12.974	12.213
4th Quantile (24.04% - 36.33%)	64.6	69.2	65.7	68.5	61.8	-4.334	13.112	11.309
Income - Male	2002-2004	2005-2007	2008-2010	2011-2014	2015-2019	% Change 2002-2004	2002 - 2004	2015 - 2019
1st Quantile (1897 - 3298)	135.4	125.8	125.2	116.6	94.9	-29.911	14.232	12.4
2nd Quantile (3299 - 3975)	107.4	112.6	105.8	102.6	85.7	-20.205	14.678	14.206
3rd Quantile (3976 - 4503)	108.6	99.2	92.7	85.5	74.7	-31.215	17.848	17.117
4th Quantile (4504 - 8632)	103.8	93.9	83.5	76.9	60	-42.197	53.242	56.276
Income - Female	2002-2004	2005-2007	2008-2010	2011-2014	2015-2019	% Change 2002-2004	2002 - 2004	2015 - 2019
1st Quantile (1897 - 3298)	65.8	69.4	66.1	67.5	62.8	-4.559	13.977	12.1
2nd Quantile (3299 - 3975)	54	52.1	59.8	57.9	50.8	-5.926	14.678	14.253
3rd Quantile (3976 - 4503)	52.9	55.3	51.9	50.8	46	-13.043	17.74	16.896
4th Quantile (4504 - 8632)	56.3	53.7	53.3	50.7	42.5	-24.512	53.605	56.751
Poverty - Male	2002-2004	2005-2007	2008-2010	2011-2014	2015-2019	% Change 2002-2004	2002 - 2004	2015 - 2019
1st quantile (5.98% - 16.98%)	104.5	94.2	86.3	77.3	61	-41.627	44.329	46.033
2nd Quantile (16.99% - 20.39%)	104.2	98.2	84.9	82.2	67.4	-35.317	26.064	26.331
3rd Quantile (20.40% - 25.37%)	110.1	111.6	106.2	103.8	84.5	-23.252	16.617	16.244
4th Quantile (25.38% - 42.50%)	138.2	126.4	127.1	115.7	98.9	-28.437	12.99	11.392
Poverty - Female	2002-2004	2005-2007	2008-2010	2011-2014	2015-2019	% Change 2002-2004	2002 - 2004	2015 - 2019
1st quantile (5.98% - 16.98%)	57.4	55	53.7	50.8	43.3	-24.564	44.71	46.381
2nd Quantile (16.99% - 20.39%)	51.6	51	51.5	50.4	42.9	-16.86	25.914	26.24
3rd Quantile (20.40% - 25.37%)	54.3	57.4	59.3	56	50.3	-7.366	16.562	16.22
4th Quantile (25.38% - 42.50%)	67.1	67	67.1	70.9	65.3	-2.683	12.814	11.16
Unemployement - Male	2002-2004	2005-2007	2008-2010	2011-2014	2015-2019	% Change 2002-2004	2002 - 2004	2015 - 2019
1st Quantile (4.10% - 6.74%)	104.3	94.7	83.1	77.1	60.1	-42.378	31.679	33.411
2nd Quantile (6.75% - 7.90%)	106.2	98.1	88.8	82	66.7	-37.194	35.372	36.249
3rd Quantile (7.91% - 10.17%)	109	108.5	105.7	98.4	82.5	-24.312	16.639	16.024
4th Quantile (10.18% - 18.95%)	129.4	120.3	118.7	112.1	93	-28.13	16.31	14.317
Unemployement - Female	2002-2004	2005-2007	2008-2010	2011-2014	2015-2019	% Change 2002-2004	2002 - 2004	2015 - 2019
1st Quantile (4.10% - 6.74%)	57.6	52	54.2	48.6	41.4	-28.125	31.524	33.33
2nd Quantile (6.75% - 7.90%)	54.2	55	52.1	52.6	44.8	-17.343	36.108	36.93
3rd Quantile (7.91% - 10.17%)	52.6	55	59.8	56.5	51.4	-2.281	16.502	15.903
4th Quantile (10.18% - 18.95%)	65	66.4	63.1	65.9	59.5	-8.462	15.866	13.838
	2002-2004	2005-2007	2008-2010	2011-2014	2015-2019	% Change 2002-2004	2002 - 2004	2015 - 2019
Appalachia - Male	125	115.8	111.7	107.6	89.4	-28.48	28.454	26.306
Not Appalachia - Male	103.4	97	87.7	80	64.2	-37.911	71.546	73.694
Appalachia - Female	60.2	60.5	62.2	61.4	56	-6.977	28.229	26.101
Not Appalachia - Female	55.2	54	53.2	51.2	43.5	-21.196	71.771	73.899



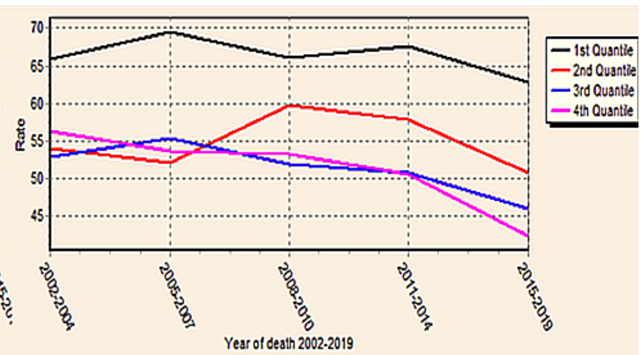
(A) % less than high school education



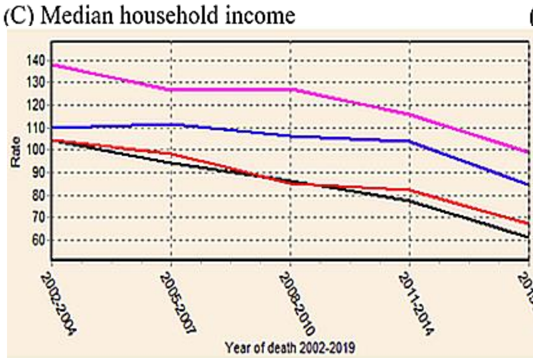
(B) % less than high school education



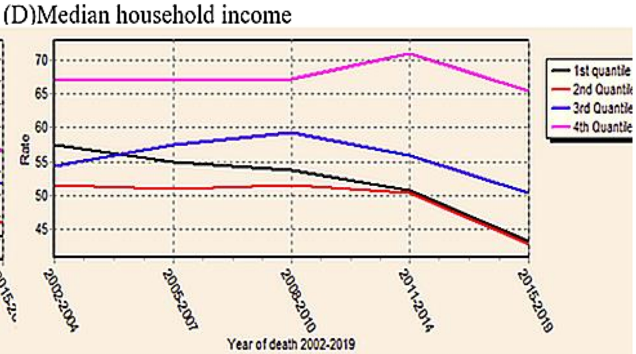
(C) Median household income



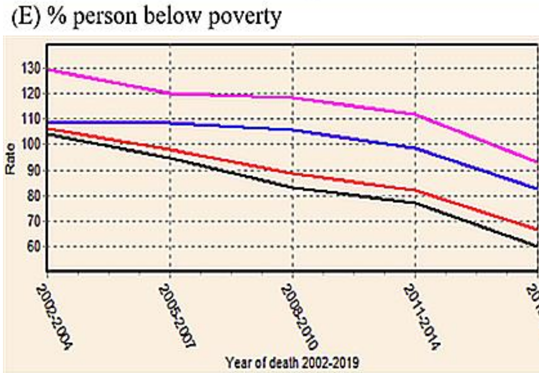
(D)Median household income



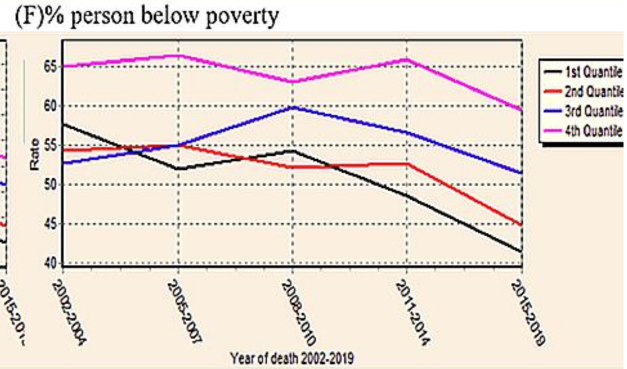
(E) % person below poverty



(F)% person below poverty

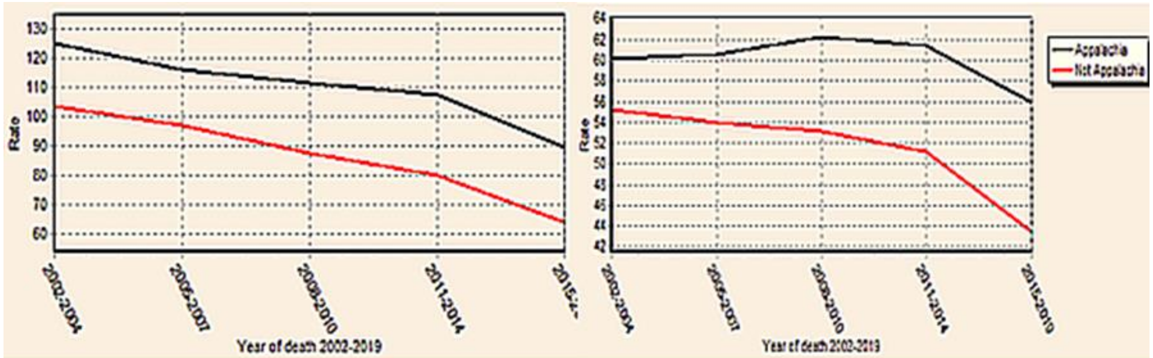


(G) % Unemployment



(H)% unemployment

Figure 7-1: Age-adjusted lung cancer mortality rate among males and females, by socio-economic variables 2002 – 2019.



(I)Male -Appalachian and Non- Appalachian (I)Female -Appalachian and Non- Appalachian

Figure 7-2 (I): Age-adjusted lung cancer mortality rate among males and females, by geographic region 2002 – 2019.

7.2 Education Disparities in Lung Cancer Mortality, 2002 - 2019, Kentucky.

To measure educational disparity for all counties in Kentucky, the counties from the SEER database (n=120) were ranked according to the percentage of the population ages 25 and over with at least a high school degree, estimated from the Census 2012-2016 American Community Survey (ACS). Less than high school graduate attainment for male and female ranged from 7.22 % in Oldham County to 36.3% in Clay County. To determine quintile cut points for a user-defined variable based on what percentage of the county population had less than high school education use the county attributes data. Based on that percentage, less than high school graduate was divided into four quantiles, such as, first quantile (7.22% - 14.20%), second quantile (14.21% - 19.67%), third quantile (19.68% - 24.03%) and fourth quantile (24.03% - 36.33%).

7.2.1 Educational Disparities in Lung Cancer Mortality by Male

Rates of lung cancer mortality from 2002 to 2019 for males with less than high school education, are shown in quantiles by percentage in Figure 7-1(A). Quantile one is the most advantageous group because it belongs to the lowest quantity of individuals with less than high school education. And the fourth quantile is considered the least advantageous group with the highest rate of less than high school education.

According to Figure 7-1(A), lung cancer mortality has declined for all education quantile groups, and the magnitude of the decline was generally between 42% - 22%. Table 7.1 illustrates the mortality rates for each year and percentage change in each education group from 2002 – 2019. 4^{th} quantile group represents the highest rate of lung cancer mortality each year, which is the least advantageous group. Conversely, the 1^{st} quantile that is the most advantageous group reveals the lowest lung cancer mortality rate.

The percentage less than high school education disparity measures for males and females is shown in Table 7.2. Four measures of absolute disparity in lung cancer mortality show a decline in disparity in ACI and SII by 49% and 54%. But RD and BGV values indicate a positive direction.

In six measures of relative disparity, RR, Idisp, MLD, and T, values show an increase in disparity, but RCI and RII reveal negative disparity around 132% - 140%, respectively.

.

Table 7.2: Education disparity in lung cancer mortality between 2002 - 2019, Kentucky – males & females.

	ABSOLUTE MEASURES				RELATIVE MEASURES					
Education - Male	**RD**	**BGV**	**ACI**	**SII**	**RR**	**Idisp**	**MLD**	**T**	**RCI**	**RII**
2002-2004	31.2	108.857	-4.75	-35.392	1.303	15.34	0.004	0.004	-0.043	-0.323
2005-2007	30.4	121.274	-5.489	-41.091	1.323	20.993	0.005	0.006	-0.054	-0.403
2008-2010	39.9	208.048	-7.331	-55.278	1.481	33.092	0.011	0.011	-0.078	-0.588
2011-2014	39.5	224.737	-7.659	-58.293	1.525	39.132	0.014	0.014	-0.088	-0.67
2015-2019	35.1	194.156	-7.105	-54.619	1.591	47.363	0.019	0.019	-0.101	-0.776
% Change 2002-2019	12.5	78.358	-49.6	-54.327	22.104	208.756	347.026	338.124	-132.672	-140.024
Education - Female	**RD**	**BGV**	**ACI**	**SII**	**RR**	**Idisp**	**MLD**	**T**	**RCI**	**RII**
2002-2004	12.3	11.66	-0.851	-6.369	1.235	13.066	0.002	0.002	-0.015	-0.112
2005-2007	17.1	27.341	-1.794	-13.502	1.328	14.587	0.004	0.004	-0.032	-0.242
2008-2010	12.8	21.845	-2.061	-15.629	1.242	13.548	0.003	0.003	-0.037	-0.281
2011-2014	18.6	36.671	-2.809	-21.49	1.373	20.441	0.006	0.006	-0.052	-0.4
2015-2019	20.1	46.814	-3.38	-26.093	1.482	31.255	0.01	0.01	-0.072	-0.56
% Change 2002-2019	63.415	301.483	-296.91	-309.669	19.984	139.215	469.174	479.133	-381.955	-397.448

Absolute measures: RD – Rate Difference, BVG – Between Group Variances, ACI – Absolute Concentration Index, SII – Slope Index Inequality.

