2016

Comparing African- and U.S.-Born Blacks at Stage of Diagnosis and Treatment for Nonsmall Cell Lung Cancer

Relindis K. Fofung
Walden University

Walden University

College of Health Sciences

This is to certify that the doctoral dissertation by

Relindis Fofung

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Donald Goodwin, Committee Chairperson, Public Health Faculty
Dr. Arti Parikh-Patel, Committee Member, Public Health Faculty
Dr. Roland Thorpe, University Reviewer, Public Health Faculty

Chief Academic Officer
Eric Riedel, Ph.D.

Walden University
2016

Abstract

Comparing African- and U.S.-Born Blacks at Stage of Diagnosis and Treatment for Nonsmall Cell Lung Cancer

by

Relindis K. Fofung

BS, University of Calabar, 1995

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Public Health

Walden University

May 2016

Abstract

Lung cancer is a disease with a high mortality rate for the U.S. Black population. There had been considerable research done on different population demographics, necessary to achieve the Healthy People 2020 overarching goals to eliminate health disparities, gain health equity and maintain quality health. Yet, the African-born Black (AFBB) population has been understudied for nonsmall cell lung cancer (NSCLC). This study sought to determine whether within race differences in stage at diagnosis and treatment of NSCLC exists between AFBB and American-born Blacks (AMBB) populations in the United States. The study data is secondary data collected as part of the National Cancer Institute's Surveillance Epidemiologic and End Result (SEER) Program from 2004-2011. Athough no significant difference was found between AFBB (n = 119) and AMBB (n = 238) relative to NSCLC stage at diagnosis, differences in treatments were found. The proportion of AFBB patients with early stage (I and II) NSCLC who underwent surgery differed significantly from that of AMBB ($p < 0.05$); AFBB patients were more likely to receive surgical therapy. The proportion of AFBB patients with stages I-IV of the disease who received radiation treatment also differed significantly from that of AMBB patients ($p < 0.05$); the latter were more likely to receive radiation therapy. Results from logistic regression analysis indicate that AFBB patients were more likely to receive surgical treatment while AMBB patients were more likely to receive radiation treatment. This study outcome can inform other NSCLC research to provide better insights to the cause of the treatment differences within the race from differing birth places, and efficient planning, evaluation of control programs and management of the disease.

Comparing African- and U.S.-Born Blacks at Stage of Diagnosis and Treatment for Nonsmall Cell Lung Cancer

by

Relindis K. Fofung

BS, University of Calabar, 1995

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Public Health

Walden University

February 2016

Dedication

This dissertation is dedicated to my paternal Grandmother, Mamie Saji Titaladinga Vando, and Great-Grandmother, Kahtasama Monica Titaladinga for their unconditional love in raising me, and laying the solid foundation for my education. They did not live to see me get to this stage in my education which they had highly anticipated, yet their spirits live on with me in all that I do till this day.

Acknowledgments

I am indebted to my dissertation Committee Chair, Dr. Donald Goodwin for his mentorship and selfless efforts from the beginning to the end to make this work a success. It was a wonderful learning process in every aspect of this dissertation and I thank him immensely for making me acquire the necessary knowledge that I would never have without his dedication. My sincere thanks to my dissertation committee member, Dr. Parikh Patel for her comments and positive suggestions that have made this work a success.

To my dear husband, Titus Fofung, and my very precious children, Gurb, Bila and Nahdia, I thank you all very much for understanding my efforts and supporting me for the past several years of my study. I love you all so much and I am so proud to say that you all mean the World to me. I would never have come this far without your patience and encouragements.

Table of Contents

List of Tables

Chapter 1: Introduction to the Study

Problem Statement

According to the National Cancer Institute (2013f), U.S. Whites and Blacks have marked disparities when it comes to cancer incidence and mortality. The Black population has higher incidence and mortality rates. According to the United States Census Bureau (2012), the Black race in the United States constitutes people born in the United States as well as immigrants. This means that the U.S Black population is heterogeneous in terms of place of birth.

Yet, cancer racial comparison from National Cancer Institute (2013f) does not stratify the Black population by place of birth to determine within-race differences in disease epidemiology that might exist. African-born Black (AFBB) immigrants constitute part of the Black race population (United States Census Bureau, 2012) and are considered in the comparative analyses between Blacks and Whites. Their population grew by 200% from the 1980s through the 1990s, then by 100% from then through the 2000s to a total of 1.1 million due to admission of large number of African refugees and the introduction of the diversity visa program by the U.S. government (Capps, McCabe, & Fix, 2011). Therefore, knowledge of the AFBB lung cancer epidemiology as compared to their American-born Black (AMBB) counterpart will be of importance in effective planning, management and evaluation of control programs of the disease for this population.

According to global cancer statistics for 2008, lung cancer was the most common among all cancers for both males and females, accounting for approximately 12.7% of all the cancers and 18.2% of cancer related mortalities (Ferlay et al., 2010). For males, lung

cancer was the most common type of cancer and cancer-related mortality, with an age-standardized incidence rate of 34.0 per 100,000 population members and an age standardized mortality rate of 29.4 per 100,000 population members (Ferlay et al., 2010). For females, lung cancer was the fourth most common type of cancer behind breast, colorectal, and cervical cancers. A comparison of lung cancer incidence and mortality rates by world region showed that the age-standardized incidence rate was 13.5 per 100,000 population members, and the age standardized mortality rate was 11 per 100,000 population members (Ferlay, et al, 2010).

In addition to sex, incidence and mortality rates for lung cancer differ by geographic region and race. According to Ferlay et al. (2010), the Middle African region had the lowest age-adjusted incidence rates for lung cancer among all African regions with 2.8 for males and 0.9 for females (per 100,000 population members), followed by the Western African region with 3.1 for males and 1.2 for females. The Middle African region also had the lowest age-adjusted mortality rates for the disease with 2.7 for males and 0.8 for females (per 100,000 population members), followed by Western Africa with 2.9 for males and 1.1 for females (Ferlay et al., 2010; Figure 6). Meanwhile, North America had the highest age-adjusted incidence rates for lung cancer with 48.5 per 100,000 population members for males and 35.8 per 100,000 population members for females (Ferlay et al., 2010). The U. S. continent also had the highest age-adjusted mortality rate from the disease with 37.9 deaths per 100,000 population members for males and 24.2 deaths per 100,000 population members for females (Ferlay et al., 2010;

Figure 6). This 2008 worldwide cancer statistics show that lung cancer is widely distributed with incidence and mortality rates varying from region to region.

Analyzing data for 2001-2010, researchers at the National Cancer Institute (2013f) reported racial and gender differences in lung cancer incidence and mortality rates among the U.S. population. Investigators observed a higher incidence of lung cancer among males (74.3 cases per 100,000 population members) than among females (51.9 cases per 100,000 population members). Meanwhile, the mortality rates per 100,000 population members for males of all races were 63.5; for females, it was 39.2. The National Cancer Institute also observed gender differences for the disease within race. Among U.S. Black males, the incidence rate per 100,000 population members was 95.8; for Black females, it was 52.2. The National Cancer Institute (2013f) report shows that the mortality rate per 100,000 population members for Black males was 78.5 while for Black females was 37.2. This finding also shows lung cancer racial trends where the incidence rate per 100,000 population members for both sexes for Whites was 63.1 and for Blacks was 69.7 while the mortality rates per 100,000 population members for Whites was 50.2 and for Blacks was 53.5 (National Cancer Institute, 2013f). Overall, from 2001-2010 in the United States, while the incidence rate for lung cancer for Black males was almost double that for Black females; the mortality rate for Black males was more than double that for Black females. Mortality rates for Black patients at late stage nonsmall cell lung cancer (NSCLC) was higher than that for White patients in a study carried out by Hardy, Xia, Liu, Cormier, Nurgalieva, and Du (2009). Yet they showed that for both black and white patients the mortality rates were significantly reduced when stage

specific standard therapies were received. Their study outcome underscores the importance of comparing the incidence, mortality rates, treatments, and outcomes for lung cancer among similar groups.

Risk Factors for lung cancer. ***Risk factors for lung cancer incidence.*** The risk factors for lung cancer even among people of the same race have been shown to vary depending on several factors. Etzel et al. (2008) studied lung cancer risks among African Americans and found that exposure to air pollutants that include tobacco smoke, asbestos, wood dusts, toluene, and xylene increased the risk for developing the disease. The researchers found that lung cancer risk increased with duration and quantity of the tobacco smoke exposure and the age of the individual as well as preexisting conditions such as chronic obstructive pulmonary disease (COPD). Individuals with preexisting diseases such as AIDS with recurrent pneumonia, and pulmonary tuberculosis, are two to five times more likely to develop lung cancer than the general population (Shebl, Engels, Goedert, & Chaturvedi, 2010; Hou, Fu, Ge, Du, & Hua, 2013).

The impact of differing early life experiences to lung cancer health has been studied and documented. Early life exposure to second-hand tobacco smoke increases the risk for lung cancer (Brown Anda, Felitti, Edwards, Malacher, Croft, & Giles, 2010). Also, the burning of biomass or solid fuel for cooking, which is most common in developing nations (Bonjour, et al. 2013), is a risk for lung cancer as it releases and exposes the household to particulate matter and CO_2 at a daily concentration that is comparable to those of active cigarette smokers (Pope et al., 2009, 2011; Smith et al., 2010). Indoor radon and workplace chemical carcinogens contribute to about 25% of all

lung cancer cases and about 300,000 lung cancer deaths worldwide (Youlden, Cramb, & Baade, 2008). Therefore, exposure to certain airborne pollutants is a risk for developing lung cancer.