Relative Measures: RR – Rate Ratio, Idisp – Index of Disparity, MLD – Mean Log Deviation, T- Theil Index, RCI- Relative Concentration Index, RII- Relative Index of Inequality.

7.2.2 Educational Disparities in Lung Cancer Mortality by Females

Rates of lung cancer mortality from 2002 – 2019 for females are presented in Figure 7-1(B). The lung cancer mortality rate declined for all education quantiles after 2014. The magnitude of decline was generally 4.35 – 25.5%. But the extent of the educational decline was usually larger for males than females.

Table 7.1 shows the mortality rates for each year and percentage change in each education quantile group from 2002 – 2019 for females. The second quantile education quartile group represents the highest decline in mortality until 2002 - 2008. But during 2010 – 2019, 1st quantile group represent the lowest mortality. The least advantageous educational group (4th quantile) displays the lowest decline in mortality, which is only a 4% decline in mortality through 2002 – 2019.

The disparity changes in percentage less than high school education for females is shown in Table 7.2. The absolute disparity in lung cancer mortality shows a decline in disparity in ACI and SII. But RD and BGV values indicate an increase in disparity. In relative disparity, RR, Idisp, MLD, and T, values increase disparity, but RCI and RII are negative, which decreases disparity by 381% and 397%.

7.3 Median Household Income Disparities in Lung Cancer Mortality, 2002 - 2019, Kentucky.

Median household income is categorized into four quantiles, and the first quantile is considered the least advantageous group at the lowest income quantile. And the most advantageous group is represented by the fourth quantile.

Median household income (in ten thousand) five quantiles were determined as first quantile ($01897 – $03298), second quantile ($03299 – $03975), third quantile ($03976 – $04503), and fourth quantile ($04504 -$08632).

7.3.1 Median Household Income Disparities in Lung Cancer Mortality by Males

Rates of lung cancer mortality from 2002 to 2019 for males, by median household income quantiles, are shown in Figure 7-1(C). Lung cancer mortality has declined for all income groups, and the magnitude of the decline was generally between 42% - 20%. Again, the amount of decrease is more prominent for males than females.

Table 7.1 shows the mortality rates for each year and percentage changed in each income quantile group from 2002 – 2019. For example, the 1st quantile group represents the highest lung cancer mortality rate, considered the lowest median household income quantile. Conversely, the most advantageous group (4th quantile) indicates the lowest mortality rate.

Health disparities change by median household income for males is shown in Table 7.3. The absolute disparity in lung cancer mortality shows a decline in disparity in ACI and SII by 54% - 59%. But RD and BGV values suggest increased disparity. In relative disparity, RR, Idisp, MLD, and T, values show an increase in disparity, but RCI and RII show a decline in disparity.

Table 7.3: Medium household income disparity in lung cancer mortality between 2002 - 2019, Kentucky – males & females

	ABSOLUTE MEASURES				RELATIVE MEASURES					
MHIncome - Male	**RD**	**BGV**	**ACI**	**SII**	**RR**	**Idisp**	**MLD**	**T**	**RCI**	**RII**
2002-2004	31.6	113.526	-4.366	-31.286	1.304	12.845	0.004	0.004	-0.04	-0.285
2005-2007	31.9	131.579	-5.621	-40.509	1.34	19.844	0.006	0.006	-0.055	-0.397
2008-2010	41.7	211.949	-7.184	-52.142	1.499	29.222	0.011	0.011	-0.076	-0.555
2011-2014	39.7	205.94	-7.113	-52.023	1.516	32.076	0.012	0.013	-0.082	-0.596
2015-2019	34.9	171.714	-6.765	-49.985	1.582	41.833	0.016	0.017	-0.096	-0.709
% Change 2002-2019	10.443	51.255	-54.932	-59.77	21.253	225.672	282.996	273.599	-141.056	-148.585
MHIncome - Female	**RD**	**BGV**	**ACI**	**SII**	**RR**	**Idisp**	**MLD**	**T**	**RCI**	**RII**
2002-2004	12.9	15.292	-0.798	-5.738	1.244	10.964	0.002	0.002	-0.014	-0.101
2005-2007	17.3	29.418	-1.771	-12.816	1.332	14.139	0.004	0.004	-0.032	-0.229
2008-2010	14.2	22.3	-1.95	-14.214	1.274	15.093	0.003	0.003	-0.035	-0.255
2011-2014	16.8	33.131	-2.489	-18.297	1.331	15.845	0.005	0.005	-0.046	-0.34
2015-2019	20.3	43.852	-3.067	-22.782	1.478	25.176	0.009	0.009	-0.066	-0.488
% Change 2002-2019	57.364	186.775	-284.296	-297.028	18.796	129.627	296.972	308.282	-366.174	-381.618

Absolute measures: RD – Rate Difference, BVG – Between Group Variances, ACI – Absolute Concentration Index, SII – Slope Index Inequality.

Relative Measures: RR – Rate Ratio, Idisp – Index of Disparity, MLD – Mean Log Deviation, T- Theil Index, RCI- Relative Concentration Index, RII- Relative Index of Inequality.

7.3.2 Median Household Income Disparities in Lung Cancer Mortality by Females

Rates of lung cancer mortality from 2002 to 2019 for females, by median household income quantiles, are displayed in Figure 7-1(D). Lung cancer mortality has declined for all income groups, and the magnitude of the decline was generally between 4.5% - 24.5%.

Table 7.1 shows the mortality rates for each year and percentage change in each income group from 2002 – 2019. For example, the 1st quantile group, lowest median household income quantile, has the highest rate of lung cancer mortality for females.

But between 2002 – 2006, the 4th quantile group (most advantaged group) showed the second-highest mortality, then, second quantile and third quantile, respectively. After 2005, this scenario changed, and the 2nd quantile became the second-highest mortality group. But most significantly, mortality rates fluctuated in the 4th and 3rd quantiles between 2005 – 2011. Between 2011 – 2019, the 4th quantile group has the lowest mortality.

The change in median household income disparity for males is shown in Table 7.3. The absolute disparity in lung cancer mortality shows a decline in disparity in ACI and SII by 54% and 59%. But RD and BGV values indicate a positive trend. Rate ratios show a 57%

increase and between-group variance is 186%. In relative disparity, RR, Idisp, MLD, and T values increase disparity, but RCI and RII are negative.

7.4 Poverty Disparities in Lung Cancer Mortality from, 2002 – 2019 in, Kentucky.

The percent of persons whose incomes are below the poverty level are calculated using tables from the Census 2012-2016 ACS data. The percentage of people below the poverty level was separated into four quantiles such as, first quantile (5.98 % - 16.98%) second quantile (16.99% - 20.39%) third quantile (20.40% - 25.37%) and fourth quantile (25.38% - 42.50%). The first quantile is considered as the most advantageous group with the lowest poverty level. The fourth quantile group is the least advantaged group with the highest poverty.

7.4.1 Poverty Disparities in Lung Cancer Mortality by Males

Rates of lung cancer mortality from 2002 to 2019 for males, by the percentage below poverty quantiles, are shown in Figure 7-1(E). Lung cancer mortality has declined for all poverty groups, and the magnitude of the decline was generally between 23% - 42%. Thus, the scale of changes is relatively large for males compared to females.

Table 7.1 shows the mortality rates for each year and percentage change in each poverty group from 2002 – 2019. The 1st quantile group represents the lowest rate of lung cancer mortality, which is considered the most advantageous poverty quantile (5.98% - 16.98%), and the 4th quantile group indicates the highest mortality rate. The 4th quantile group belongs to the highest poverty rate. (25.38% - 42.50%).

Disparity changes in percentage below poverty for males are shown in Table 7.4. The absolute disparity in lung cancer mortality shows a decline in disparity in ACI and SII by around 50%. But RD and BGV values indicate an increase in disparity. In relative disparity, RR, Idisp, MLD, and T values have increases in disparity, but RCI and RII represent a 130% decline in disparity.

Table 7.4: Percentage below poverty disparity in lung cancer mortality between 2002 - 2019, Kentucky – males & females

	ABSOLUTE MEASURES				RELATIVE MEASURES					
Poverty - Male	RD	BGV	ACI	SII	RR	Idlsp	MLD	T	RCI	RII
2002-2004	34	125.408	-4.332	-29.254	1.326	12.86	0.005	0.005	-0.039	-0.267
2005-2007	32.2	121.58	-5.399	-36.54	1.342	18.967	0.005	0.006	-0.053	-0.358
2008-2010	42.2	207.927	-6.281	-42.625	1.497	25.481	0.01	0.011	-0.067	-0.452
2011-2014	38.4	193.115	-6.848	-46.609	1.497	30.099	0.012	0.012	-0.078	-0.533
2015-2019	37.9	167.694	-6.463	-44.146	1.621	37.049	0.015	0.016	-0.091	-0.623
% Change 2002-2019	11.471	33.718	-49.21	-50.903	22.244	188.099	226.242	222.787	-131.189	-133.813
Poverty - Female	RD	BGV	ACI	SII	RR	Idlsp	MLD	T	RCI	RII
2002-2004	15.5	21.767	-0.556	-3.765	1.3	15.504	0.003	0.003	-0.01	-0.066
2005-2007	16	22.223	-1.368	-9.28	1.314	17.255	0.003	0.003	-0.024	-0.166
2008-2010	15.6	24.16	-1.867	-12.7	1.303	16.57	0.004	0.004	-0.034	-0.228
2011-2014	20.5	41.979	-2.562	-17.477	1.407	17.526	0.006	0.007	-0.048	-0.324
2015-2019	22.4	49.854	-2.859	-19.571	1.522	23.465	0.01	0.01	-0.061	-0.418
% Change 2002-2019	44.516	129.03	-413.861	-419.765	17.053	51.352	200.877	216.875	-521.947	-529.093

Absolute measures: RD – Rate Difference, BVG – Between Group Variances, ACI – Absolute Concentration Index, SII – Slope Index Inequality.

Relative Measures: RR – Rate Ratio, Idisp – Index of Disparity, MLD – Mean Log Deviation, T- Theil Index, RCI- Relative Concentration Index, RII- Relative Index of Inequality.

7.4.2 Poverty Disparities in Lung Cancer Mortality by Females

Rates of lung cancer mortality from 2002 to 2019 for females, by the percentage below poverty quantiles, are shown in Figure 7-1(F). Lung cancer mortality has declined for all poverty groups, and the magnitude of the decline was generally between 2.6% - 24.56%.

Table 7.1 shows the mortality rates for each year and percentage change in each poverty group from 2002 – 2019. For example, the 2nd quantile group represents the lowest rate of lung cancer mortality between 2002 – 2019, which is the second most advantages poverty quantile. Most advantage groups (1st quantile) represent the second-lowest mortality rate between 2003-2014.

The fourth quantile group indicates the highest mortality rate between 2002 - 2019. The fourth quantile group belongs to the least advantaged group (25.38% - 42.50%).

The disparity measure in percentage below poverty disparity for females is shown in Table 7.4. The absolute disparity in lung cancer mortality shows a 400% decline in disparity in ACI and SII. But RD and BGV values indicate an increase in disparity. In relative disparity, RR, Idisp, MLD, and T values show an increase in disparity, but RCI and RII show a 500% decline.

7.5 Unemployment Disparities in Lung Cancer Mortality from, 2002-2019 in, Kentucky.

The percent of people ages 16 and over who are unemployed is calculated using the Census 2012-2016 ACS data. Based on the amount of the county population unemployed four quantiles was defined as, first quantile (4.10% - 6.74%) second quantile (6.75% - 7.90%) third quantile (7.91% - 10.17%) and fourth quantile (10.18% - 18.60%). The first quantile is the most advantageous group with the lowest unemployment, and the 4^{th} quantile group is considered the least advantaged group.

7.5.1 Unemployment Disparities in Lung Cancer Mortality by Males

Rates of lung cancer mortality from 2002 to 2019 for males, by the percentage of unemployment quantiles, are shown in Figure 7-1(G). Lung cancer mortality has declined for all unemployment groups, and the magnitude of the decline was generally between 24.3% - 42.37%.

Table 7.1 shows the mortality rates for each year and percentage changed in each unemployment group from 2002 – 2019. The 1^{st} quantile group represents the lowest rate of lung cancer mortality, which is considered the most advantageous group (4.10% - 6.74%), and the 4^{th} quantile group indicates the highest mortality rate. In addition, the 4^{th} quantile group belongs to the highest unemployment rate (10.18% - 18.95%).

The disparity measure in percentage of unemployment disparity for males is shown in Table 7.5. The absolute disparity in lung cancer mortality shows a 60% decline in disparity in ACI and SII. But RD and BGV values that indicates an increase in disparity. In relative disparity, RR, Idisp, MLD, and T values show an increase in disparity, but RCI and RII indicate a 150% decline in disparity.

Table 7.5: Unemployment disparity in lung cancer mortality between 2002 - 2019, Kentucky – males & female

	ABSOLUTE MEASURES				RELATIVE MEASURES					
Unemployement - Male	RD	BGV	ACI	SII	RR	Idisp	MLD	T	RCI	RII
2002-2004	25.1	76.928	-3.814	-25.012	1.241	10.131	0.003	0.003	-0.035	-0.228
2005-2007	25.6	82.62	-4.597	-30.185	1.27	15.065	0.004	0.004	-0.045	-0.295
2008-2010	35.6	164.875	-6.618	-43.527	1.428	25.632	0.009	0.009	-0.07	-0.461
2011-2014	35	156.581	-6.353	-41.863	1.454	26.459	0.009	0.01	-0.073	-0.478
2015-2019	32.9	136.839	-6.096	-40.279	1.547	34.332	0.013	0.013	-0.086	-0.569
% Change 2002-2019	31.076	77.878	-59.808	-61.037	24.726	238.876	332.601	329.379	-147.975	-149.881
Unemployement - Female	RD	BGV	ACI	SII	RR	Idisp	MLD	T	RCI	RII
2002-2004	12.4	16.216	-0.571	-3.752	1.236	12.041	0.002	0.002	-0.01	-0.066
2005-2007	14.4	22.253	-2.137	-14.066	1.277	13.077	0.003	0.003	-0.038	-0.252
2008-2010	11	16.396	-1.616	-10.653	1.211	13.308	0.003	0.003	-0.029	-0.191
2011-2014	17.3	31.841	-2.878	-19.004	1.356	20.027	0.005	0.005	-0.053	-0.353
2015-2019	18.1	36.877	-3.1	-20.534	1.437	25.362	0.008	0.008	-0.066	-0.439
% Change 2002-2019	45.968	127.418	-442.93	-447.226	16.302	110.641	224.421	228.992	-558.725	-563.938

Absolute measures: RD – Rate Difference, BVG – Between Group Variances, ACI – Absolute Concentration Index, SII – Slope Index Inequality.

Relative Measures: RR – Rate Ratio, Idisp – Index of Disparity, MLD – Mean Log Deviation, T- Theil Index, RCI- Relative Concentration Index, RII- Relative Index of Inequality.

7.5.2 Unemployment Disparities in Lung Cancer Mortality by Females

Rates of lung cancer mortality from 2002 to 2019 for females, by the percentage of unemployment quantiles, are shown in Figure 7-1(H). Lung cancer mortality has declined for all unemployment groups, and the magnitude of the decline was generally between 8.4% - 28.1%. The amount of decrease is considerably more significant for men compared to women.

Table 7.1 shows the mortality rates for each year and percentage changed in each unemployment group from 2002 – 2019. The 4th quantile group indicates the highest mortality rate, and the 4th quantile group belongs to the highest unemployment rate (10.18% - 18.95%).

The 1st quantile and 2nd quantile groups represent the highest mortality rate between 2002-2005. The 1st and 2nd quantile belong to the most advantaged group. But after 2005 the 3rd quantile became the second-highest mortality group. Between 2005 -2019, the 1st and 2nd quantile represent the lowest mortality rate.

Disparity measures in the percentage of unemployment for females are shown in Table 7.5. The absolute disparity in lung cancer mortality shows a 450% decline in disparity in ACI and SII. But RD and BGV values and that indicates an increase in disparity. RR, Idisp, MLD, and T values increase the disparity in relative disparity, but RCI and RII are negative.

7.6 Appalachian and Non-Appalachian Disparities in Lung Cancer Mortality from, 2002-2019 in, Kentucky.

This section obtained male and female mortality estimates for two geographic regions, including the Appalachian and Not Appalachian regions.

7.6.1 Appalachian and Non -Appalachian Disparities in Lung Cancer Mortality for Males

Rates of lung cancer mortality from 2002 to 2019 for males, by Appalachian and not Appalachian quantiles, are shown in Figure 7-1(I). Lung cancer mortality has declined for two geographic regions. The magnitude of the decline for males in the Appalachian region was 28.48%, and for the non-Appalachian area, it was 37.91%.

Table 7.1 shows the mortality rates for each year and percentage change in each geographic region from 2002 – 2019. Again, the non-Appalachian region represents the lowest lung cancer mortality rate, and the Appalachian region shows the highest mortality rate for males.

The change in the Appalachian region and non-Appalachian region disparity for males is shown in Table 7.6. The absolute disparity in lung cancer mortality shows an increase in RD and BGV values. Rate ratios show a 16% increase, and the between-group variance is 30%. In relative disparity, RR, Idisp, MLD, and T values also appear to increase the disparity in both geographic regions.

Table 7.6: Geographic regional disparity in lung cancer mortality between 2002 - 2019, Kentucky – males & females

	ABSOLUTE MEASURES				RELATIVE MEASURES					
Applachian or not - Male	**RD**	**BGV**	**ACI**	**SII**	**RR**	**Idisp**	**MLD**	**T**	**RCI**	**RII**
2002-2004	21.6	94.981			1.209	20.89	0.004	0.004		
2005-2007	18.8	71.394			1.194	19.381	0.003	0.003		
2008-2010	24	115.11			1.274	27.366	0.006	0.006		
2011-2014	27.6	150.221			1.345	34.5	0.009	0.009		
2015-2019	25.2	123.107			1.393	39.252	0.011	0.012		
% Change 2002-2019	16.667	29.613			15.19	87.902	196.856	202.879		

	ABSOLUTE MEASURES				RELATIVE MEASURES					
Applachian or not- Female	**RD**	**BGV**	**ACI**	**SII**	**RR**	**Idisp**	**MLD**	**T**	**RCI**	**RII**
2002-2004	5	5.065			1.091	9.058	0.001	0.001		
2005-2007	6.5	8.476			1.12	12.037	0.001	0.001		
2008-2010	9	16.062			1.169	16.917	0.002	0.003		
2011-2014	10.2	20.381			1.199	19.922	0.003	0.003		
2015-2019	12.5	30.138			1.287	28.736	0.006	0.007		
% Change 2002-2019	150	495.018			18.043	217.241	729.272	749.324		

Absolute measures: RD – Rate Difference, BVG – Between Group Variances, ACI – Absolute Concentration Index, SII – Slope Index Inequality.

Relative Measures: RR – Rate Ratio, Idisp – Index of Disparity, MLD – Mean Log Deviation, T- Theil Index, RCI- Relative Concentration Index, RII- Relative Index of Inequality.

7.6.2 Appalachian and Non-Appalachian Disparities in Lung Cancer Mortality by Females

Rates of lung cancer mortality from 2002 to 2019 for females, by Appalachian and not Appalachian quantiles, are shown in Figure 7-1(J). Lung cancer mortality has declined for two geographic regions. The magnitude of the decline for females in the Appalachian region was 6.9% and for the non-Appalachian region, it was 21.1%. For the Appalachian region, the female lung cancer mortality rate increased between 2002 – 2008, and it started to decline after 2010.

Table 7.1 shows the mortality rates for each year and percentage change in each geographic region from 2002 – 2019. The non – Appalachian region represents the lowest rate of lung cancer mortality. The Appalachian region shows the highest mortality rate for females.

Disparity measures in the Appalachian region and non-Appalachian region disparity for females are shown in Table 7.6. The absolute disparity in lung cancer mortality shows an increase in RD and BGV values. Rate difference indicates disparity increase in 150% and between-group variance is 495%.

In relative disparity, RR, Idisp, MLD, T values show an increase in disparity. The rate ratio demonstrates an 18% increase, the index of disparity is 217%, and the mean log deviation and the index increase by 730%.

7.7 Change in Socioeconomic and Geographic Disparity

7.7.1 Mortality Trends

Table 7.1 presents age-adjusted mortality rates and population distribution for the first and last years of observation by gender, area-socioeconomic quantile, and geographic regions, as well as the percent change from 2002 to 2019.

To give a complete picture of changes in lung cancer mortality, Figure 7-1 shows the trends in the age-adjusted mortality of lung cancer, by gender, for social characterized by socioeconomic quantile and geographic regions.

Among males, mortality generally declined for all socioeconomic variables and geographic regions. But the magnitude of the decline was considerably more significant for men compared to women. The picture was more mixed among females, with rates decreasing among some groups and increasing among others. In addition, there is considerably more variation in mortality visible in females.

7.7.2 Change in Socioeconomic Disparity

Table 7.2 to 7.6 shows the male and female absolute and relative disparity change for each socioeconomic variable from 2002 to 2019. Generally, all the relative and absolute measures of disparity suggest that socioeconomic inequality for all the variables in lung cancer mortality increases among males and females. However, the magnitude of the increase differed widely across disparity measures. They range from a 10% increase in the rate difference to a 479.13% percent increase in the Theil index.

In addition, the four measures of disparity are sensitive to the direction of the gradient (RCI/ACI and RII/SII) suggest that disparity for all the socioeconomic variables in lung cancer mortality decreased among males and females. This is a clear example of the importance of selecting a disparity measure based on appropriate standards. However, the empirical result cannot notify the reader about which measure is "right." Any substantive conclusion is therefore entirely dependent on which measure is chosen.

In this case, the value position rests on whether disparity measures should be weighted by population size. Population-weighted methods allow for incorporating information about the size of the social group by weighting, which measures the relationship between a group's health and its relative socioeconomic rank. Where population-weighted, regression-based methods differ from un-weighted methods, they enable us to incorporate information about the size of the social group by weighting. These measures are interpreted as the effect on health moving from the lowest to the highest socioeconomic group (Table 7.7).

The study unweighted disparity measures (RR, IDisp, and RD) would generally suggest that the socioeconomic disparity increased during 2002 – 2019. On the other hand, population-weighted disparity measures (RCI/ACI and RII/SII) indicate an improvement a moderate decrease in relative disparity and absolute disparity.

For example, the disparity of income by males for 2002 – 2019 is RCI -141.056, ACI -54.932, RII -148.585, and SII -59.77 9 (Table 7.8). The negative disparity value suggests that the disparity in lung cancer mortality favors better income. One of the explanations for the ACI and RCI (and, by extension, the SII/RII indices) are preferred by some researchers is that they "reflect the socioeconomic dimension to inequalities in health" (Wagstaff, Paci, & van Doorslaer, 1991). A downward health gradient (health degrades with increasing social-group rank) results in a positive index, whereas an upward health gradient results in a negative index.

Table 7.7: Characteristics of health disparity measures

Disparity Measure	Absolute or Relative	Reference Group	All Social Groups	Reflect SES Gradient	Social Group Weighting
Rate difference (RD)	Absolute	Best	No	Yes	No
Between group variance (BGV)	Absolute	Average	Yes	No	Yes
Absolute concentration index (ACI)	Absolute	Average	Yes	Yes	Yes
Slope index of inequality (SII)	Absolute	Average	Yes	Yes	Yes
Relative difference (RR)	Relative	Best	No	Yes	No
Index of disparity (Idisp)	Relative	Best	Yes	No	No
Mean log deviation (MLD)	Relative	Average	Yes	No	Yes
Theil index (T)	Relative	Average	Yes	No	Yes
Relative concentration index (RCI)	Relative	Average	Yes	Yes	Yes
Relative index of inequality (RII)	Relative	Average	Yes	Yes	Yes

7.7.3 Changes in Geographic Disparity

The male and female disparity between different geographic regions showed agreement with the direction of change for the Appalachian and non-Appalachian regions. Furthermore, measures of absolute and relative disparity for males and females increased during 2002 – 2019. Therefore, answering whether male and female disparity in the

Appalachian and non-Appalachian regions is straightforward. But the magnitude of disparity for females is significantly larger than for males.

But even among the magnitude of the measures of relative and absolute disparity, there was disagreement, with the index of disparity suggesting that Appalachian region disparities increased by 87% among males. Still, the measurements of Theil index and mean log deviation indicate an increase of 202% and 197%, respectively (Table 7.8).

7.8 Conclusion

This chapter presents the results of five separate analyses in lung cancer mortality trends in selected socioeconomic and geographic health disparities, which empirically compared various summary measures of health disparities. In addition, the study included assessments of socioeconomic and geographic disparities in lung cancer mortality. These analyses aimed to examine the consistency of different measures of health disparity across a range of lung cancer-related outcomes.

Summaries of selected results are shown in Table 7.8. The numbers in the table represent percentage changes in the values of the disparity measure between 2002 -2019. Blue color means the disparity has increased by more than 30%. Pink indicates the disparity increased between 10-29%, gray means an increase of changes less than 10%, yellow means a decrease of change less than 10%, dark green indicates declines in the disparity of 10-29% and light green means that disparity has declined by more than 30%.

Overall, these graphical examples emphasize the conclusion that the way in which disparity is measured matters. For instance, for education, income, poverty, and unemployment disparity in lung cancer mortality, no conclusion can be drawn about whether disparity got better or worse between 2002 and 2019. There are pink, blue, or green cells indicating increases and decreases depending on which measure is used.

Therefore, the best way to determine socioeconomic disparity trend in lung cancer mortality is to define whether disparity should be measured relative or absolute. But this is not a problem of the disparity in geographic regions because it indicates an increase in lung cancer disparity in Appalachian and non-Appalachian areas. Where reasonably concluded that geographic disparity increased regardless of which measure was used (Table 7.8).

Table 7.8: Graphical summary of selected disparity trends.