Lung cancer etiology and mortality perceptions and beliefs differ by race. Here, the likelihood of Black Americans race to hold a belief that would interfere with the prevention and treatment of lung cancer is higher than that of White Americans, *OR* = 2.05, 95% *CI* [1.19-3.53] (Lathan, Okechukwu, Drake, and Bennett,2010). The Black American race would more likely avoid lung cancer disease screening and evaluation due to the fear that they will be diagnosed with the disease.

***Risk factors for lung cancer mortality*.** Lung cancer mortality rate is influenced by several factors. Lung cancer at its earlier stage is asymptomatic, but, gradually, those with the disease develop symptoms that are nonspecific, thus masking the cancer which would be later diagnosed at an advanced stage of development (Youlden, et al., 2008). Diagnosis at a late stage of the disease is the primary reason for the low survival high mortality rates (Youlden, et al., 2008). Nonspecific symptoms coupled with lack of adequate and qualified diagnostic health care personnel and tools contribute to late stage diagnosis and impact the disease prognosis (Youlden, et al., 2008).

The class and stage of lung cancer which constitute the outcome variable in this study is obtained through what is known as the tumor, node, metastasis staging, primarily by the technique called computer tomography and positron emission tomography scans (Mirsadraee, Oswal, Alizadeh, Caulo, & van-Beek, 2012). Furthermore, this staging

process provides the extension of the tumor anatomy, from which the disease prognosis and therapy are based. The two major forms of the disease are the small-cell lung cancer (SCLC), and the NSCLC (Mirsadraee et al., 2012). While SCLC is the more aggressive or highly metastatic form, NSCLC is of higher incidence and less aggressive consisting of three subtypes. The three subtypes for the NSCLC are: adenocarcinoma that is often found in the pleura or outer area of the lung; squamous cell carcinoma which forms in the center of the lung by the bronchi; and the large cell carcinoma, which is the fastest growing of the three subtypes, and usually forms indiscriminately in any part of the lung. Lung cancer prognosis is very poor when it is diagnosed at an advance stage than when diagnosed at an earlier stage. Differences in health insurance have been shown to affect lung cancer diagnosis and care in which case the unavailability or inadequate health insurance policies influence lung cancer mortality (Bradley, Dahman, & Given, 2008). Studies examining risk factors have shown that high mortality rates in Blacks as compared to Whites diminish after controlling for health insurance type (Elchoufaniz, et al, 2013). Thus, The reasons for late stage diagnosis have been linked to several factors that include socioeconomic status, culture, health beliefs, and preexisting lung diseases confounding the symptoms.

African Born Population Health Studies. Health studies with African population had been overwhelmingly involved with infectious diseases such as malaria, tuberculosis, and HIV/AIDS, draining both resources and manpower from other disease studies (Galukande & Kiguli-Malwadde, 2010) providing a wealth of statistics. Such studies help reveal infectious diseases patterns for the African-born population, and guide

policies that allow for appropriate treatment of those identified as infected, and control of the disease. Since such intense studies have been lacking with most other chronic diseases among which is cancer, the pattern of cancer epidemiology in Africa as well as the African-born black immigrants to the U.S. is not well documented and is not well understood. Thus, infectious diseases stand out as the only diseases plaguing Africans. Claims that cancer incidence and mortality rates are high among Africans; with only available statistics being estimated from the limited data that are collected by the sparse urban services for cancer diagnosis, treatment, registry, as well as death registries which serve only about 8% of the total African population (Sitas et al., 2006). Data used in developing the cancer statistics for Africa by the International Agency for Research on Cancer are collected by cancer registries located in 12 countries namely; Algeria, Egypt, Mali, Mozambique, Nigeria, Senegal, South Africa, Gambia, Tunisia, Uganda, and Zimbabwe, out of the 53 African countries (International Agency for Research on Cancer, 2011, and United Nations, 2013a).

U S. Cancer Registries. In the United States, the contemporary estimate of national cancer statistics is provided through the Surveillance, Epidemiology, and End Results (SEER) program. Created by the National Cancer Institute (NCI), it consists of 18 registries located in 14 states (Alaska, Arizona, Connecticut, Michigan, Georgia, California, Hawaii, Iowa, Kentucky, Louisiana, New Jersey, New Mexico, Washington, and Utah; National Cancer Institute, 2013a). Its website contains data on race, date of birth, sex, stage and age at diagnosis, treatment type, place of residence, marital status, country, and state of birth from the cancer surveillance reports for patients diagnosed

with cancer (National Cancer Institute, 2013d, and 2013e).These registries account for about 28% of the U.S. population and include 26% of its Black population. The population that the SEER program covers compares to the general U.S. population in terms of poverty and education level, and consists of a higher proportion of foreign-born persons (National Cancer Institute, 2013a, 2013b, 2013c). For the purpose of lowering cost and improving quality and consistency of data collected, all the registries are supported by a centralized data management system (National Cancer Institute, 2013a, 2013b, 2013c). Thus, the U.S. centralized system of cancer registries provides a unified method of collecting and managing cancer data.

African Registries. African countries participating in the World Health Organization (WHO) cancer data collection lack a unified authoritative cancer registry. Unlike the United States where all 14 participating states operate cancer registries under a unified central organization, SEER, those countries in Africa from which WHO gathers data used for world cancer statistics operate their own independent cancer registries. The few African countries that operate their own cancer registries are Algeria, Egypt, Gambia, Mali, Mozambique, Nigeria, Senegal, South Africa, Tunisia, Uganda and Zimbabwe (International Agency for Research on Cancer, 2011).

Dependence on anecdotes is insufficient in determining how lung cancer epidemiology for the African-born Black subgroup compares with the American-born Black subgroup. Most health studies for population subgroups are usually unavailable because of the difficulties with appropriate and sufficient data collection, the cost, manpower and time. Cancer, including lung cancer data for the African-born Blacks in

the U.S is readily available through the SEER program and is used only when considered in African American cancer epidemiologic studies. The data include variables such as cases' race, sex, age, place of birth, and type, stage and date of cancer diagnosis, treatment type received as well as the disease outcome (National Cancer Institute, 2013d, and 2013e) that can be used to generate cancer statistics on a sample of African born Black population. One of the cancers showing the highest incidence and mortality disparity between the White and Black populations is lung cancer (National Cancer Institute, 2013f). It therefore gives a great opportunity to examine whether there are any differences in the stage at which the cancer is diagnosed and the type of treatment received between the Black populations of African and American nativities in the United States. Therefore, in this study, I seek to determine whether there is any difference in the stage at diagnosis and the type of treatment received for NSCLC among AFBB and AMBB populations in the United States.

This type of analysis requires the disease incidence cases to serve as the population under study. The World Health Organization collects cancer data from different countries in the world using standardized methods (International Agency for Research on Cancer, 2013a). These data have been used to study the global epidemiology of cancers, including lung cancer, providing countries' cancer incidence and mortality rates to the International Agency for Research on Cancer (IARC) for their GLOBOCAN publication (Ferlay, Shin, Bray, Forman, Mathers, & Parkin , 2010; International Agency for Research on Cancer, 2013b). Sources for mortality data include cancer registries, vital records, and verbal autopsy surveys. Then for incidence rates, the WHO obtains

countries' data by: (1) high quality regional data classified in 3 alpha numeric categories as A, B, and C according to the percentage of the cases that it covers which are respectively: greater than 50%; between 10% and 50%; and lower than 10%; (2) National data or category D; (3) Regional data or category E; (4) Frequency data or category F; and (5) no data or category G. According to the type of data available, various methods are used to estimate each country's incidence and mortality rates. These methods include: rates projection to 2012, applying most recent rates to the 2012 population, estimates using modelled survival, using weighted average of the local rates to estimate the national rate, use of data from a single cancer registry or use of the weighted average of local rates, and the use of neighboring countries' rates or data to estimate another country's rates (International Agency for Research on cancer, 2013b). The data source and methods outlined by International Agency for Research on Cancer, (2013b) indicate that data from the United States fall in category A and it covers more than 50% of the cancer cases; but most African countries (including Angola, Burundi, Cameroon, Cape Verde, Central African Republic, Chad, Congo, Democratic Republic of Congo, Cote d Ivoire, Ethiopia, Gabon, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Mali, Morocco, Mozambique, Niger, Nigeria, Rwanda, Senegal, Sierra Leone, Somalia, Sudan, Tanzania, Togo, Uganda, Western Sahara and Zambia) fall in categories E, F and G for which data availability is less than 10% or is not available at all, while the only high quality data from Africa come from Egypt, Uganda and Zimbabwe that fall under category C.

Nature of the Study

I will use a retrospective study design using secondary data acquired from the SEER database, the underlying cancer data source for the study. For this study, the frequency data for lung cancer from 2004-2011 for Black cases of American and African nativities shall be obtained from the SEER database using Black race, place of birth and NSCLC as filters for the specific states that operate a SEER Cancer Registry. The queried data shall be edited to make sure that only patients with information that include the age, gender, the NSCLC stage at diagnosis, treatment type as radiation and surgery, shall be selected for the study. The Statistical Packages for the Social Sciences (SPSS) or the Epi Info software shall use to perform the necessary data analysis.