		ABSOLUTE MEASURES					RELATIVE MEASURES				
Education	RD	BGV	ACI	SII		RR	Idisp	MLD	T	RCI	RII
Male	12.5	78.358	-49.6	-54.327		22.104	208.76	347	338.1	-132.7	-140
Female	63.415	301.48	-296.91	-309.669		19.984	139.22	469.2	479.1	-382	-397.4
MHIncome											
Male	10.443	51.255	-54.932	-59.77		21.253	225.67	283	273.6	-141.1	-148.6
Female	57.364	186.78	-284.296	-297.028		18.796	129.63	297	308.3	-366.2	-381.6
Poverty											
Male	11.471	33.718	-49.21	-50.903		22.244	188.1	226.2	222.8	-131.2	-133.8
Female	44.516	129.03	-413.861	-419.765		17.053	51.352	200.9	216.9	-521.9	-529.1
Unemployement											
Male	31.076	77.878	-59.808	-61.037		24.726	238.88	332.6	329.4	-148	-149.9
Female	45.968	127.42	-442.93	-447.226		16.302	110.64	224.4	229	-558.7	-563.9
Applachian or not											
Male	16.667	29.613				15.19	87.902	196.9	202.9		
Female	150	495.02				18.043	217.24	729.3	749.3		

≥30%	11% to 29%	10% to 0	0 to (-)10%	(-)11% to (-)29%	≤(−)30

Absolute measures: RD – Rate Difference, BVG – Between Group Variances, ACI – Absolute Concentration Index, SII – Slope Index Inequality.

Relative Measures: RR – Rate Ratio, Idisp – Index of Disparity, MLD – Mean Log Deviation, T- Theil Index, RCI- Relative Concentration Index, RII- Relative Index of Inequality.

Most of the cases of disagreement between measures of disparity differed on two issues. One is the scale on which disparity should be evaluated. In many situations, relative disparity measures show an increase or decline in disparity, while absolute estimates show the opposite of relative measures. For example, income disparity in lung cancer mortality among males (BVG = 51.2 and RCI = -141.1) is getting better or worse depending on whether the disparity section is absolute or relative. Therefore, specifying whether absolute or relative disparities are more critical before undertaking any analyses will minimize disagreement about disparity trends.

The second source of disagreement among disparity methods was whether they weighed social groups by population size. Several researchers found that population-weighted disparity measures differed in either magnitude or direction from unweighted disparity measures (Harper et al., 2008a). In particular, and as might be expected, unweighted measures of disparity appear to be more sensitive to the movement of disease rates, especially those of smaller population groups whose disease rates may be less stable over time.

The differences observed in this study are from different conceptions of disparity on which these measures are based. Thus, our results suggest that attempts to evaluate trends in health disparities require judgments about which conception of disparity is essential for the question at hand.

Therefore, choices about the appropriate reference point from which to measure disparity, whether disparity should be measured in absolute or relative terms, whether to weight social groups according to the fraction of the population they represent, and whether to place additional weight on subgroups of interest (e.g., the poor or the least healthy) should be explicit when assessing health disparity change.

Chapter 8

Conclusion

8.1 Introduction

This chapter includes a discussion of the results of this dissertation, contributions to literature, limitations of the research, areas for future research, and a conclusion. The discussion and findings are organized under the four broad questions this research sought to answer:

1. What are the lung cancer incidence and mortality trends in Kentucky? And what are the disparities in male and female lung cancer trends?

2. Is there any statistical significance between lung cancer mortality and socio-economic factors across Kentucky? And what are the geographic patterns of lung cancer mortality in the Appalachian region versus the non-Appalachian region?

3. What is the spatial variation analysis of lung cancer in Kentucky? Are there any interactive effects on lung cancer risk factors? What are the highest lung cancer

risk areas? What type of risk factors are mainly responsible for lung cancer? And what are their relative importance? Do lung cancer risk factors interact or lead to disease independently?

4. Is there lung cancer disparity across Kentucky?

This chapter outlines the significant findings and discusses how the question was addressed in this research for each research question. Throughout the discussion, the study findings are discussed that support or counter the available literature in this field and the conceptual model discussed in the methodology chapter. In addition, this chapter elaborates on the broader implications of these findings of lung cancer mortality and risk factors and how results from this research contribute to the general understanding of lung cancer epidemiology. Finally, the chapter also outlines future areas of research in this field that could provide additional findings and answer questions that were not addressed by this research.

8.2 Lung Cancer Incidence and Mortality Trends in Kentucky

The first section of this dissertation aims to understand lung cancer incidence and mortality trends in Kentucky. Our results indicate that men's lung cancer incidence and mortality peaked at the beginning of the 20th century and decreased until 2016. However, the cancer mortality rate rose during most of the 20th century, mainly because of the tobacco epidemic's rapid increase in lung cancer deaths among men (Siegel et al., 2021). According to the literature among adults in 2000, smoking prevalence was highest in Kentucky (30.5%), Nevada (29.1%), and Missouri (27.2%); prevalence was lowest in Utah (12.9%), Puerto Rico (13.1%), and California (17.2%). In 2000, 31.3% of men and 21.3% of women used tobacco in any form (Giovino, 2002). This increasing trend appears at the beginning of a lung cancer incidence and mortality during 2000 -2002 in Kentucky for men and women. For example, the highest lung cancer incidence and mortality for men were present during 2000 -2002.

Also, previous studies suggest very high lung cancer incidence in several southeastern Kentucky counties could be related to coal-mining activity during 1996–2006 (W. Jay Christian et al., 2011). This increasing trend appears in the male and female incidence trend during 2000-2008.

According to epidemiology, the lung cancer epidemic is associated with tobacco use because of the continued decline in the prevalence of smokers in recent decades. However, declines in smoking and improvements in early detection and treatment have resulted in a continuous reduction between 2014 – 2016 in the cancer incidence and death rate (Siegel et al., 2020). This declining trend appears in Kentucky's lung cancer incidence and mortality during 2014-2016 for men and women. For example, between 2014-2016, lung cancer mortality and incidence declined around 4.5% for men in Kentucky.

A previous study evaluated lung cancer incidence and survival according to cancer subtype, sex, and trends in incidence-based mortality. Results reveal that Among men, incidence-based mortality from Non-Small Cell Lung Cancer (NSCLC) decreased 6.3% annually from 2013 through 2016, whereas the incidence decreased 3.1% annually from 2008 through 2016 (Howlader et al., 2020). This declining trend also appears in Kentucky, suggesting the highest 4.5% mortality decline during 2014-2016 among men.

Also, population-level mortality from Non-Small Cell Lung Cancer (NSCLC) in the United States fell sharply from 2013 to 2016, and survival after diagnosis improved substantially (Howlader et al., 2020). Our analysis suggests that a reduction in incidence and treatment advances, particularly approvals for and use of targeted therapies, is likely to explain the decline in mortality observed during this period.

Although in the U.S., smoking rates have historically been lower among women than men, smoking rates have not declined as quickly for women as for men. For example, since 2005, smoking rates among women have reduced by 25.4 % compared with a 26.8 % decline among men. Additionally, smoking rates among women have dropped by about 59% since 1965, compared with a 66 % drop among men (truth initiative, 2019). This trend can be found in the women lung cancer incidence and mortality in Kentucky. Because of these female lung cancer incidence and mortality trends, declining rates were considerably lower compared to the men.

The Institute of Medicine and others have found that smoke-free ordinances help to reduce lung cancer. For these reasons, smoke-free laws inarguably benefit public health. Despite the persistence of high smoking rates throughout the state, many local communities in Kentucky have enacted smoke-free regulations that prohibit smoking in workplaces and enclosed buildings open to the public. Smoke-free ordinance covered 23 counties (of 120) starting in December 2009. The effect of these regulations can be found in lung cancer incidence and mortality trends. The research study explores the impact of local smoke-free ordinances on Kentucky's smoking prevalence, revealing that smoking prevalence was approximately 5% lower in counties with smoke-free laws (W. Jay Christian, Walker, Huang, & Hahn, 2019).

Further reductions in the lung cancer burden will require continued efforts to develop, deliver, and surveil effective cancer prevention, early screening, and treatment strategies.

8.2.1 Limitation

However, this study has some limitations. First, the underlying relationship cannot be established because Joinpoint regression consists of trend analysis in incidence and mortality. Therefore, study results require further confirmation with individual-level data.

8.2.2 Implication

The significant contribution of this research is adopting the trend analysis measure at the county level. Since it was proposed by Guraga (1997) existing research primarily focused on the trends in a specific county or region (Guarga et al., 2021). But very few studies have focused on specific geographic areas. Research analyzing lung cancer incidence and mortality trends in Kentucky is lacking in the current literature. This study results on incidence and mortality trends in lung cancer are similar to those observed in Spain (Izarzugaza et al., 2010) and other European countries such as Austria, France, Iceland, Italy, the Netherlands, and Switzerland, that argued that lung cancer incidence rates reflect the progress in smoking cessation, first observed in men and then also in women (Lortet-Tieulent et al., 2014).

Second, looking at lung cancer incidence and mortality rates over time is essential. Epidemiologists can track changes in the risk of developing and dying from specific cancers and information about survival chances and forecasts. Researchers also show how the trend analysis results can be adopted in developing lung cancer prevention programs (Dela Cruz, Tanoue, & Matthay, 2011).

Third, the results presented in this research could offer references to governments, policymakers, health professionals, and researchers to understand the impact of lung cancer on the population. Also, there is a need for help to develop strategies to address lung cancer challenges. Finally, statistical trends analysis is also crucial for measuring the success of efforts to control and manage cancer. Therefore, this study contributes to empirical applications for cancer epidemiology like Didkowska et al (2016) (Didkowska, Wojciechowska, Mańczuk, & Łobaszewski, 2016).

8.2 Associations of Lung Cancer Mortality with Socio-Economic Factors

Chapter five found statistical significance between lung cancer mortality and socio-economic factors across Kentucky and geographic patterns of lung cancer mortality in the Appalachian region versus the non-Appalachian area.

Appalachian counties showed the strongest statistical association between lung cancer mortality rates with median household income and high school graduation rate, which may explain higher lung cancer mortality. A study conducted in Kentucky also indicates that lung cancer incidence and mortality are higher when socioeconomic factors such as education and income are low (Berlia, 2016). According to the literature, an analysis including 16 European populations reported higher lung cancer mortality rates in groups with the lowest educational attainment (Van der Heyden et al., 2009). Also, a study conducted in Sweden found an association between household disposable income and lung cancer survival(Sachs, Jackson, & Sartipy, 2020).

Education level can influence occupation, income, adherence to healthy behaviors, and participation in health promotion and screening programs. This finding indicates that Kentucky's emphasis on improving graduation rates may reduce lung cancer mortality and increase personal income and address health disparities between Appalachian counties. Using the Institute of Medicine's 2007 report data, a study conducted in Kentucky reveals that high school graduation rates showed the strongest statistical association with lung cancer mortality (Gross, 2010). This result indicates that continued improvements in Kentucky's diploma attainment rate may contribute to future reductions in lung cancer mortality statewide.

In the non-Appalachian region, adult smoking rates showed the strongest association with lung cancer mortality. In addition, statewide adult smoking rate, high school graduation rate, median household income, and the number of coal mine employment showed a strong relationship with lung cancer mortality rates. Thus, according to the literature, smoking undoubtedly contributes more than any other factor to the high rates of lung cancer found throughout the state of Kentucky (Hopenhayn et al., 2003). Therefore, reducing smoking is essential for individual health, and reducing secondhand smoking is also necessary.

Recent research has suggested that coalmine employment increases lung cancer risk (W. J. Christian, B. Huang, J. Rinehart, & C. Hopenhayn, 2011). Study results found that the number of coal mine employment contributes to increasing lung cancer risk in all counties in Kentucky.

In conclusion, Kentucky areas with low education and income have the highest smoking and lung cancer levels. Prevention efforts should be focused on these areas since the counties have some of the highest smoking rates in the country and contribute significantly to the overall smoking and lung cancer rate for Kentucky. Kentucky has a long way to go to address these significant health issues, and one of the first steps could be raising policies and programs to reduce tobacco use and implementing strict smoke-free laws.

8.2.1 Limitation

A limitation of this study was the inability to determine and examine all social determinants of health that contribute to higher mortality rates and poorer health outcomes for Kentuckians. There are many background variables like public policy, family history of smoking, and greater prevalence of tobacco farming and marketing that could contribute to the higher mortality rates in Appalachia instead of non-Appalachia.

Although the fundamental goal of public health research is to thoroughly understand the interaction between cancer and physical and socioeconomic conditions, this study focused on only a few socioeconomic variables. Therefore, to understand the relationship between lung cancer and socioeconomic variables, it is possible to apply a framework to analyze different years instead of one epidemic year. By doing so, the study could predict the spatial-temporal pattern of lung cancer and socioeconomic relationship.

Furthermore, linear regression cannot handle non-linear relationships. Therefore, certain transformations will be necessary if any non-linearity in the variables is identified. However, the study did not examine any non-linearity in this research.

8.2.2 Implication

This dissertation has several important implications for public policy. First, the study results identify the socioeconomic variables responsible for Kentucky's disproportionate lung cancer impacts. This information can assist local advocacy groups, and government organizations develop programs to reduce income inequalities and increase educational attainment in specific geographic regions. Second, the GWR methodology used in this study provided detailed information about locations that are disproportionately impacted by socioeconomic factors.

Education and income are generally associated with lung cancer mortality, but differences in the strength and direction of these associations exist depending on geographic location. Therefore, improving high school graduation rates and household income in Appalachia could result in a meaningful long-term reduction in lung cancer mortality. This research supports the conclusion made by Castro et al (2021) which suggested a decline in lung cancer screening rates among patients with lower income and education (Castro et al., 2021). Also, a study done by Hovanec et al (2018) found a constant relationship between socioeconomic status and cancer (Hovanec et al., 2018).

Conversely, the relative importance of adult smoking to lung cancer outcomes was greater outside the Appalachian regions. According to Schoenberg et al., (2015) lower educational attainment is a robust and independent predictor of smoking in Kentucky (Schoenberg, Huang, Seshadri, & Tucker, 2015). Given this, equitable investment in public education might be considered an “upstream” strategy for reducing the prevalence of tobacco use. However, public education in Kentucky has historically been underfunded because of its ties to local property taxes (Cardarelli et al., 2021). This is particularly true for schools in lower-resourced communities, including Appalachian, Kentucky. School districts in the lowest quintile of funding are concentrated mainly in Appalachian Kentucky (Wewers et al., 2000).

It would be interesting to see whether Kentucky's smoke-free policy and indoor smoking bans significantly affect smoking and lung cancer levels. As of 2016, a few counties and cities across Kentucky have implemented indoor smoking bans, and some that have implemented them have significant exemptions attached. However, a comprehensive smoke-free policy is essential for Kentucky because it has one of the country's highest rates of smoking and lung cancer. Legislators and policymakers should consider it because, in the long run, such a ban will save a tremendous amount of money that goes towards health care spending, increase worker productivity, and decrease the burden on Medicare and Medicaid. Also, reducing miners’ exposure to respirable airborne contaminants minimizes the risk of developing lung disease. Historically, such policies and practices have not been widely implemented in rural communities (York et al., 2010).

8.3 Influence of Lung Cancer Risk Factors.

The third section of this study investigated the spatial distribution pattern of Kentucky's lung cancer mortality rate. Results showed that the lung cancer mortality rate is heterogeneous in Kentucky. It is highly autocorrelated in space; a large quantity of the counties with high mortality rates are distributed in the eastern region of Kentucky.

Furthermore, using the new Geographical Detector method, the study examined the potential determinants of the lung cancer mortality rate. This study’s findings suggest that adult smoking, median household income, unemployment, number of coal mine employment, and physically unhealthy days played a much more significant role in increasing lung cancer mortality rate than other studied factors. The combined effects of pairs of factors are also described and can be compared with their separate impacts. What is remarkable is the interactive effect between different factors. Since all the interactive effects influenced the values of the Power of Determinant, combinations of the studied factors will be more efficient at explaining the spatial variability of the lung cancer mortality rate compared with different factors.

Most existing research have investigated the independent effects of various factors on lung cancer disease; however, the causes of lung cancer mortality are complex. This study explained that the Geographical Detector technique could measure the different effects of two or even more factors on the lung cancer mortality rate and the interactive impact between various determinants.

8.3.1 Implication

This study has implications for future research. Firstly, existing research primarily focused on the independent effects of various risk factors on lung cancer mortality; insufficient attention was paid to the interactive effect between risk factors (Ghasemi, Mahaki, Dreassi, & Aghamohammadi, 2020). As for two risk factors, this study examined their influences and understand their interactive effects.

Findings from this research support conclusions made by He et al., (2013) which suggested that smoking status and radon exposure have significant interactive effects on lung cancer (He et al., 2013; Ridge et al., 2013). Additionally, adult smoking and high school graduation, high school graduation with medium household income, high school graduation rate, and unemployment increased the interactive effect of lung cancer mortality. Previous research reveals that people with a high school education smoke cigarettes for more than twice as many years as people with at least a bachelor's degree (Siahpush et al., 2010). Also, people in the most socioeconomically deprived groups, such as low income and educational attainment, have higher lung cancer risk than those in the most affluent groups (Clegg et al., 2009).

This study shows that the differences between the high school graduation rate and household medium income factors statistically significantly influence lung cancer mortality. In addition, a study was done by Sidorchuk et al (2009) also found that lung cancer incidence was associated with low education, occupational, and income (Sidorchuk et al., 2009).

Secondly, the results presented in this research could offer a reference for understanding the spatial distribution patterns and epidemiological characteristics of the lung cancer mortality rate. Finally, implications from this study provide clues for policymakers to develop strategies to prevent and control lung cancer. For example, high priority should be paid to regions with high adult smoking, median household income, and education.

8.3.2 Limitation

One limitation of this study is the discretization of quantitative data. The Geographical Detector method requires a discretization of the impact factors before they are input into the model. For qualitative data, it is easy to obtain their classifications according to their categorical attributes. The study used multiple sorting techniques, such as the Jenks Natural Breaks classification method and Geometric interval method. However, sorting methods tend to be subjective; therefore, variable discretization using these methods may weaken the Geographic Detector's ability to characterize the actual associations between lung cancer mortality rate and risk factors. The problem of how to discretize quantitative data effectively should be considered in future studies.

8.4 Health Disparity in Lung Cancer

Chapter 7 presents the results of five separate analyses in lung cancer mortality trends in selected socioeconomic and geographic health disparities, which empirically compared various summary measures of health disparities. In addition, the study included assessments of socioeconomic and geographic differences in lung cancer mortality. These analyses aimed to examine the consistency of different measures of health disparity across a range of lung cancer-related outcomes.

Among males, mortality generally declined for all socioeconomic variables and geographic regions. But the magnitude of the decline was considerably more significant for men compared to women. The picture was more mixed among females, with rates decreasing among some groups and increasing among others. In addition, there is considerably more variation in mortality visible in females.

According to existing literature, 50% of women diagnosed with lung cancer worldwide are never-smokers compared with only 15–20% of men, and these proportions have been rising in both genders (MacRosty & Rivera, 2020). In addition, a study conducted in the US found that among never-smokers, women were at higher risk of developing lung cancer than men (Wakelee et al., 2007). Therefore, a higher rate of lung cancer among never-smoking women compared with men is a crucial driver of the changing lung cancer demographics cancer worldwide (Fidler-Benaoudia, Torre, Bray, Ferlay, & Jemal, 2020).

The study result shows the absolute and relative disparity change for each socioeconomic variable from 2002 to 2019. Generally, all the relative and absolute measures of disparity suggest that socioeconomic inequality for all the variables in lung cancer mortality increases among males and females. However, the magnitude of the increase differed widely across disparity measures. In addition, the four measures of disparity are sensitive to the direction of the gradient (RCI/ACI and RII/SII) suggest that disparity for all the socioeconomic variables in lung cancer mortality decreased among males and females. This is a clear example of the importance of selecting a disparity measure based on appropriate standards. However, the empirical result cannot notify the reader about which measure is "right." Any substantive conclusion is therefore entirely dependent on which measure is chosen.

In this case, the value position rests on whether disparity measures should be weighted by population size. Population-weighted methods allow for incorporating information about the size of the social group by weighting, which measures the relationship between a group's health and its relative socioeconomic rank. Where population-weighted, regression-based methods differ from un-weighted methods, they enable us to incorporate information about the size of the social group by weighting. These measures are interpreted as the effect on health moving from the lowest to the highest socioeconomic group.

8.4.1 Implications

There is currently a strong emphasis in the US public health policymaking community on monitoring progress toward eliminating health disparities. This is one of the first studies to examine socioeconomic disparities on lung cancer mortality in Kentucky using multiple disparity metrics on both the absolute and relative scales. Findings from this research support Caposole et al (2014) which suggested the importance of eliminating socioeconomic and racial disparities in lung cancer (Caposole, Miller, Kim, Steward, & Bauer, 2014).

Findings from this research support assumptions made by Elkbuli et al (2020) that individuals of higher socioeconomic status experienced higher survivorship than those of lower socioeconomic status. Interventions aimed at public education and access to high-quality healthcare are needed to improve socioeconomic and gender-based disparities in lung cancer survivorship (Elkbuli et al., 2020).

However, the results of the case studies presented in this dissertation demonstrate that it is possible to come to fundamentally different conclusions about the extent of progress toward eliminating health disparities using the same data but various measures of health disparity. For example, the study done by Harper et al (2008) suggested that result summary measures can confuse policymakers and researchers about whether disparities are increasing or decreasing. This confusion will be minimized, and health disparity measurement will be advanced by increased debate and discussion of the issues that generate differences among measures of health disparity (Harper et al., 2008b).

Thus, study results suggest that attempts to evaluate trends in health disparities require judgments about which picture of disparity is essential for the question at hand. When assessing health disparity, selection of the appropriate reference point, absolute or relative disparity, and use of un-weighted or weighted social groups should be specific. Firebaugh et al (2008) suggested that the decision of whether or not to use a population-weighted measure of disparity is a decision about how much value to place on the health of individuals: Population-weighted measures count all individuals equally, while unweighted estimates count all groups equally and weight individuals inversely concerning the size of their social group (Firebaugh, 2009; Harper et al., 2008b).

Population-weighted measures, therefore, capture changes in the distribution of social groups over time and would serve to complement a view that regards this as an essential aspect of health disparity. Alternatively, unweighted measures would complete a statement that social groups with normative importance should be weighted equally, regardless of their population size. One of these choices may be justifiable. Still, because this is likely to have consequences on one's conclusions about the magnitude of disparity, the reasons for choosing one versus another conception of disparity should be made clear at the outset.

One strength of this study was using SEER-linked data, a population-based, high-quality data source. In addition, the analytic approach was robust, including both absolute and relative measures of disparities. Finally, the project presents a more comprehensive framework for comparing health disparities to determine the impacts more comprehensively.

8.4.2 Limitation

Despite the strengths of the study, there were some limitations. First, analysis is restricted to male and female disparity only. Estimates using different ethnic groups could explore different aspects of disparity trends. Second, the study used a limited few socioeconomic variable: education, income, unemployment, and poverty. Different measures of health behaviors and built environment could generate other aspects of lung cancer disparity.

Concerning the disparity metrics selected, because most metrics have been adapted from economic applications, they use either the population average or the "best" group's rate as the reference. This can imply that equity is the primary goal, even if it is achieved by reducing health for the most advantaged group. Moreover, the more sophisticated disparity metrics lack clear interpretations regarding the magnitude of existing disparities beyond determining if the metric is significantly different from zero. More methodological research is needed on metrics that allow for comparisons to an "ideal" rate or value and more easily interpretable metrics (e.g., Healthy People 2020 target).

Finally, more research is needed to explore the contribution of neighborhood factors (e.g., segregation, SES) to disparities in lung cancer and how differences may vary geographically.

8.5 Area for Future Research

This study presents suggestive evidence of the association between the social environment, physical environment, and health behaviors on lung cancer mortality in Kentucky.

Future research should focus on collecting primary data from Appalachian and non-Appalachian regions to evaluate the significance. In addition to collecting primary data, future research should aim to test more social determinants of health. For example, research should include the collection of biological specimens (e.g., toenails, urine, and blood), health records (CT lung screening, smoking history), and environmental samples (e.g., air, water, and soil) to determine the presence of trace elements and other lung carcinogens.

Furthermore, future studies should also address the possibility that exposure to relatively low levels of contaminants might be interacting with other factors to increase risk. Smoking, for example, has been shown to interact synergistically with arsenic so that smokers are at greater risk of arsenic-related metabolic and health effects than nonsmokers (Hopenhayn-Rich, Biggs, Smith, Kalman, & Moore, 1996). The population of Appalachian Kentucky might thus be especially sensitive to this or similar environmental exposures due to the high prevalence of heavy tobacco use.

In the future, researchers should focus on assessing the multilevel determinants of health and health care disparities, including individual, provider, and organizational factors, on understanding the root causes of cancer inequalities. Finally, when developing and implementing interventions designed to eliminate disparities, researchers should consider study designs that yield generalizable data on the effectiveness of the intervention and encourage the participation of vulnerable populations. Ultimately, researchers should be encouraged to publish their findings, so they are available to communities, policymakers, and other stakeholders to maximize benefit in the field and strengthen the policy implications of their work.

References

American Cancer Association. (2021). Lung Cancer Fact Sheet. *Mortality*. Retrieved from https://www.lung.org/lung-health-diseases/lung-disease-lookup/lung-cancer/resource-library/lung-cancer-fact-sheet

American Cancer Society. (2020). Lung Cancer Risks for People Who Don't Smoke. *Secondhand smoke.* Retrieved from https://www.cancer.org/latest-news/why-lung-cancer-strikes-nonsmokers.html

American Cancer Society. (2022). Radon and Cancer. *What is radon?* Retrieved from https://www.cancer.org/cancer/cancer-causes/radiation-exposure/radon.html

Anselin, L. (2010). Thirty Years of Spatial Econometrics. *Papers in Regional Science, 89*, 3-25. doi:10.1111/j.1435-5957.2010.00279.x

Appalachian Regional Commission. Appalachian Counties Served by ARC. *Counties in Appalachia.* Retrieved from https://www.arc.gov/appalachian-counties-served-by-arc/

Appalachian Regional Commission. (2017). Retrieved from https://www.arc.gov/tax_year/2017/page/2/

Appalachian Regional Commission. (2019). Issue brief: health disparities related to smoking in Appalachia. Retrieved from https://www.arc.gov/report/issue-brief-health-disparities-related-to-smoking-in-appalachia-practical-strategies-and-recommendations-for-communities/

Appalachian Regional Commission. (2021). Kentucky. *Counties in Appalachia.* Retrieved from https://www.arc.gov/

Arnold, M., Rentería, E., Conway, D. I., Bray, F., Van Ourti, T., & Soerjomataram, I. (2016). Inequalities in cancer incidence and mortality across medium to highly developed countries in the twenty-first century. *Cancer Causes Control, 27*(8), 999-1007. doi:10.1007/s10552-016-0777-7

Bangal, R., Giri, P., Bangal, S., More, M., & Singh, K. (2014). Socio-demographic profile and associated risk factors in cancer patients attending the Oncology OPD of a tertiary care teaching hospital in Western Maharashtra, India. *International Journal of Medical Science and Public Health, 3*, 1. doi:10.5455/ijmsph.2014.200820142

Beenackers, M. A., Oude Groeniger, J., van Lenthe, F. J., & Kamphuis, C. B. M. (2018). The role of financial strain and self-control in explaining health behaviours: the GLOBE study. *Eur J Public Health, 28*(4), 597-603. doi:10.1093/eurpub/ckx212

Berlia, S. (2016). *"The Link Between Smoking, Lung Cancer and Socioeconomic Factors in Kentucky"* ((M.P.H. & Dr.P.H.)). University of Kentucky, https://uknowledge.uky.edu/cph_etds/108. Retrieved from https://uknowledge.uky.edu/cph_etds/108 (108)

Braveman, P. (2014). What are health disparities and health equity? We need to be clear. *Public health reports (Washington, D.C. : 1974), 129 Suppl 2*(Suppl 2), 5-8. doi:10.1177/00333549141291S203

Braveman, P., Egerter, S., & Williams, D. R. (2011). The social determinants of health: coming of age. *Annu Rev Public Health, 32*, 381-398. doi:10.1146/annurev-publhealth-031210-101218

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin, 68*(6), 394-424. doi:10.3322/caac.21492

Breen, N., Scott, S., Percy-Laurry, A., Lewis, D., & Glasgow, R. (2014). Health disparities calculator: a methodologically rigorous tool for analyzing inequalities in population health. *American journal of public health, 104*(9), 1589-1591. doi:10.2105/AJPH.2014.301982

Buettner-Schmidt, K., Miller, D. R., & Maack, B. (2019). Disparities in Rural Tobacco Use, Smoke-Free Policies, and Tobacco Taxes. *Western journal of nursing research, 41*(8), 1184-1202. doi:10.1177/0193945919828061

Burbank, F., & Fraumeni, J. F., Jr. (1972). U.S. cancer mortality: nonwhite predominance. *J Natl Cancer Inst, 49*(3), 649-659.