Research Questions and Hypotheses

I will use the following research questions to guide my investigation.

RQ1: Among Black patients newly diagnosed with NSCLC and reported to SEER between 2004 and 2011, do the proportions of NSCLC stage (I, II, III and IV) differ significantly between AFBB and AMBB patients?

RQ2: Among Black patients diagnosed with early stage (I, II) NSCLC and reported to SEER between 2004 and 2011, does the proportion of AFBB patients who underwent surgery differ significantly from the corresponding proportion of AMBB patients?

RQ3: Is there any significant difference in the proportion of AFBB NSCLC (stages I, II, III and IV) patients treated with radiation to the corresponding proportion of AMBB patients, reported to SEER from 2004-2011?

RQ4: Among Black early stage (I, II) NSCLC patients reported to SEER between 2004 and 2011, does the proportion of AFBB patients who received neither surgery nor radiation differ significantly from the corresponding proportion of AMBB patients? The null hypothesis is that between the African-born and American-born Black populations living in the U.S., there are no differences in their NSCLC stage at diagnosis and the type of treatment they receive for the years 2004-2011. This retrospective study will utilize secondary analysis of quantitative data. With funding from the National Cancer Institute (NCI) the SEER program has created cancer registries around the United States that cover about 28% of the national population. These authoritative data sources provide data with variables that include the cases' age, sex, date of birth, place of birth, race, geographic area, cancer type, and year of diagnosis are collected and submitted to a unified database maintained by SEER (National Cancer Institute, 2013a, 2013b, and 2013c).

Theoretical Framework

I will draw from Ecological Theory which says that the ecology involved in human development describes that science which deals with the growing human's exposure or active engagement in the properties of the changing living environment, as well as their larger interconnecting settings (Bronfenbrenner, 1979). The environment in this context involves the immediate as well as the wider society that include governmental policies, culture and economic structures, that shape the growing person's psychology, social and biological development (Bronfenbrenner, 1979, p. 21). Researchers have shown that adverse childhood experiences lead to increased risk for

lung cancer (Brown et al., 2010). To support this childhood experience risk factor,the effect of place of birth was shown to be a factor that influenced the receipt of breast cancer surgery when Chavan,Goodman, Jemal, & Fedewa (2014) compared U.S. resident breast cancer women born in 6 different countries. The theory held true after controlling for several covariates. Place of birth was the only variable that was linked to the observed outcome that showed within-race differences in which Foreign-born Asian women received less breast cancer surgery when compared to their counterpart American-born Asian women, as well as Non-Hispanic White women with *OR* of .76 at 95% and *CI* = .72 – 0.80. In another study the incidence rate of NSCLC is shown to be approximately 35% higher for foreign born Asians than U.S. born Asians living in the State of California. This NSCLC incidence pattern by nativity is consistent with the populations' tobacco smoking prevalence (Raz, Gomez, & Chang, 2008). Following these observations, whether place of birth influences the NSCLC stage at diagnosis as well as the type of treatment received by AFBB compared to their counterpart AMBB cases will be determined by the data analysis.

Types and stages of lung cancer and treatment options.

The SEER cancer staging manual requires lung cancer to be classified into the different types of tumors that are found. But, these different tumors fall under two main categories; the small cell lung cancer (SCLC) and the non-small cell lung cancer (NSCLC) (National Cancer Institute, 2013d, 2013e, and 2014f). The stage at diagnosis for lung cancer specifically applies to one type, the non-small-cell lung cancer (NSCLC) which constitutes a significantly higher incidence of all lung cancers cases (Mirsadraee, et al,

2012). It is detected at five major stages namely: occult or stage 0, stage I, stage II, stage III and stage IV and thus reported as required by the SEER Program coding and staging manual, though the numbers diagnosed with stage 0 is usually too small to be considered in statistical analysis (National Cancer Institute, 2013d, and 2013e, and 2014f). Meanwhile, there are 5 major types of treatments possibilities for lung cancer and are namely: no treatment, surgery, radiation, chemotherapy and others to include experimental methods. Of these treatments, chemotherapy is only applied to SCLC though it is usually diagnosed at very advanced stage, unsuitable for surgery. Meanwhile NSCLC does not respond to chemotherapy and so is not reported to SEER. But, the earlier stages (I and II) are curable by surgery and radiation since they are localized. Meanwhile, the late stages (III and IV) are not suitable for surgery but do get radiation for palliative purposes (National Cancer Institute, 2014f).

Significance of the study

This study is basically an evaluation of the effect of birth place on the stage (I, II, III and IV) at which non-small cell lung cancer is diagnosed; and the type of treatment (no treatment, surgery or radiation therapy) received between U.S. resident AFBB and AMBB lung cancer patients. Between U.S and Africa, differences in the rate of screening, health services availability, treatment, follow-up care, and record keeping do exist. Therefore, it is best to compare data from samples of the two populations when they are living in the same environment where the opportunity for screening, treatment, care, record keeping and environmental conditions are more similar than when they are living in separate continents. According to the ecological theory on which this study is

based, these factors constitute the ecological factors that affect the wellbeing of an actively growing child (Bronfenbrenner, 1979) and so studying their effects on disease outcome in different populations can help detect where help is needed.

Implications for Social Change

This may help fulfill the goal of reducing cancer disparity for population groups as required by the Healthy People 2020 overarching goal (United States Department of Health and Human Services, 2010 and 2013). This study has a great potential for a social change. The study outcome could indicate whether there is a significant difference in the proportions of lung cancer stage at diagnosis, and treatment between of AFBB and AMBB NSCLC patients. Also, it could be a basis for in-depth investigation to understand each group's cultural, social, belief system, and ease of access to prevention education and treatment, all of which could be potential reasons why such differences were observed. This could aid in the development and implementation of interventions that may influence cultural, social and behavioral modifications that can in turn encourage early diagnosis for the population that is more at risk for late stage NSCLC diagnosis. This will help to appropriately achieve the goal of the Healthy People 2020, in the context of lung cancer that seeks to eliminate health disparities, gain health equity and maintain quality health for different population groups through the assessment of various demographics such as race/ethnicity and geography (United States Department of Health and Human Services, 2010, and 2013). The study outcome will also be a key source of information to Africans seeking knowledge about African immigrant lung cancer

statistics as well as providing research opportunities to interested epidemiologists wishing to examine the risk factors that influence such within-race disparities.

Chapter 2: Literature Review

In this chapter I will discuss research literature for lung cancer risk factors, incidence rates, treatment methods, mortality rates and stages at diagnosis for better understanding of the sample used and interpretation of the study findings. The review is subdivided into the literature search, epidemiological timeline of lung cancer, lung cancer risk factors, incidence and mortality

Research Literature Search

I searched for full-text research literature by using the major biomedical research databases PubMed Central, PLos One, the CDC's Mortality Morbidity and Mortality Weekly Report database, and the National Library of Medicine. I also searched Google Scholar. The NCBI database includes titles such as *Journal of Cancer Epidemiology*, *Journal of Clinical Oncology*, *Journal of Thoracic Oncology*, *CA: A Cancer Journal for Clinicians*, *Journal of the National Cancer Institute*, , and *Journal of Biomarkers & Prevention*. I restricted my search to articles published within the past 5 years.

I reviewed references in the articles I found for other relevant research. I found no primary research articles comparing lung cancer incidence and mortality and the stages at diagnosis as well as the treatment type received among AMBB and AFBB. Instead, the majority of the articles that I found concerned lung cancer risk factors; a few were epidemiological studies. I selected a total of 120 articles of much relevance to my research questions from the total of 8,803 that my general search returned. I further screened the 120 research articles for those with free access to the full text, and contacted

the Walden University Library for help with access to those with restricted access. Meanwhile I excluded some that were too recent and did not yet have any access from the literature review.

Epidemiological Trend and Evidence of the Etiology: The Timeline

Lung cancer and tobacco smoke. Researchers began to describe malignant growth of the lung and bronchi and the possible etiology in the early 1900s. Adler (1912) reviewed hundreds of autopsy and pathology reports in Europe and the United States. He found an association between malignant growths of the lung and tuberculosis and suggested that subjects' occupation and tobacco smoking might also be involved (Adler, 1912).

After Adler's (1912) report, there were reports of noticeable increase in the incidence of lung cancer in England and Wales from 1921-1932 where the absolute number of lung cancer cases rose from 361 cases in 1921 to 2095 cases in 1932 for men and from 186 cases in 1921 to 680 cases in 1932 for women (Kenneway & Kenneway, 1936). The death ratio for women increased from 1:1.94 in 1921 to 1:3.08 in 1934 (Kenneway, & Kenneway, 1936). Clinically and statistically, Wynder and Graham (1950) demonstrated tobacco smoking as the major contributing factor to this rapid increase. In the cases that they analyzed, Wynder and Graham assessed individuals' occupations, history of previous lung disease, hereditary components, and smoking habits.

The researchers Wynder, & Graham (1950) found out that 98.7% of the 605 cases involved in their study were smokers. Following Wynder & Graham (1950) finding, Doll, & Hill (1950) conducted a study involving a much larger participants pool made up of

1731 lung cancer cases, and control consisting of 1762 patients with cancer-of-other-sites, respiratory and other diseases in England. The Doll, & Hill (1950) study participants were from varied residential cities but had similar access to medical care. Doll & Hill (1950) assessed their smoking history and habits in many different scenarios after eliminating sampling errors and interviewer bias. They considered estimated amount of tobacco that the subjects regularly and recently smoked before illness onset, the approximate maximum amount of tobacco smoked on a daily basis, and the approximate total smoked since starting to smoke. Doll & Hill (1950) then compared lung carcinoma patients and the study control group for each of these three scenarios. They found that heavy smokers made up a significantly larger proportion in the lung carcinoma group; 26.0 % men and 14.6% women regularly smoked 25 or more cigarettes a day as compared to 13.5% men and 0% women in the control group. From their results, they concluded that carcinoma of the lung patients smoked the maximum amount for each scenario than the control, indicating that smoking might have played a role in causing their lung cancer (Doll & Hill (1950). Their study therefore showed that cigarette smoke causes lung cancer.

Doll and Hill as well as other researchers continued to study the association between tobacco smoke and lung cancer, and in 1954 they initiated a prospective study involving 40,564 British mmedical ddoctors as the study subjects (Doll & Hill, 1954). They stratified the subjects during the baseline data collection into continuing smokers, ex-smokers, and never-smokers. They observed at intervals the subsequent changes in smoking habits and lung cancer mortality, as well as other diseases for the study subjects.

Lung cancer death rate increased as the amount of cigarette smoked per day increased after 29 months observation (Doll & Hill, 1954). The lung cancer death rate for nonsmokers was 0.00 per 1000 population, while for those who smoked at least 25g of tobacco or more daily was 1.14 per 1000 population (Doll & Hill, 1954). After 10 years observation, the death rate from lung cancer for continuing smokers doubled that for nonsmokers (Doll & Hill, 1964).

In 1965, Hill compared their findings to that of the report of a study that was done several decades earlier involving scrotal cancer and its association to chimney sweeping (Waldron, 1983; Hill, 1965). He found a similarity in association between each disease and the environmental factors to which those patients were exposed (Hill, 1965). After elimination of chance occurrence, and despite these strong associations observed between lung cancer and smoking; and scrotal cancer and chimney sweep, to conclude that such environmental factors caused the diseases Hill (1965) pointed out a number of aspects of such associations that must be fulfilled. In his address to the Royal Society of Medicine during a meeting, Hill (1965) using the results of their tobacco smoking and lung cancer studies indicated that causation can only be concluded from (1) the strength of the association between the disease and the factor in question in which case such an association must be in excess when compared to the disease association with other possible causes, (2) consistency of such an association, in which case the same result must be obtained if the study is repeated by different researchers at different times and locations, (3) specificity of association in which instance the association must be limited to a particular disease and population at a particular place, where no excess of the disease

is observed without the presence of the factor in consideration, exemplified by the results of the smoking and lung cancer study results (Doll, & Hill, 1964), (4) temporality of association in which case the exposure to the factor must occur before the onset of the disease; (5) Biological gradient exemplified by the fact that in their study, lung cancer death rate rose linearly with increase in amount of cigarette smoked; (6) plausibility; (7) coherence exemplified by the fact that the association between lung cancer and cigarette smoking has been coherent with the histopathology results for the disease in such patients; (8) experiment which he exemplified by the fact that if cigarette smoking is eliminated, will the association between it and lung cancer reduce; and (9) analogy where the effect of similar factors leading to the same kind of association should be considered (Hill, 1965). These have since been known as the Bradford Hill Criteria for causation and are widely applied in epidemiological investigations of causality.

During the 20th year Doll and Hill observation of the study population showed tobacco consumption had markedly reduced, and lung cancer death rate also reduced, but death rate from other diseases did not reduce accordingly (Doll, Peto, 1976). By the 40th year those who ceased smoking by middle age had substantially low risk of dying from lung cancer as compared to those who continued smoking (Doll, Peto, Wheatley, Gray, & Sutherland, 1994)

With these revelations, studies focused on tobacco smoking have revealed its trend and relation to other diseases, including lung cancer incidence and mortality. In the United States, stratifying study cohorts by time periods as 1960s, 1980s and contemporary (2000s) in the comparison of those who smoked to those who never

smoked, showed the relative risks of death from lung cancer in women to be 2.73, 12.65, and 25.66 respectively, and in men to be 12.22, 23.81, and 24.97 respectively (Thun, et al 2013). In 2009, the prevalence of cigarette smoking among adult US population 18 years and older was 20.6% with prevalence of 23.5% for men and 17.9% for women (Centers for Disease Control and Prevention, 2010). For the 5 years period, from 2005-2009, there was no significant difference in cigarette smoking prevalence which was recorded at 20.9% and 20.6% respectively for U.S adults 18 years and older (Centers for Disease Control and Prevention, 2010).

Tobacco use is also common in Africa, exemplified by Ghana, a West African nation where studies have shown that smoking intention among the youths were 27.7% for those exposed to tobacco smoking commercials as compared to 19% for those not exposed to such commercials (Doku, Raisamo, & Wiium, 2012). Among Ghanaian adolescents 14-19 years old, tobacco smoking prevalence of 1.0% was lower when compared with the older adults 60-69 years with prevalence of 6.1% (Owusu-Dabo, Lewis, McNeill, Gilmore, & Britton, 2009, Table 2). But, a later study showed the average daily smoking prevalence among the older adults estimated at 7.6%; meanwhile, the prevalence among men was higher at 11.3% as compared to female with 3.7% (Yawson, et al., 2013, Table 1). The countries, Rwanda, Uganda , Tanzania and Kenya, all located in Eastern Africa have the highest prevalence of cigarette smoking among men of 14.2%, 18.7%, 21.0%, and 22.9% respectively, with the lowest prevalence of 8.0% seen in men of the Western Africa state of Nigeria (Pampel, 2008).

Lung cancer in never-smokers.

***Preexisting diseases as lung cancer risk factors*.** Despite the compelling evidence that has linked tobacco smoking to lung cancer incidence and mortality since the early 20th Century to contemporary times (Wynder, & Graham, 1950; Doll & Hill, 1950, and Thun, et al, 2013), nonsmokers have other lung cancer risk factors too. These include environmental air pollutants and previous inflammatory lung disease (Samet,et al, 2009, and Moldoveanu et al, 2009). Common among the inflammatory lung diseases which could also be seen in smokers are tuberculosis, emphysema, pneumonia and chronic bronchitis. But independent of tobacco smoking these conditions have been studied by the International Lung Cancer Consortium and shown that the relative risk for them to influence lung cancer are 1.48 (95% *CI*: 1.17-1.87), 2.44 (95% *CI*: 1.64 - 3.62), 1.57 (95% *CI*: 1.22 - 2.01), and 1.47 (95% *CI*: 1.29 -1.68) respectively (Brenner, et al., 2012). In a tuberculosis study cohort, 26.3 per 100,000 persons of patients with tuberculosis developed lung cancer; which is 10.9 times more, compared to the 2.41 per 100,000 persons for non-tuberculosis patients who developed lung cancer (Yu, et al. 2011). Another preexisting disease, HIV infection, is also a risk factor for lung cancer with an incidence relative risk of 1.7 (95% *CI*: 1.5 to 1.9) after 37,294 HIV positive U.S Veterans and 75,750 uninfected controls, composed of about 50% non-Hispanic African Americans and 40% non-Hispanic Whites were followed for 5.8 years in a prospective cohort study (Sigel, et al., 2012).

Environmental Air pollutants. Apart from preexisting diseases, chronic and acute exposure to some ambient air pollutants increases the risk for lung cancer (Turner, et al.,

2011). One such air pollutant is silica and those exposed to it have a lifetime risk of 0.51% for lung (Liu, et al., 2013). Nitrogen oxides derived from traffic is also associated with lung cancer with incidence relative risk (*IRR)* of 1.30 (95% *CI*: 1.05–1.61), (Raaschou-Nielsen, et al., 2011).

Incidence and mortality. In South Africa, for the general population, the annual lung cancer mortality did not significantly change as it only minimally decreased from 24.3 per 100,000 population in 1995 to 23.8 per 100,000 population in 2006 (Bello, Fadahun, Kielkowski, & Nelson, 2011). But, by gender, for corresponding years, the rate increased for women from 10.8 per 100,000 populations to 13.4 per 100,000 populations, though it decreased minimally for men from 44.2 per 100,000 populations to 39.4 per 100,000 populations (Bello, et al., 2011). However for the general South African population, there was a significant change for the 5 years' time frame of 2000-2005 with an annual decrease of 129 deaths per 100,000 populations though from 1999-2006 there was a significant annual increase of 0.34 per 100,000 populations for women (Bello, et al., 2011).

Even though 1 out of 4 men in the East African country of Malawi smoke tobacco (Msyamboza, et al., 2011), a study of cancer burden in the country revealed that the least common cancer was lung cancer, and it accounted for only 0.2% of the 18,946 new cancer cases that were registered from 2007-2010 (Msyamboza, et al., 2012). A retrospective study spanning 15 years (1993-2007) identified 1,882 registered lung cancer cases in the North African nation of Tunisia which accounted for 10.9% of all the cancers registered during that period for the general population, but it contributed for 21.7% of all

cancer cases registered for men, and 1.4% for women (Missaoui, et al. 2011). The trend for lung cancer in Tunisia during the study period decreased significantly by - 6.5% (95% *CI*: -12.9% to - 0.2%), (Missaoui, et al. 2011).