Cao, Q., Rui, G., & Liang, Y. (2018). Study on PM2.5 pollution and the mortality due to lung cancer in China based on geographic weighted regression model. *BMC Public Health, 18*(1), 925. doi:10.1186/s12889-018-5844-4

Caposole, M. Z., Miller, K., Kim, J.-N., Steward, N. A., & Bauer, T. L. (2014). Elimination of socioeconomic and racial disparities related to lung cancer: Closing the gap at a high volume community cancer center. *Surgical Oncology, 23*(2), 46-52. doi:https://doi.org/10.1016/j.suronc.2014.02.001

Cardarelli, K., Westneat, S., Dunfee, M., May, B., Schoenberg, N., & Browning, S. (2021). Persistent disparities in smoking among rural Appalachians: evidence from the Mountain Air Project. *BMC Public Health, 21*(1), 270. doi:10.1186/s12889-021-10334-6

Casebeer, A. W., Antol, D. D., Hopson, S., Khoury, R., Renda, A., Parikh, A., . . . Bunce, M. (2019). Using the Healthy Days Measure to Assess Factors Associated with Poor Health-Related Quality of Life for Patients with Metastatic Breast, Lung, or Colorectal Cancer Enrolled in a Medicare Advantage Health Plan. *Popul Health Manag, 22*(5), 440-448. doi:10.1089/pop.2019.0054

Castro, S., Sosa, E., Lozano, V., Akhtar, A., Love, K., Duffels, J., . . . Erhunmwunsee, L. (2021). The impact of income and education on lung cancer screening utilization, eligibility, and outcomes: a narrative review of socioeconomic disparities in lung cancer screening. *Journal of thoracic disease, 13*(6), 3745-3757. doi:10.21037/jtd-20-3281

Cavelaars, A. E., Kunst, A. E., Geurts, J. J., Crialesi, R., Grötvedt, L., Helmert, U., . . . Mackenbach, J. P. (2000). Educational differences in smoking: international comparison. *Bmj, 320*(7242), 1102-1107. doi:10.1136/bmj.320.7242.1102

CDC. (2017a). Behavioral Risk Factor Surveillance System, 2017. Retrieved from https://www.cdc.gov/brfss/index.html

CDC. (2017b). State Tobacco Activities Tracking and Evaluation (STATE) System. Retrieved from https://www.cdc.gov/statesystem/index.html

CDC. (2017c). Youth Risk Behavior Surveillance System (YRBSS). Retrieved from https://www.cdc.gov/healthyyouth/data/yrbs/index.htm centers for Disease Control and Prevention. United States Cancer Statistics (USCS). Retrieved from https://www.cdc.gov/cancer/uscs/index.htm

centers for Disease Control and Prevention. (2018). Health-Related Quality of Life (HRQOL). Retrieved from https://www.cdc.gov/hrqol/methods.htm

centers for Disease Control and Prevention. (2019). Smoking and Tobacco Use Retrieved from https://www.cdc.gov/tobacco/data_statistics/fact_sheets/fast_facts/index.htm#toll

centers for Disease Control and Prevention. (2021). What Are the Risk Factors for Lung Cancer? *Smoking* Retrieved from https://www.cdc.gov/cancer/lung/basic_info/risk_factors.htmS

Chawińska, E., Tukiendorf, A., & Miszczyk, L. (2013). Unemployment and Lung Cancer Incidence in the Province of Opole - Brief Report. *Central European journal of public health, 21*, 118-120. doi:10.21101/cejph.a3872

Christian, W. J., Huang, B., Rinehart, J., & Hopenhayn, C. (2011). Exploring geographic variation in lung cancer incidence in Kentucky using a spatial scan statistic: elevated risk in the Appalachian coal-mining region. *Public health reports (Washington, D.C. : 1974), 126*(6), 789-796. doi:10.1177/003335491112600604

Christian, W. J., Huang, B., Rinehart, J., & Hopenhayn, C. (2011). Exploring geographic variation in lung cancer incidence in Kentucky using a spatial scan statistic: elevated risk in the Appalachian coal-mining region. *Public health reports (Washington, D.C. : 1974), 126*(6), 789-796. doi:10.1177/003335491112600604

Christian, W. J., Walker, C. J., Huang, B., & Hahn, E. J. (2019). Effect of Local Smoke-Free Ordinances on Smoking Prevalence in Kentucky, 2002-2009. *Southern medical journal, 112*(7), 369-375. doi:10.14423/SMJ.0000000000001000

Christian, W. J., Walker, C. J., Huang, B., Levy, J. E., Durbin, E., & Arnold, S. (2020). Using residential histories in case-control analysis of lung cancer and mountaintop removal coal mining in Central Appalachia. *Spat Spatiotemporal Epidemiol, 35*, 100364. doi:10.1016/j.sste.2020.100364

Clegg, L. X., Reichman, M. E., Miller, B. A., Hankey, B. F., Singh, G. K., Lin, Y. D., . . . Edwards, B. K. (2009). Impact of socioeconomic status on cancer incidence and stage at diagnosis: selected findings from the surveillance, epidemiology, and end results: National Longitudinal Mortality Study. *Cancer causes & control : CCC, 20*(4), 417-435. doi:10.1007/s10552-008-9256-0

Cohen, R., & Velho, V. (2002). Update on respiratory disease from coal mine and silica dust. *Clin Chest Med, 23*(4), 811-826. doi:10.1016/s0272-5231(02)00026-6

Columbia Public Health. (2019). Geographically Weighted Regression. Retrieved from https://www.publichealth.columbia.edu/research/population-health-methods/geographically-weighted-regression

Consonni, D., Carugno, M., De Matteis, S., Nordio, F., Randi, G., Bazzano, M., . . . Landi, M. T. (2018). Outdoor particulate matter (PM10) exposure and lung cancer risk in the EAGLE study. *PLoS One, 13*(9), e0203539. doi:10.1371/journal.pone.0203539

County Health Rankings & Roadmaps. (2021a). Health Factors Retrieved from https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model/health-factors

County Health Rankings & Roadmaps. (2021b). Social and Economic Factors. Retrieved from https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model/health-factors/social-and-economic-factors

County Health Rankings & Roadmaps. (2022a). Air and Water Quality. Retrieved from https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model/health-factors/physical-environment/air-and-water-quality

County Health Rankings & Roadmaps. (2022b). Health Behaviors. Retrieved from https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model/health-factors/health-behaviors

County Health Rankings & Roadmaps. (2022c). Physical Environment. Retrieved from https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model/health-factors/physical-environment

Davis, R. M. (2000). Healthy People 2010: objectives for the United States. Impressive, but unwieldy. *BMJ (Clinical research ed.)*, *320*(7238), 818-819. doi:10.1136/bmj.320.7238.818

Dela Cruz, C. S., Tanoue, L. T., & Matthay, R. A. (2011). Lung cancer: epidemiology, etiology, and prevention. *Clinics in chest medicine*, *32*(4), 605-644. doi:10.1016/j.ccm.2011.09.001

Denham, S. A., Meyer, M. G., Toborg, M. A., & Mande, M. J. (2004). Providing health education to Appalachia populations. *Holist Nurs Pract, 18*(6), 293-301. doi:10.1097/00004650-200411000-00005

Didkowska, J., Wojciechowska, U., Mańczuk, M., & Łobaszewski, J. (2016). Lung cancer epidemiology: contemporary and future challenges worldwide. *Annals of Translational Medicine, 4*(8), 2. Retrieved from https://atm.amegroups.com/article/view/9532

Division of Cancer Prevention and Control, C. f. D. C. a. P. (2020, September 22, 2020). What Are the Risk Factors for Lung Cancer? *Smoking.* Retrieved from https://www.cdc.gov/cancer/lung/basic_info/risk_factors.htm

Doogan, N. J., Roberts, M. E., Wewers, M. E., Stanton, C. A., Keith, D. R., Gaalema, D. E., . . . Higgins, S. T. (2017). A growing geographic disparity: Rural and urban cigarette smoking trends in the United States. *Prev Med, 104*, 79-85. doi:10.1016/j.ypmed.2017.03.011

Ekberg-Aronsson, M., Nilsson, P. M., Nilsson, J. A., Pehrsson, K., & Löfdahl, C. G. (2006). Socio-economic status and lung cancer risk including histologic subtyping--a longitudinal study. *Lung Cancer, 51*(1), 21-29. doi:10.1016/j.lungcan.2005.08.014

Elkbuli, A., Byrne, M. M., Zhao, W., Sutherland, M., McKenney, M., Godinez, Y., . . . Koru-Sengul, T. (2020). Gender disparities in lung cancer survival from an enriched Florida population-based cancer registry. *Annals of Medicine and Surgery, 60*, 680-685. doi:https://doi.org/10.1016/j.amsu.2020.11.081

Esri. Geographically Weighted Regression Retrieved from https://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/geographically-weighted-regression.htm

Fidler-Benaoudia, M. M., Torre, L. A., Bray, F., Ferlay, J., & Jemal, A. (2020). Lung cancer incidence in young women vs. young men: A systematic analysis in 40 countries. *Int J Cancer, 147*(3), 811-819. doi:10.1002/ijc.32809

Finke, I., Behrens, G., Weisser, L., Brenner, H., & Jansen, L. (2018). Socioeconomic Differences and Lung Cancer Survival-Systematic Review and Meta-Analysis. *Frontiers in oncology, 8*, 536-536. doi:10.3389/fonc.2018.00536

Firebaugh, G. (2009). Harvard University Press.

Fotheringham, A., Brunsdon, C., & Charlton, M. (2000). *Quantitative Geography: Perspectives on Spatial Data Analysis*.

Fotheringham, A., Brunsdon, C., & Charlton, M. (2002). Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. *John Wiley & Sons, 13*.

Fotheringham, A. S., Charlton, M. E., & Brunsdon, C. (1998). Geographically Weighted Regression: A Natural Evolution of the Expansion Method for Spatial Data Analysis. *Environment and Planning A: Economy and Space, 30*(11), 1905-1927. doi:10.1068/a301905

Ghasemi, S., Mahaki, B., Dreassi, E., & Aghamohammadi, S. (2020). Spatial Variation in Lung Cancer Mortality and Related Men-Women Disparities in Iran from 2011 to 2014. *Cancer management and research, 12*, 4615-4624. doi:10.2147/CMAR.S247178

Giovino, G. A. (2002). Epidemiology of tobacco use in the United States. *Oncogene, 21*(48), 7326-7340. doi:10.1038/sj.onc.1205808

GIS/Data Center Guides. (2021). Analyzing Spatial Patterns. Retrieved from https://wiki.rice.edu/confluence/display/GDCGUIDES/Analyzing+Spatial+Patterns

Gomez, S. L., Shariff-Marco, S., DeRouen, M., Keegan, T. H., 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

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

Gross, D. A. (2010). The relationship between educational attainment and lung cancer mortality in Kentucky: implications for nurses. *Online Journal of Rural Nursing & Health Care*, *10*, 75+. Retrieved from https://link.gale.com/apps/doc/A245393205/HRCA?u=nysl_oweb&sid=googleScholar&xid=1a9eb058

Guarga, L., Ameijide, A., Marcos-Gragera, R., Carulla, M., Delgadillo, J., Borràs, J. M., & Galceran, J. (2021). Trends in lung cancer incidence by age, sex and histology from 2012 to 2025 in Catalonia (Spain). *Scientific Reports*, *11*(1), 23274. doi:10.1038/s41598-021-02582-8

Hahn E, B. K., Rayens M, Riker C. (2008). *Workplace tobacco policy study* Retrieved from Lexington:

Hamra, G. B., Guha, N., Cohen, A., Laden, F., Raaschou-Nielsen, O., Samet, J. M., . . . Loomis, D. (2014). Outdoor particulate matter exposure and lung cancer: a systematic review and meta-analysis. *Environ Health Perspect, 122*(9), 906-911. doi:10.1289/ehp/1408092

Harper, S., & Lynch, J. (2004). Methods for Measuring Cancer Disparities: Using Data Relevant to Healthy People 2010 Cancer-Related Objectives. *NCI Cancer surveill Monogr Ser, 6*.