In a comparison of U.S non-Hispanic Black and White population aged 20 to 39 from 1992-2006 the death rate from lung cancer per year decreased by 3.6% in white men and 7.9% in Black men, meanwhile, the decrease was 1.9% in White women and 4.8% in Black women (Jemal, Center, & Ward 2009). There has been a marked reduction in the prevalence of tobacco smoking in the United States, more so for women and men of the West with average of 33.3% and 28.5% respectively while the smallest decrease is in the Midwest with 20.3% and 18.6% respectively (Jemal, et al., 2011). Yet, in most U.S. States, lung cancer mortality rates for White women have been on the rise except for California where there has been a decline (Jemal, et al., 2012).

Chapter 3: Research Method

In this chapter I will discuss the study objectives, research questions, and hypotheses. I will also describe the study method to provide a clear understanding of the research questions and hypotheses, the methods that I will use to obtain the data, the sample size calculation and instrumentation, as well as the statistical test methods for the data analysis.

Study Objectives

Reflecting on this within-race study, as stated in the research questions below the objective is two folds. Firstly, the objective is to gather lung cancer data collected from 2004-2011 for AFBB and AMBB; and secondly, to use the acquired data to answer the research questions (RQ) following the stated hypotheses (H_0) and alternative hypotheses (H_1). The populations in the study are AFBB and AMBB who have been diagnosed as having NSCLC, and recorded in the SEER database.

Research Questions, Hypotheses, and Alternative Hypotheses.

I used the following research questions to guide my investigation:

RQ1: Do the proportions of NSCLC stage (I, II, III and IV) in the SEER database for 2004-2011 differ between AFBB and AMBB patients?

$H_0$1: There is no significant difference between the proportion of AFBB and AMBB diagnosed with NSCLC by stage (I, II, III and IV) and reported to SEER between 2004 and 2011.

$H_1$1 There is a significant difference between the proportion of AFBB and AMBB diagnosed with NSCLC by stage (I, II, III and IV) and reported to SEER between 2004-2011.

RQ2: Among Black patients diagnosed with early stage (I, II) NSCLC reported to SEER between 2004-2011, does the proportion of AFBB patients who underwent surgery differ from the corresponding proportion of AMBB patients?

$H_0$2: There is no significant difference between the proportion of AFBB early stage (I, II) NSCLC who underwent surgery and the corresponding proportion of AMBB patients that were reported to SEER between 2004-2011

$H_1$2: There is a significant difference between the proportion of AFBB early stage (I, II) NSCLC who underwent surgery and the corresponding proportion of AMBB patients reported to SEER between 2004-2011

RQ3: Is there a significant difference in the proportion of AFBB NSCLC (stages I, II, III and IV) patients treated with radiation to the corresponding proportion of AMBB patients, reported to SEER from 2004-2011?

$H_0$3: The proportion of AFBB NSCLC (stages I, II, III and IV) patients treated with radiation differs from the corresponding proportion of AMBB patients, reported to SEER from 2004-2011.

$H_1$3: The proportion of AFBB NSCLC (stages I, II, III and IV) patients treated with radiation differs from the corresponding proportion of AMBB patients, reported to SEER from 2004-2011.

RQ4: Among the early stage (I, II) NSCLC patients reported between 2004 and 2011, does the proportion of AFBB patients who received neither surgery nor radiation differ significantly from the corresponding proportion of AMBB patients?

$H_0$4: The proportion of AFBB early stage (I, II) NSCLC patients who received neither surgery nor radiation does not differ significantly from the corresponding proportion of AMBB patients, reported between 2004-2011.

$H_1$4: The proportion of AFBB early stage (I, II) NSCLC patients who received neither surgery nor radiation differs from the corresponding proportion of AMBB patients, reported between 2004-2011; there is no significant difference in their proportion

Research Population and Setting.

Based on my research questions and hypotheses, I limited eligibility of participants to patients who are Black and who were born in the United States or an African country or region. I excluded cases with unknown values for age, gender, stage at diagnosis, marital status, country of birth, radiation, and surgical treatment methods.

Sample Size Determination. For this retrospective study, I determined the sample size using the Open Epi software calculator. I was interested in analyzing both male and female AFBB and AMBB cases. I based my sample size on a number of risk ratios (1.2, 1.5, and 1.75), a 95% *CI*, and the chance of detecting a meaningful outcome or the conventional statistical power for the study at least 80% (Fosgate, 2009). From information gathered from the Georgia Comprehensive Cancer Registry (GCCR), one of the SEER Program Registries, the ratio of AFBB lung cancer cases to AMBB lung cancer

cases is approximately 1:20. I used this ratio in the Open Epi software calculator for the "Ratio of unexposed to exposed" in the sample size calculation. For the pair of sample size calculation done, the percent of exposed with outcome is assumed at four possible values being 20%, 30%, 40%, and 50%; and the risk ratio at three different levels being 1.2, 1.5 and 1.75. I used these figures in order to generate a wide range of eligible cases within which to choose the best sample size. Table 1 shows the sample size ratio for AFBB to AMBB patients with lung cancer. The ratio ranges from a low of 15:293 to a high of 850:16984.

Table 1 *Sample Size Ratios for Lung Cancer AFBB to AMBB Subjects*

% Unexposed with outcome	RR=1.2	RR=1.5	RR=1.75
20	850/16984	136/2713	61/1210
30	491/9810	78/1557	35/697
40	317/6332	50/997	23/445
50	207/4121	33/669	15/293

Note. Assumptions: power = 80%, two sided type error = 0.05
RR is Risk ratio

This AFBB to AMBB ratio of 1:20 calculated using the original sample presents a very wide size discrepancy which will inevitably influence a significant difference in the calculated proportions. Therefore, from the AMBB population a random sample shall be approximated to reduce the sample size ratio to 2 AMBB: 1 AFBB. In this process, patients' age and gender which are confounding variables to the dependent variables shall be used to determine the sampling method for selecting the AMBB study sample. One of

two possible methods shall be used depending on the outcome of the preliminary population age and gender proportion comparisons for the two groups. If the population age and gender proportions comparisons yield a significant difference, between AFBB and AMBB, then from the unsorted AMBB original population, a number of cases that double the number of AFBB shall be selected from the top of the list. A second method of sampling if the preliminary population age and gender proportions comparison for AFBB and AMBB do not yield any significant difference will be purposive sampling. In this case 2 AMBB subjects shall be matched to 1 AFBB for all the AFBB subjects, by age and gender.

***SEER data source*.** The required secondary data variables for this study are all included in the SEER primary data information collected, which fall into 3 categories: (1) Cancer identification which include the primary site, stage, diagnostic procedure and date; (2) patient demographics to include the race, gender, age, and birth place, but void of personal identifiers; and (3) NSCLC disease treatment and outcome of the patients (National Cancer Institute, 2014b).

Ethical consideration. For the SEER Program, patient confidentiality is highly respected during primary data collection. The gathering, reporting and accessing of the data are guided by specific laws, and are only allowable to eligible individuals (National Cancer Institute, 2014c). Furthermore, personal identifiers are coded before any transmission is done (National Cancer Institute, 2014c). Thus, the data for this study shall contain assigned or coded identities for each patient, rather than specific patient's identifiers. To further protect the patients' privacy, during abstracting of the lung cancer

data for this study, the SEER database shall be accessed following stipulated guidelines and permission from the SEER Program. According to the SEER Program website, among all the primary data information collected, some basic ones such as cancer type and stage, date of cancer diagnosis, race, sex, age, type of treatment, and the date and status of last follow up are readily available to be abstracted (National Cancer Institute, 2014d). Among these available information those appropriate to create the NSCLC study frequency data shall include the Black race cases, the year of diagnosis, sex, age, place of birth, treatment received, and marital status. The country of birth shall be decoded, and for those born in Africa, selection shall go by names of specific countries as well as the African Region (Northern, Southern, Middle, Eastern and Western Africa). For those born in the United States, all the SEER's State-of-birth codes (National Cancer Institute, 2014e) shall be recoded to the one code assigned to the country of birth as the United States of America (U.S.A.); then the SEER's Government Services Administration codes for African Country of Birth (National Cancer Institute, 2013, and National Cancer Institute, 2014e). These abstracted data shall be presented in a table, void of personal identifiers; and, being a secondary data analysis, there shall be no direct contact with the involved subjects.

Statistical Analysis of Data

For the data, the total number of NSCLC cases will serve as the subjects under study. Then, the data analysis shall be performed using the Statistical Package for the Social Sciences (SPSS) software (IBM Corporation, 2014). The statistical test to answer each of the four questions RQ1 through RQ4 will be the Chi-square test; and each p-

value obtained compared to the alpha (0.05) to determine whether or not to reject the null hypothesis. This statistical test is suitable because each of the questions involves the difference between 2 groups only. The categorical independent variables is place of birth of Black race (represented by AFBB and AMBB), and the dependent variables are the stage-at-diagnosis of NSCLC; and the type of treatment received. The assumption here is that for each of the research questions, the 2 sets of samples are random and independent; and that the 2 populations from which the 2 groups, AFBB and AMBB NSCLC cases are derived are normally distributed. It is also assumed that the test is going to be a 2-tailed test, meaning that there will be two rejection regions, one on each tail of the assumed normal distribution. This software comparison of the 2 groups' proportions using the Chi square test will yield the p (or 2-tailed significance) values from which the 2 samples proportions differences can be interpreted by comparison to the critical p-value. Other possible predictors of the outcome under study include age, and marital status which can be possible confounding variables. For any of the Chi square tests that show a significant difference between the two groups, a logistic regression analysis shall be performed. For this test, the study subjects' ages gender and the marital status shall be used as covariates to the place of birth.

Study Limitations. The potential limitations to this study will include but not limited to missing data due to unclassified cases by place of birth, inability to control for other relevant covariates such as culture, preexisting health condition, and type of healthcare system involved, lack of data collected for the duration of stay either in the

United States or country of birth for the AFBB population and duration of prior out-of-U.S. stay for both the AMBB and AFBB.

Chapter 4: Results

Introduction to data analysis and Results

In this chapter, I present quantitative analyses of the NSCLC cases in order to profile the two study groups (AMBB and AFBB patients) and enable comparisons of their demographic characteristics, stage at diagnosis, and treatments. I matched the AMBB cases to the AFBB on their age and gender. I used a specific statistical tool and techniques to investigate the research questions and hypotheses and to predict the treatment type received. I also used logistic modeling. I performed the descriptive statistics for both samples prior to and after the matching process. Furthermore, after descriptive, Pearson's chi-square are used to test the study's four hypotheses and the chi-square value and the p-value are used to make decisions regarding the hypotheses. In order to make a decision to either accept or reject the null hypotheses, I compared the obtained *p* values to the alpha value of 0.05. Then, I estimated the binary logistic regressions to investigate which factors best predicted surgical and radiation treatment. Finally, I have presented a summary of the results at the end of the chapter.

Data Acquisition and Recoding. I used the SEER.STAT software to query the data from the SEER-18 Registries database. The data query process was not straightforward due to the large number of differential codes used for the cancer histologic types and stages. I sought help from SEER by phone call and I asked for a complete list of the codes and for help in transferring the queried dataset from the SEER database into an Excel spread sheet. With the queried dataset, I recoded the age ranges, marital status, surgical therapy, radiation therapy, the country of birth to align with the

study questions. There were 135 AFBB and 25,139 AMBB. I eliminated 16 AFBB and 2846 AMBB cases because of missing values. This process yielded 119 AFBB and 22293 AMBB eligible cases using my study criteria, as outlined in Chapter 3. I then recategorized participants using 10-year intervals. Table 2 shows the age distribution of the cases. It excludes the 0-29 years old age group since that range had no AFBB cases. My recoding produced 5 age categories: 30-39 years, 40-49 years, 50-59 years, 60-69 years, and >69 years. To provide more simplicity than that provided by SEER, I recoded marital status into four categories: single, married, Divorced, and Widowed. I classified the NSCLC stages at diagnosis as "Early Stage" (to include stages I and II) and "Late stage" (to include stages III and IV). Also, I dichotomized the therapy types received were surgery and radiation of which surgical therapy to "Yes", if any of the 24 different types of surgical procedures reported was received, and "No" if none of the 24 different types was received. In the same way, I dichotomized radiation therapy to "Yes" to incorporate any of the various forms of radiation procedures received, and "No" if no form of radiation treatment was received. I renamed birth country as United States for AMBB patients and Africa for AFBB patients. This categorization was simpler and easy to understand the different groups and variables.

Sampling of Data. Among the 135 AFBB and 25139 AMBB SEER cases for 2004-2011, 16 AFBB and 2846 AMBB cases were missing some values. I eliminated these cases from the study dataset, which left 119 AFBB and 22293 AMBB NSCLC cases for the study sample. This left a study cohort with very large population ratio difference. I then considered a 1 AFBB to 2 AMBB sampling to reduce the sample size

difference and eliminate statistical differences that the large size difference would influence. In this case, I considered age and gender as the prime confounding variables and I used them for the match. Thus, I compared the age and gender distribution for AFBB and AMBB prior to the sampling, in order to determine if there were any age and gender distributions differences between the two groups prior to sampling.

Age distribution prior to sampling AMBB study subjects. For the 5 age groups comparison, the result of the analysis is presented in Table 2 below.

Table 2 *Age Distribution for AFBB and AMBB Study Subjects*

Age (Years)	Birth country		
	AMBB	AFBB	p value
30-39	139 (1%)	7 (6%)	0.00*
40-49	1572 (7%)	12 (10%)	0.2
50-59	5229 (24%)	38 (32%)	0.03*
60-69	7100 (32%)	43 (36%)	0.32
> 69	8253 (37%)	19 (16%)	0.00*
Total	22293 (100%)	119 (100%)	

Note. Percent values were rounded.
*$p < 0.05$

Table 2 shows that there is a significant difference in age categories. Of the five age categories, three are significantly different between the AFBB and AMBB. AFBB patients were comparatively younger than the AMBB patients.

Gender distribution prior to sampling AMBB study subjects. The result of gender distribution is presented in Table 3 below. It shows that the gender distribution between AMBB and AFBB patients was not significantly different.

Table 3 *Gender distribution for all AFBB and prior to sampling AMBB study subjects*

	Gender	Birth Country (USA or Africa)		
		AMBB	AFBB	p-value
	Female	9089(41%)	59(50%)	0.05
	Male	13204(59%)	60(50%)	0.05
Total		22293(100%)	119(100%)	

Note. Percent values were rounded.
*$p < 0.05$

Conclusively, for the before sampling test for age and gender distribution, the analysis found that the age distribution was significantly different between AMBB and AFBB cases prior to the sampling. But, for gender, the distribution was not significantly different. Table 3, shows that for AFBB the gender was 50% male and 50% female; while for AMBB the gender was 59% male and 41% female; this difference in gender distribution was not statistically significant. Random sampling for this study was then performed based on this result. Due to the low frequency count of 119 for AFBB I made a decision to include all 119 in the study. I did the sampling following the results of the age and gender distribution tests. From the non-ordered original 22293 AMBB data I selected the first 238 so as to provide 2 AMBB for each of the AFBB study subjects. I entered the study cohort consisting of 119 AFBB cases and 238 AMBB cases into SPSS and appropriately coded for the study hypotheses assessment tests.

Descriptive Statistics.

Marital Status of the NSCLC cases. After recoding the marital status for this study sample as "single", "married", "divorced" and 'widowed", I computed the NSCLC disease frequency and percentages by marital status and presented the results in Table 4. Among the AFBB patients, 25% were single, while 24% of AMBB were single, and there is no significant difference between the two groups considering the single status. However, about 35% of AMBB cases were married, while 57% of the AFBB were married, this shows the proportion of AFBB cases that were married are higher, and there is a significant difference in proportion for married cases between AFBB and AMBB patients ($p < 0.05$).

Table 4 *Marital Status of AMBB and AFBB Subjects*

		Birthplace (Africa or USA)		
		AMBB	AFBB	p-value
Marrital Status	Single	57(24%)	30(25%)	0.78
	Married	82(35%)	68(57%)	<0.05*
	Divorced	54(23%)	8(7%)	<0.01*
	Widowed	45(19%)	13(11%)	0.05
Total		238(100%)	119(100%)	

Note. Percent values were rounded.
*$p < 0.05$

Table 4 also shows that for AFBB, about 7% were divorced while for AMBB 23% were divorced patients. This indicates that the proportion of divorced AMBB patients is considerably higher compared to AFBB, and these proportions are significantly different ($p < 0.05$) as shown in Table 4. Finally, 11% of the AFBB patients were widowed and

19% of AMBB were widowed, and these proportions were not significantly different as shown on Table 4.

NSCLC Stage at diagnosis. The NSCLC stage at diagnosis by birth place is presented in Table 5. Here, for AMBB patients 25% of the patients were at early stage, while for AFBB it was 24%.

Table 5 *NSCLC Stage at diagnosis by birth place (Africa or USA) for study subjects*

		Birth place (Africa or USA)		
		AMBB	AFBB	P-value
NSCLC Stage at diagnosis	Early Stage	60(25%)	28(24%)	0.72
	Late stage	178(75%)	91(77%)	0.72
Total		238(100%)	119(100%)	

Note. Percent values were rounded.
Critical value: $p > 0.05$

Also, 75% of the AMBB patients were diagnosed at late stage, while 77% of AFBB were at late stage. The test of proportion for the NSCLC diagnosis at late stage shows that there is no significant difference between AMBB and AFBB patients as presented on Table 5.

Surgical Treatment. The patients received two types of therapies, surgery and radiation. Table 6 illustrates the information regarding surgical therapy of the study subjects data set. As presented, 29% of the AFBB patients received surgical therapy, whereas 20% of the AMBB patients had surgical therapy as treatment, and these

proportions are significantly different ($p < 0.05$). Furthermore, as presented on Table 6, the proportion of AFBB patients that did not have surgery compared to the corresponding proportion of AMBB patients were significantly different ($p < 0.05$). The same conclusion applies to the proportions AFBB and AMBB that received surgical therapy ($p < 0.05$) as shown on Table 6.