Harper, S., Lynch, J., Meersman, S. C., Breen, N., Davis, W. W., & Reichman, M. E. (2008a). An overview of methods for monitoring social disparities in cancer with an example using trends in lung cancer incidence by area-socioeconomic position and race-ethnicity, 1992-2004. *American journal of epidemiology, 167*(8), 889-899. doi:10.1093/aje/kwn016

Harper, S., Lynch, J., Meersman, S. C., Breen, N., Davis, W. W., & Reichman, M. E. (2008b). An Overview of Methods for Monitoring Social Disparities in Cancer with an Example Using Trends in Lung Cancer Incidence by Area-Socioeconomic Position and Race-Ethnicity, 1992–2004. *American journal of epidemiology, 167*(8), 889-899. doi:10.1093/aje/kwn016

He, Y., Li, S., Ren, S., Cai, W., Li, X., Zhao, C., . . . Zhou, C. (2013). Impact of family history of cancer on the incidence of mutation in epidermal growth factor receptor gene in non-small cell lung cancer patients. *Lung Cancer, 81*(2), 162-166. doi:10.1016/j.lungcan.2013.05.004

Health Disparities Calculator, V. (September 12, 2019).

Healthy people.gov. (2022). Disparities. Retrieved from https://www.healthypeople.gov/2020/about/foundation-health-measures/Disparities

Heidary, F., Rahimi, A., & Gharebaghi, R. (2013). Poverty as a risk factor in human cancers. *Iranian journal of public health, 42*(3), 341-343. Retrieved from https://pubmed.ncbi.nlm.nih.gov/23641414

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3633807/

Hendryx, M., O'Donnell, K., & Horn, K. (2008). Lung cancer mortality is elevated in coal-mining areas of Appalachia. *Lung Cancer, 62*(1), 1-7. doi:10.1016/j.lungcan.2008.02.004

Hoffman, P. C., Mauer, A. M., & Vokes, E. E. (2000). Lung cancer. *Lancet, 355*(9202), 479-485. doi:10.1016/s0140-6736(00)82038-3

Holt, D., Steel, D., Tranmer, M., & Wrigley, N. (2010). Aggregation and Ecological Effects in Geographical Based Data. *Geographical Analysis, 28,* 244-261. doi:10.1111/j.1538-4632.1996.tb00933.x

Hopenhayn-Rich, C., Biggs, M. L., Smith, A. H., Kalman, D. A., & Moore, L. E. (1996). Methylation study of a population environmentally exposed to arsenic in drinking water. *Environmental health perspectives, 104*(6), 620-628. doi:10.1289/ehp.96104620

Hopenhayn-Rich, C., Stump, M. L., & Browning, S. R. (2002). Regional assessment of atrazine exposure and incidence of breast and ovarian cancers in Kentucky. *Arch Environ Contam Toxicol, 42*(1), 127-136. doi:10.1007/s002440010300

Hopenhayn, C., Jenkins, T. M., & Petrik, J. (2003). The burden of lung cancer in Kentucky. *J Ky Med Assoc, 101*(1), 15-20.

Hosseinpoor, A. R., Bergen, N., Mendis, S., Harper, S., Verdes, E., Kunst, A., & Chatterji, S. (2012). Socioeconomic inequality in the prevalence of noncommunicable diseases in low- and middle-income countries: results from the World Health Survey. *BMC Public Health, 12*, 474. doi:10.1186/1471-2458-12-474

Hovanec, J., Siemiatycki, J., Conway, D. I., Olsson, A., Stücker, I., Guida, F., . . . Behrens, T. (2018). Lung cancer and socioeconomic status in a pooled analysis of case-control studies. *PLoS One, 13*(2), e0192999-e0192999. doi:10.1371/journal.pone.0192999

Howlader, N., Forjaz, G., Mooradian, M. J., Meza, R., Kong, C. Y., Cronin, K. A., . . . Feuer, E. J. (2020). The Effect of Advances in Lung-Cancer Treatment on Population Mortality. *New England Journal of Medicine, 383*(7), 640-649. doi:10.1056/NEJMoa1916623

Islami, F., Torre, L. A., & Jemal, A. (2015). Global trends of lung cancer mortality and smoking prevalence. *Translational lung cancer research, 4*(4), 327-338. doi:10.3978/j.issn.2218-6751.2015.08.04

Izarzugaza, M. I., Ardanaz, E., Chirlaque, M. D., Font, C., Ameijide, A., & Linares, C. (2010). Tobacco-related tumours of the lung, bladder and larynx: changes in Spain. *Ann Oncol, 21 Suppl 3*, iii52-60. doi:10.1093/annonc/mdq084

Kawachi, I., & Kennedy, B. P. (1997). Health and social cohesion: why care about income inequality? *Bmj, 314*(7086), 1037-1040. doi:10.1136/bmj.314.7086.1037

Kentucky Cancer Registry. (2019). Kentucky Cancer Incidence and Mortality Rates, 2011–2015. . Retrieved from https://www.cancer-rates.info/ky/

Kim, H. J., Fay, M. P., Feuer, E. J., & Midthune, D. N. (2000). Permutation tests for joinpoint regression with applications to cancer rates. *Stat Med, 19*(3), 335-351. doi:10.1002/(sici)1097-0258(20000215)19:3<335::aid-sim336>3.0.co;2-z

Kinsey, T., Jemal, A., Liff, J., Ward, E., & Thun, M. (2008). Secular Trends in Mortality From Common Cancers in the United States by Educational Attainment, 1993–2001. *Journal of the National Cancer Institute, 100*, 1003-1012. doi:10.1093/jnci/djn207

Klassen, A. C., Hsieh, S., Pankiewicz, A., Kabbe, A., Hayes, J., & Curriero, F. (2019). The association of neighborhood-level social class and tobacco consumption with adverse lung cancer characteristics in Maryland. *Tobacco induced diseases, 17*, 06-06. doi:10.18332/tid/100525

Knight, J. R., Williamson, L. H., Armstrong, D. K., & Westbrook, E. A. (2019). Understanding Lung Cancer Resources and Barriers Among Worksites With Mostly Male Employees in Eight Rural Kentucky Counties: A Focus Group Discussion. *American journal of men's health, 13*(6), 1557988319882585-1557988319882585. doi:10.1177/1557988319882585

Krieger, N. (2005). Defining and investigating social disparities in cancer: critical issues. *Cancer Causes Control, 16*(1), 5-14. doi:10.1007/s10552-004-1251-5

Laaksonen, M., Rahkonen, O., Karvonen, S., & Lahelma, E. (2005). Socioeconomic status and smoking: analysing inequalities with multiple indicators. *Eur J Public Health, 15*(3), 262-269. doi:10.1093/eurpub/cki115

Lantz, P. M., Mendez, D., & Philbert, M. A. (2013). Radon, smoking, and lung cancer: the need to refocus radon control policy. *American journal of public health, 103*(3), 443-447. doi:10.2105/AJPH.2012.300926

Lengerich, E. J., Tucker, T. C., Powell, R. K., Colsher, P., Lehman, E., Ward, A. J., . . . Wyatt, S. W. (2005). Cancer incidence in Kentucky, Pennsylvania, and West Virginia: disparities in Appalachia. *J Rural Health, 21*(1), 39-47. doi:10.1111/j.1748-0361.2005.tb00060.x

Levi, F., Lucchini, F., La Vecchia, C., & Negri, E. (1999). Trends in mortality from cancer in the European Union, 1955-94. *Lancet, 354*(9180), 742-743. doi:10.1016/s0140-6736(99)01909-1

Li, C., Wang, C., Yu, J., Fan, Y., Liu, D., Zhou, W., & Shi, T. (2020). Residential Radon and Histological Types of Lung Cancer: A Meta-Analysis of Case–Control Studies. *International Journal of Environmental Research and Public Health, 17*(4), 1457. doi:10.3390/ijerph17041457

Li, J., Li, W. X., Bai, C., & Song, Y. (2017). Particulate matter-induced epigenetic changes and lung cancer. *Clin Respir J, 11*(5), 539-546. doi:10.1111/crj.12389

Lin, C.-H., & Wen, T.-H. (2011a). Using Geographically Weighted Regression (GWR) to Explore Spatial Varying Relationships of Immature Mosquitoes and Human Densities with the Incidence of Dengue. *International Journal of Environmental Research and Public Health, 8*(7). doi:10.3390/ijerph8072798

Lin, C.-H., & Wen, T.-H. (2011b). Using geographically weighted regression (GWR) to explore spatial varying relationships of immature mosquitoes and human densities with the incidence of dengue. *International Journal of Environmental Research and Public Health, 8*(7), 2798-2815. doi:10.3390/ijerph8072798

Lortet-Tieulent, J., Soerjomataram, I., Ferlay, J., Rutherford, M., Weiderpass, E., & Bray, F. (2014). International trends in lung cancer incidence by histological subtype: Adenocarcinoma stabilizing in men but still increasing in women. *Lung Cancer, 84*(1), 13-22. doi:https://doi.org/10.1016/j.lungcan.2014.01.009

Lubin, J. H., Boice, J. D., Jr., Edling, C., Hornung, R. W., Howe, G. R., Kunz, E., . . . et al. (1995). Lung cancer in radon-exposed miners and estimation of risk from indoor exposure. *J Natl Cancer Inst, 87*(11), 817-827. doi:10.1093/jnci/87.11.817

Lynge, E. (1997). Unemployment and cancer: a literature review. *IARC Sci Publ*(138), 343-351.

Mackenbach, J. P., Stirbu, I., Roskam, A. J., Schaap, M. M., Menvielle, G., Leinsalu, M., & Kunst, A. E. (2008). Socioeconomic inequalities in health in 22 European countries. *N Engl J Med, 358*(23), 2468-2481. doi:10.1056/NEJMsa0707519

MacRosty, C. R., & Rivera, M. P. (2020). Lung Cancer in Women: A Modern Epidemic. *Clin Chest Med, 41*(1), 53-65. doi:10.1016/j.ccm.2019.10.005

Mao, Y., Hu, J., Ugnat, A.-M., Semenciw, R., Fincham, S., & Group, a. t. C. C. R. E. R. (2001). Socioeconomic status and lung cancer risk in Canada. *International Journal of Epidemiology, 30*(4), 809-817. doi:10.1093/ije/30.4.809

Mao, Y., Hu, J., Ugnat, A. M., Semenciw, R., & Fincham, S. (2001). Socioeconomic status and lung cancer risk in Canada. *Int J Epidemiol, 30*(4), 809-817. doi:10.1093/ije/30.4.809

Marí-Dell'Olmo, M., Gotsens, M., Palència, L., Burström, B., Corman, D., Costa, G., . . . Borrell, C. (2015). Socioeconomic inequalities in cause-specific mortality in 15 European cities. *J Epidemiol Community Health, 69*(5), 432-441. doi:10.1136/jech-2014-204312

Marmot, M. (2005). Social determinants of health inequalities. *Lancet, 365*(9464), 1099-1104. doi:10.1016/s0140-6736(05)71146-6

Matthews, S. A., & Yang, T.-C. (2012). Mapping the results of local statistics: Using geographically weighted regression. *Demographic research*, *26*, 151-166. doi:10.4054/DemRes.2012.26.6

Miller, B. G., & MacCalman, L. (2010). Cause-specific mortality in British coal workers and exposure to respirable dust and quartz. *Occup Environ Med, 67*(4), 270-276. doi:10.1136/oem.2009.046151

Mohamed, M. K., Herndon, D., Schmidt, M., & Manning, M. A. (2020). The effect of under and uninsured status on survival in lung cancer while adjusting for other mortality risk factors. *Journal of Clinical Oncology, 38*(15_suppl), e21734-e21734. doi:10.1200/JCO.2020.38.15_suppl.e21734

Montazeri, A., Gillis, C. R., & McEwen, J. (1998). Quality of life in patients with lung cancer: a review of literature from 1970 to 1995. *Chest, 113*(2), 467-481. doi:10.1378/chest.113.2.467

Moore, J. X., Akinyemiju, T., & Wang, H. E. (2017). Pollution and regional variations of lung cancer mortality in the United States. *Cancer epidemiology, 49*, 118-127. doi:10.1016/j.canep.2017.05.013

Mustafic, H., Jabre, P., Caussin, C., Murad, M. H., Escolano, S., Tafflet, M., . . . Jouven, X. (2012). Main air pollutants and myocardial infarction: a systematic review and meta-analysis. *Jama, 307*(7), 713-721. doi:10.1001/jama.2012.126

Nagelhout, G. E., de Korte-de Boer, D., Kunst, A. E., van der Meer, R. M., de Vries, H., van Gelder, B. M., & Willemsen, M. C. (2012). Trends in socioeconomic inequalities in smoking prevalence, consumption, initiation, and cessation between 2001 and 2008 in the Netherlands. Findings from a national population survey. *BMC Public Health, 12*, 303. doi:10.1186/1471-2458-12-303

National Cancer Insitute. (2022). Joinpoint Trend Analysis Software. Retrieved from https://surveillance.cancer.gov/joinpoint/

National Cancer Institute. (2011). Radon and Cancer. *How does radon cause cancer?* Retrieved from https://www.cancer.gov/about-cancer/causes-prevention/risk/substances/radon/radon-fact-sheet

National Center for Health Statistics. National Vital Statistics System, m. d. (2018). QuickStats: Age-Adjusted Lung Cancer Death* Rates,† by State Retrieved from https://www.cdc.gov/mmwr/volumes/69/wr/mm6936a8.htm#suggestedcitation

National Institutes of, H. (2006). The NCI strategic plan for leading the nation to eliminate the suffering and death due to cancer. Retrieved from http://catalog.hathitrust.org/api/volumes/oclc/76871874.html

Neuberger, J. S., Mahnken, J. D., Mayo, M. S., & Field, R. W. (2006). Risk factors for lung cancer in Iowa women: implications for prevention. *Cancer detection and prevention, 30*(2), 158-167. doi:10.1016/j.cdp.2006.03.001

O'Connor, J. M., Sedghi, T., Dhodapkar, M., Kane, M. J., & Gross, C. P. (2018). Factors Associated With Cancer Disparities Among Low-, Medium-, and High-Income US Counties. *JAMA Network Open, 1*(6), e183146-e183146. doi:10.1001/jamanetworkopen.2018.3146

Parker, J., Rich, D., Glinianaia, S., Leem, J.-H., Wartenberg, D., Bell, M., . . . Woodruff, T. (2011). The International Collaboration on Air Pollution and Pregnancy Outcomes: Initial Results. *Environmental health perspectives, 119*, 1023-1028. doi:10.1289/ehp.1002725

Powell, H. A. (2019). Socioeconomic deprivation and inequalities in lung cancer: time to delve deeper? *Thorax, 74*(1), 11-12. doi:10.1136/thoraxjnl-2018-212362

Preston, S. H., Haines, M. R., & Pamuk, E. R. (1981). *Effects of industrialization and urbanization on mortality in developed countries*.