Table 6 *Surgical therapy by birthplace (Africa or USA) for study subjects*

		Birth place		
		AMBB	AFBB	P-value
Surgical Therapy	No	191(80%)	84(71%)	0.04*
	Yes	47(20%)	35(29%)	0.04*
Total		238(100%)	119(100%)	

Note. Percent values were rounded
*critical value: $p < 0.05$

Radiation Treatment. Table 7 shows that among AMBB patients 50% did not receive radiation therapy, compared with a corresponding 63% of AFBB patients. Statistically, these proportions are significantly different ($p<0.05$). Additionally, 37% of the AFBB patients received radiation therapy compared with a corresponding 50% of the AMBB patients. These proportions are significantly different ($p < 0.05$).

Table 7 *Radiation therapy by birthplace (Africa or USA) for study subjects*

		Birth place (Africa or USA)		
		AMBB	AFBB	P-value
Radiation Therapy	No	118(50%)	75(63%)	0.02*
	Yes	120(50%)	44(37%)	0.02*
Total		238(100%)	119(100%)	

Note. Percent values were rounded
*critical value: $p < 0.05$

Hypotheses Testing.

Based on the four research questions (RQ), the null hypothesis (H_0) and alternative hypotheses (H_1) for each corresponding RQ were formulated in order to objectively answer the questions. The alpha or predetermined level of statistical significance in this study was 0.05 to which all the p-values were compared to determine whether or not to reject the null hypothesis. In each case, I used the Chi-square test of independence to determine which hypothesis to accept and which to reject. The hypotheses test results are discussed in this section.

In order to answer RQ1 and judge the hypotheses objectively, I applied the Pearson Chi-Square Analysis. Table 8 presents the expected and observed counts for each category of NSCLC stage for the AFBB and AMBB patients. I assessed the difference

between the observed count and expected counts and then calculated the p-value for each proportion. The result of Pearson Chi-Square Analysis is also presented in Table 8.

Table 8 *Pearson Chi-Square to assess the difference in NSCLC Stage at Diagnosis for AMBB and AFBB patients*

			Birth place (Africa or USA)		
			AMBB	AFBB	p-value
NSCLC Stage at Diagnosis	Early Stage	Count	60(25%)	28(24%)	0.72*
		Expected Count	58.7	29.3	
	Late stage	Count	178(75%)	91(77%)	0.72*
		Expected Count	179.3	89.7	
	Total	Count	238(100%)	119(100%)	

Note. Percent values were rounded
Pearson Chi-Square, $X^2 = 0.1206$
Degree of freedom = 1
*Critical Value: $p > 0.05$

The result of the Pearson Chi-Square test indicates that there is no significant difference between the proportion of AFBB and AMBB diagnosed with NSCLC by stage (early and late) and reported to SEER from 2004-2011.

To answer the RQ2 and assess the hypotheses objectively, I applied the Pearson Chi-Square Analysis. Table 9 is the contingency table which presents the expected and observed counts for each category by NSCLC early stage for the AFBB and AMBB patients who underwent surgery.

Table 9 *Pearson Chi-Square to assess the differences between AMBB and AFBB patients diagnosed with early stage NSCLC who received surgical therapy*

			Birth place (Africa or USA)		p-value
			AMBB	AFBB	
Surgical Therapy	No	Count	28(47%)	6(21%)	0.02*
		Expected Count	23.2	10.8	
	Yes	Count	32(53%)	22(79%)	0.02*
		Expected Count	36.8	19.7	
	Total	Count	60(100%)	28(100%)	

Note. Percent values were rounded.
Pearson Chi-Square, X2 = 5.129
Degree of freedom = 1
*Critical Value: $p < 0.05$

Table 9 shows that there is a significant difference between the proportion of AFBB diagnosed with early stage (I, II) NSCLC and the corresponding proportion of AMBB patients who underwent surgery ($p < 0.05$). This means that the null hypothesis is rejected and the alternative hypothesis is accepted. As a whole, for surgical therapy as shown in Table 9, the proportion of AFBB patients who underwent surgery differs from the corresponding proportion of AMBB patients ($p < 0.05$). AFBB patients were found to be more likely to have received surgical therapy compared with AMBB patients.

To objectively answer RQ3 by assessing the hypotheses, I applied the cross-tab Pearson Chi-Square tests, first to those that were diagnosed at the early as well as the late stages and then to the overall disease stages at diagnosis, early and late combined. Amongst the patients diagnosed with NSCLC at the early stages, there was a significant

difference in the receipt of radiation therapy ($p < 0.05$) between AMBB and AFBB. The proportion of AMBB patients (37%) that received radiation treatment was higher compared with the corresponding proportion of AFBB patients (14%) as shown on Table 10. But, for those diagnosed at the late stages of the disease, there was no significant difference in the proportion of AMBB (55%) that received radiation therapy when compared with the corresponding proportion of AFBB patients (44%) as shown on Table 11. When I compared the two groups for radiation treatment for the overall disease stages at diagnosis, early and late combined the result of the analysis is summarized and presented in Table 12. This Table 12 shows that per the Chi-Square analysis, the null hypothesis is rejected since the proportion of AFBB NSCLC (stages I, II, III and IV) patients treated with radiation therapy does differ significantly from the corresponding proportion of AMBB patients, reported to SEER from 2004-2011 ($p < 0.05$).

Table 10 *Pearson Chi-Square to assess the difference in radiation therapy between AMBB and AFBB patients diagnosed at early stage NSCLC*

			Birthplace (Africa or USA)		p-value
			AMBB	AFBB	
Radiation Therapy (early stage)	No	Count	38(63%)	24(86%)	0.03*
		Expected Count	42.3	19.7	
	Yes	Count	22(37%)	4(14%)	0.03*
		Expected Count	17.7	8.3	
	Total	Count	60(100%)	28(100%)	

Note. Percent values were rounded.
Pearson Chi-Square, X2 = 4.594
Degree of freedom = 1
*Critical Value: $p < 0.05$

Table 11 *Pearson Chi-Square to assess the difference in radiation therapy between AMBB and AFBB patients diagnosed at late stage NSCLC*

			Birthplace (Africa or USA)		p-value
			AMBB	AFBB	
Radiation Therapy (late stage)	No	Count	80(45%)	51(56%)	0.08*
		Expected Count	86.7	44.3	
	Yes	Count	98(55%)	40(44%)	0.08*
		Expected Count	91.3	46.7	
	Total	Count	178 100%)	91(100%)	

Note. Percent values were rounded.
Pearson Chi-Square, $X^2 = 2.97$
Degree of freedom = 1
*Critical Value: $p > 0.05$

Table 12 *Pearson Chi-Square to assess the difference in radiation therapy between AMBB and AFBB patients for all stages (Early and Late)*

			Birthplace (Africa or USA)		p-value
			AMBB	AFBB	
Radiation Therapy (all stages)	No	Count	118(50%)	75(63%)	0.02*
		Expected Count	128.7	64.3	
	Yes	Count	120(50%)	44(37%)	0.02*
		Expected Count	Table 119.3	54.7	
	Total	Count	238(Table 110%)	119(Table 110%)	

Note. Percent values were rounded.
Pearson Chi-Square, X2 = 5.77
Degree of freedom = 1
*Critical Value: $p < 0.05$

As presented in Table 12, among AMBB patients 50% did not receive radiation therapy, compared to a corresponding 63% of AFBB patients. Statistically, these proportions are significantly different ($p < 0.05$) as shown on Table 12with the corresponding 50% of the AMBB patients; and this difference was found to be statistically significant ($p < 0.05$). AMBB patients were more likely to receive radiation therapy at all the stages (I, II, III and IV) of the NSCLC than AFBB patients.

For research question 4 (RQ 4), I queried the data and obtained a sample with total of eight early stage (I, II) NSCLC patients who received neither surgery nor radiation for both AFBB and AMBB. This was composed of 1 AFBB and 7 AMBB cases as shown in the Appendix. The attempt to use any statistical tools to test the hypothesis

was unsuccessful. Therefore, the hypotheses could not be tested and so the research question was left unanswered.

Logistic Regression Analysis

In modeling the relation between treatment types and demographic variables in relation to country of birth of the patients, I conducted the analysis for both surgical and radiation treatment. I used the variables available on the SEER database that are possibly responsible for influencing the receipt of surgery or radiation therapy as a treatment choice for the analysis. These include age, gender, marital status, country of birth (Africa or USA), and the stage at diagnosis for the NSCLC. For both surgery and radiation treatments the regression results are discussed in the sections below.

Surgery treatment prediction. I performed a logistic regression analysis to ascertain the effect of age, gender, marital status, country of birth, and stage at diagnosis for the NSCLC in predicting the probability of a patient receiving surgery as treatment. In this case, the dependent variable is the surgery, dichotomized into "No Surgery" and "Surgery". In the SPSS data analysis software I coded "No Surgery" as "1" and Surgery as "2". Meanwhile, I coded the independent variables as follows: age of the patient at diagnosis (5 categories), gender (Female =1, Male =2), Marital Status (single = 1, married = 2, divorced = 3 and widowed = 4), Country of birth (AMBB = 1, AFBB = 2), and Stage at diagnosis (1= Early, 2 = Late). The result of the model is presented in summary in Table 13 below.