Raaschou-Nielsen, O., Andersen, Z. J., Beelen, R., Samoli, E., Stafoggia, M., Weinmayr, G., . . . Hoek, G. (2013). Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). *Lancet Oncol, 14*(9), 813-822. doi:10.1016/s1470-2045(13)70279-1

Ridge, C. A., McErlean, A. M., & Ginsberg, M. S. (2013). Epidemiology of lung cancer. *Semin Intervent Radiol, 30*(2), 93-98. doi:10.1055/s-0033-1342949

Rivera, M. P., Katki, H. A., Tanner, N. T., Triplette, M., Sakoda, L. C., Wiener, R. S., . . . Aldrich, M. C. (2020). Addressing Disparities in Lung Cancer Screening Eligibility and Healthcare Access. An Official American Thoracic Society Statement. *Am J Respir Crit Care Med, 202*(7), e95-e112. doi:10.1164/rccm.202008-3053ST

Robbins, H. A., Pfeiffer, R. M., Shiels, M. S., Li, J., Hall, H. I., & Engels, E. A. (2015). Excess cancers among HIV-infected people in the United States. *J Natl Cancer Inst, 107*(4). doi:10.1093/jnci/dju503

Roberts, M. E., Doogan, N. J., Kurti, A. N., Redner, R., Gaalema, D. E., Stanton, C. A., . . . Higgins, S. T. (2016). Rural tobacco use across the United States: How rural and urban areas differ, broken down by census regions and divisions. *Health Place, 39*, 153-159. doi:10.1016/j.healthplace.2016.04.001

Ross, M. H., & Murray, J. (2004). Occupational respiratory disease in mining. *Occup Med (Lond), 54*(5), 304-310. doi:10.1093/occmed/kqh073

Sachs, E., Jackson, V., & Sartipy, U. (2020). Household disposable income and long-term survival after pulmonary resections for lung cancer. *Thorax, 75*(9), 764-770. doi:10.1136/thoraxjnl-2019-214321

Schaap, M. M., van Agt, H. M., & Kunst, A. E. (2008). Identification of socioeconomic groups at increased risk for smoking in European countries: looking beyond educational level. *Nicotine Tob Res, 10*(2), 359-369. doi:10.1080/14622200701825098

Schabath, M. B., & Cote, M. L. (2019). Cancer Progress and Priorities: Lung Cancer. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 28*(10), 1563-1579. doi:10.1158/1055-9965.EPI-19-0221

Schoenberg, N. E., Huang, B., Seshadri, S., & Tucker, T. C. (2015). Trends in cigarette smoking and obesity in Appalachian Kentucky. *Southern medical journal, 108*(3), 170-177. doi:10.14423/smj.0000000000000245

Shi, R., Meacham, S., Davis, G. C., You, W., Sun, Y., & Goessl, C. (2019). Factors influencing high respiratory mortality in coal-mining counties: a repeated cross-sectional study. *BMC Public Health, 19*(1), 1484. doi:10.1186/s12889-019-7858-y

Shu, Y., Zhu, L., Yuan, F., Kong, X., Huang, T., & Cai, Y. D. (2016). Analysis of the relationship between PM2.5 and lung cancer based on protein-protein interactions. *Comb Chem High Throughput Screen, 19*(2), 100-108. doi:10.2174/1386207319666151110123345

Siahpush, M., Singh, G. K., Jones, P. R., & Timsina, L. R. (2010). Racial/ethnic and socioeconomic variations in duration of smoking: results from 2003, 2006 and 2007 Tobacco Use Supplement of the Current Population Survey. *J Public Health (Oxf), 32*(2), 210-218. doi:10.1093/pubmed/fdp104

Sidorchuk, A., Agardh, E. E., Aremu, O., Hallqvist, J., Allebeck, P., & Moradi, T. (2009). Socioeconomic differences in lung cancer incidence: a systematic review and meta-analysis. *Cancer Causes Control, 20*(4), 459-471. doi:10.1007/s10552-009-9300-8

Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2021). Cancer Statistics, 2021. *CA Cancer J Clin, 71*(1), 7-33. doi:10.3322/caac.21654

Siegel, R. L., Miller, K. D., & Jemal, A. (2016). Cancer statistics, 2016. *CA Cancer J Clin, 66*(1), 7-30. doi:10.3322/caac.21332

Siegel, R. L., Miller, K. D., & Jemal, A. (2017). Cancer Statistics, 2017. *CA Cancer J Clin, 67*(1), 7-30. doi:10.3322/caac.21387

Siegel, R. L., Miller, K. D., & Jemal, A. (2018). Cancer statistics, 2018. *CA Cancer J Clin, 68*(1), 7-30. doi:10.3322/caac.21442

Siegel, R. L., Miller, K. D., & Jemal, A. (2019). Cancer statistics, 2019. *CA Cancer J Clin, 69*(1), 7-34. doi:10.3322/caac.21551

Siegel, R. L., Miller, K. D., & Jemal, A. (2020). Cancer statistics, 2020. *CA Cancer J Clin, 70*(1), 7-30. doi:10.3322/caac.21590

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. *Journal of Cancer Epidemiology, 2011*, 107497. doi:10.1155/2011/107497

Slatore, C. G., Au, D. H., & Gould, M. K. (2010). An official American Thoracic Society systematic review: insurance status and disparities in lung cancer practices and outcomes. *Am J Respir Crit Care Med*, *182*(9), 1195-1205. doi:10.1164/rccm.2009-038ST

Smith, G. D., Leon, D., Shipley, M. J., & Rose, G. (1991). Socioeconomic differentials in cancer among men. *Int J Epidemiol, 20*(2), 339-345. doi:10.1093/ije/20.2.339

Strand, B. H., Grøholt, E. K., Steingrímsdóttir, O. A., Blakely, T., Graff-Iversen, S., & Naess, Ø. (2010). Educational inequalities in mortality over four decades in Norway: prospective study of middle aged men and women followed for cause specific mortality, 1960-2000. *Bmj, 340*, c654. doi:10.1136/bmj.c654

Tetzlaff, F., Epping, J., Tetzlaff, J., Golpon, H., & Geyer, S. (2021). Socioeconomic inequalities in lung cancer – a time trend analysis with German health insurance data. *BMC Public Health, 21*(1), 538. doi:10.1186/s12889-021-10576-4

Theil, H. (1967). *Economics and information theory*. Amsterdam: North-Holland.

truth initiative. (2019). he facts about women and tobacco. Retrieved from https://truthinitiative.org/research-resources/targeted-communities/facts-about-women-and-tobacco

Turner, M. C., Krewski, D., Pope, C. A., 3rd, Chen, Y., Gapstur, S. M., & Thun, M. J. (2011). Long-term ambient fine particulate matter air pollution and lung cancer in a large cohort of never-smokers. *Am J Respir Crit Care Med, 184*(12), 1374-1381. doi:10.1164/rccm.201106-1011OC

Twiss, P. C., & Mueller, T. R. (2004). Housing Appalachians: Recent Trends. *Journal of Appalachian Studies, 10*(3), 389-406. Retrieved from http://www.jstor.org/stable/41446647

U.S. Cancer Statistics: Data Visualizations. (November 2017). US Cancer Statistics Working Group / US Department of Health and Human Services / Centers for Disease Control and Prevention and National Cancer Institute. (1999. –2015). Retrieved from https://gis.cdc.gov/grasp/USCS/DataViz.html

Van der Heyden, J. H., Schaap, M. M., Kunst, A. E., Esnaola, S., Borrell, C., Cox, B., . . . Van Oyen, H. (2009). Socioeconomic inequalities in lung cancer mortality in 16 European populations. *Lung Cancer, 63*(3), 322-330. doi:10.1016/j.lungcan.2008.06.006

Vander Weg, M. W., Cunningham, C. L., Howren, M. B., & Cai, X. (2011). Tobacco use and exposure in rural areas: Findings from the Behavioral Risk Factor Surveillance System. *Addict Behav, 36*(3), 231-236. doi:10.1016/j.addbeh.2010.11.005

Wagstaff, A., Paci, P., & van Doorslaer, E. (1991). On the measurement of inequalities in health. *Soc Sci Med, 33*(5), 545-557. doi:10.1016/0277-9536(91)90212-u

Wakelee, H. A., Chang, E. T., Gomez, S. L., Keegan, T. H., Feskanich, D., Clarke, C. A., . . . West, D. W. (2007). Lung cancer incidence in never smokers. *J Clin Oncol*, *25*(5), 472-478. doi:10.1200/jco.2006.07.2983

Wang, J.-F., Zhang, T.-L., & Fu, B.-J. (2016). A measure of spatial stratified heterogeneity. *Ecological Indicators, 67*, 250-256. doi:https://doi.org/10.1016/j.ecolind.2016.02.052

Wang, J. F., Li, X. H., Christakos, G., Liao, Y. L., Zhang, T., Gu, X., & Zheng, X. Y. (2010). Geographical Detectors-Based Health Risk Assessment and its Application in the Neural Tube Defects Study of the Heshun Region, China. *International Journal of Geographical Information Science, 24*(1), 107-127. doi:10.1080/13658810802443457

Wang, L., Sun, W., Zhou, K., Zhang, M., & Bao, P. (2019). Spatial Analysis of Built Environment Risk for Respiratory Health and Its Implication for Urban Planning: A Case Study of Shanghai. *International Journal of Environmental Research and Public Health*, *16*(8), 1455. doi:10.3390/ijerph16081455

Watt, R. G. (2002). Emerging theories into the social determinants of health: implications for oral health promotion. *Community Dent Oral Epidemiol*, *30*(4), 241-247. doi:10.1034/j.1600-0528.2002.300401.x

Watt, R. G., & Sheiham, A. (2012). Integrating the common risk factor approach into a social determinants framework. *Community Dent Oral Epidemiol*, *40*(4), 289-296. doi:10.1111/j.1600-0528.2012.00680.x

Wewers, M. E., Ahijevych, K. L., Chen, M. S., Dresbach, S., Kihm, K. E., & Kuun, P. A. (2000). Tobacco use characteristics among rural Ohio Appalachians. *J Community Health, 25*(5), 377-388. doi:10.1023/a:1005127917122

Wilson, R. J., Ryerson, A. B., Singh, S. D., & King, J. B. (2016). Cancer Incidence in Appalachia, 2004-2011. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 25*(2), 250-258. doi:10.1158/1055-9965.EPI-15-0946

Wilson, S. H., & Walker, G. M. (1993). Unemployment and health: a review. *Public Health, 107*(3), 153-162. doi:10.1016/s0033-3506(05)80436-6

Wong, D. W. S., & Lee, J. (2005). Statistical analysis of geographic information with ArcView GIS and ArcGIS. *Geographic Information Sciences, 11*, 1-3.

Xing, D. F., Xu, C. D., Liao, X. Y., Xing, T. Y., Cheng, S. P., Hu, M. G., & Wang, J. X. (2019). Spatial association between outdoor air pollution and lung cancer incidence in China. *BMC Public Health, 19*(1), 1377. doi:10.1186/s12889-019-7740-y

Xu, Y., & Wang, L. (2014). GIS-based analysis of obesity and the built environment in the US. *Cartography and Geographic Information Science, 42*, 9-21. doi:10.1080/15230406.2014.965748

Yen, I. H., & Kaplan, G. A. (1998). Poverty area residence and changes in physical activity level: evidence from the Alameda County Study. *American journal of public health, 88*(11), 1709-1712. doi:10.2105/ajph.88.11.1709

York, N. L., Rayens, M. K., Zhang, M., Jones, L. G., Casey, B. R., & Hahn, E. J. (2010). Strength of tobacco control in rural communities. *J Rural Health, 26*(2), 120-128. doi:10.1111/j.1748-0361.2010.00273.x

Yue, H., & Hu, T. (2021). Geographical Detector-Based Spatial Modeling of the COVID-19 Mortality Rate in the Continental United States. *Int J Environ Res Public Health, 18*(13). doi:10.3390/ijerph18136832