Table 13 *Logistic model summary for surgery treatment*

Independent Variables	Beta (unstandardized)	S.E.	df	Sig./ P-value	Exp(B) (AFBB/AMBB)	95.0% *CI* for EXP(B)	
						Lower	Upper
Age	-0.279	0.156	1	0.075	0.757	0.557	1.028
Gender	-0.335	0.309	1	0.278	0.716	0.391	1.31
Marital Status	0.023	0.163	1	0.89	1.023	0.743	1.408
Stage	-2.806	0.32	1	0	0.06	0.032	0.113
Country of birth (USA or Africa)	0.663	0.326	1	0.042	1.94	1.024	3.677
Constant	3.992	1.084	1	0	54.156		

Note. Description of table labels

Model $\chi 2 = 99.86$, $p<0.05$

Mode $df = 5$

Pseudo R2 = 0.369

$N = 357$

The Beta (un- standardized) represents the coefficients for the independent variables.

S.E. is the standard error of the beta values;

Df is the degree of freedom,

Sig. is the p-value to show the significance of the independent variables in the model,

Exp (B), is the Odds Ratio (*OR*)

The logistic regression model was statistically significant ($p < 0.05$). The model explained 36.9% of the variance (R^2) in surgical treatment for the NSCLC.

In this model, stages at diagnosis and country of birth (USA or Africa) have significant influence on the model ($p < 0.05$). In the model, the most important variable turns out to be the stage at diagnosis which has a negative influence on the chance of a patient receiving surgery, (*OR* (AFBB/AMBB) = 0.06; 95% *CI* 0.032-0.11) as shown on Table 13. This indicates that for one unit increase in stage variable (Early stage, to Late stage)

here would be a 0.06 times reduction in the probability for a patient to receive surgery. In other words, if the stage goes from early to late, there would be 94% reduction in the likelihood of having surgery. This implies that late stage patients rarely ever have surgery as a treatment choice, which is as expected, considering the guidelines or recommendations for NSCLC treatment (National Cancer Institute, 2015; and National Comprehensive Cancer Network, 2015). Also in this model, country of birth (USA or Africa) has a positive beta value (*OR* (AFBB/AMBB) = 1.94; 95% *CI* 1.02-3.68) as shown on Table 13. This means that AFBB patients would receive surgery about 1.94 times more compared to AMBB patients.

The summary of this regression analysis shows that the stage at diagnosis is one important factor in determining if a patient will receive surgical treatment, with surgery most likely to happen at early stage of NSCLC than at late stage. Furthermore, AFBB patients are more likely to receive surgery compared to the AMBB patients.

Radiation treatment prediction. I performed another logistic regression analysis to determine the effect of age, gender, marital status, country of birth, and the disease stage in predicting the probability of a patient receiving radiation as treatment. In this case, the dependent variable is radiation treatment, dichotomized as "No radiation", and "radiation", respectively coded in SPSS as "0" and "1". I entered the independent variables, age (with 5 categories), gender (1 = Female, 2 = Male), Marital Status (single = 1, married = 2, divorced = 3 and widowed = 4), Country of birth (AMBB= 1, AFBB = 2) and Stage (1= Early, 2 = Late) in the model. The result summary of the model is presented in Table 14 below.

Table 14 *Logistic model summary for radiation treatment*

Independent Variables	Beta (unstandardized)	S.E.	df	Sig./P-value	Exp(B) (AFBB/AMBB)	95.0% *CI* for EXP(B)	
						Lower	Upper
Age	0.019	0.113	1	0.88	1.019	0.817	1.272
Gender	0.082	0.225	1	0.71	1.086	0.699	1.688
Marital Status	0.087	0.116	1	0.46	1.091	0.869	1.37
Stage at diagnosis	0.947	0.267	1	0	2.578	1.528	4.35
Country of birth (USA or Africa)	-0.544	0.246	1	0.03	0.581	0.358	0.941
Constant	-1.512	0.845	1	0.07	0.22		

Note. Description of table labels
Model $\chi 2 = 20.17$
Model $df = 5$
$n = 357$
The logistic regression model was statistically significant, $p < 0.05$.
The model explains 7.3% of the variance in surgery probability based on the independent variables.
The Beta (un- standardized) represents the coefficients for the independent variables.
S.E. is the standard error of the beta values;
df is the degree of freedom,
Sig. is the P-value to show the significance of the independent variables in the model,
Exp (B) is the Odds Ratio (*OR*)

In this particular model in predicting radiation treatment probability, Stage at diagnosis and Country of birth (USA or Africa) are the variables that are significant as predictors with the respective p-values of 0.00 and 0.03. The other 3 variables; age, gender, and marital status are not statistically significant in predicting radiation treatment. The most important variable in this model is the "Stage at diagnosis", which has a positive influence on receiving radiation as a form of treatment. Correspondingly, for early to late stage, there would be 2.58 times increase in the probability of receiving radiation as

treatment. This implies that late stage patients would have 2.58 times higher probability of receiving radiation treatment than early stage. This is an expected result given the NSCLC treatment guidelines or recommendations that early stage NSCLC patients should get radiation therapy only when they cannot have surgery, or be treated by radiation therapy in combination with surgery for clinical trial purposes (Howington, Blum, Chang, Balekian, & Murthy, 2013). Meanwhile, the late stage disease patients are to receive radiation therapy without any restrictions (National Cancer Institute, 2015). Additionally, the country of birth (USA or Africa) of the patients in the model shows that AMBB patients will receive radiation therapy more compared to AFBB patients. The regression analysis for radiation treatment prediction indicates that, for radiation therapy the patients are only dependent upon the stage of NSCLC and country of birth. Other demographics included in the model (age, gender, marital status) were not predictive of radiation treatment. Conclusively, at the late stage of NSCLC, there is very low probability of receiving surgery; meanwhile there is higher probability for radiation treatment, which is an expected result, following the NSCLC treatment guidelines or recommendations (National Cancer Institute, 2015; and National Comprehensive Cancer Network, 2015).

Results summary to the research questions.

Research Question 1. The results indicate that statistically there is no significant difference in the proportions of AFBB and AMBB patients for the NSCLC stage (early and late) at diagnosis as shown in Table 8.

***Research Question 2*.** The results show in Table 6 that the proportions of AFBB patients who underwent surgery differ significantly from the corresponding proportion of AMBB patients ($p < 0.05$), the AFBB patients were more likely to have surgical treatment compared to AMBB patients. There are no previous studies available that compared these two populations; therefore this result cannot be compared to any other and so must be interpreted with caution.

***Research Question 3*.** The analysis found a significant difference in the proportion of AFBB NSCLC (stages I, II, III and IV) patients treated with radiation to the corresponding proportion of AMBB patients ($p < 0.05$) as shown on Table 12, where the AMBB patients were more likely to receive radiation therapy compared to AFBB patients.

Research Question 4. Among the early stage (I, II) NSCLC patients reported to SEER between 2004 and 2011, does the proportion of AFBB patients who received neither surgery nor radiation differ significantly from the corresponding proportion of AMBB patients? I did not perform any statistical analysis for this research question because the subgroup sample counts were very low as shown in Appendix.

***Regression model for surgery*.** In predicting surgery as a treatment, the stage at diagnosis of the NSCLC is the most significant predictor followed by the country of birth (USA or Africa). Patients diagnosed with early stage NSCLC have a higher probability of receiving surgery as a treatment than those diagnosed at late stage of the disease, which is as expected, given the recommendations for treatment of NSCLC patients (National Cancer Institute, 2015; and National Comprehensive Cancer Network, 2015). In addition,

The AFBB patients were younger than AMBB patients as shown on Table 2, and the AFBB patients had a higher probability of getting surgery than AMBB patients (OR [AFBB/AMBB] = 1.94; 95% CI 1.02-3.68) as shown on Table 13.

Regression model for radiation. The variable that seems to influence radiation treatment more is the stage at diagnosis of the patient. In this case with diagnosis at late stage, the NCSLC patients are administered radiation therapy more, compared to the early stage which is an expected outcome due to the guidelines or recommendations for which treatment is suitable for what stage of the disease (National Cancer Institute, 2015; and National Comprehensive Cancer Network 2015). Also, the country of birth (USA or Africa) is an important predictor; here the AMBB patients are more likely to receive radiation therapy compared to AFBB patients. (OR [AFBB/AMBB] = 0.58; CI 0.36 – 0.94) as shown on Table 14. Other demographics tested showed no influence on predicting the radiation therapy receipt.

Summary

In this chapter, I used descriptive, hypotheses and regression models to analyze the NCSLC patient data. The study used all of the AFBB and a sample drawn from the AMBB population for the analysis to obtain the statistical tests results. I obtained the AMBB sample using a random sampling method after doing a comparative age and gender proportions of the two groups void of those with missing data values. The study found significant differences between the proportions of AMBB and AFBB NSCLC cases ($p < 0.05$) in 3 out of the 5 age categories, as shown in Table 2. But, the difference in gender distribution was not statistically significant as shown in Table 3. The study also

modeled the relationship for different treatment methods (surgery and radiation) and found that the stage at diagnosis and country of birth (USA or Africa) are important predictors of surgical and radiation treatments options.

Chapter 5: Discussion, Conclusions, and Recommendations

Strengths and Weaknesses

This study's strength lies in comparison of NSCLC patients who are of the same race, AFBB and AMBB. In conducting this study, I strove to increase the knowledge about within-race similarities and differences in relation to where NSCLC patients were born, in Africa or in the United States. I did not have access to some data including socioeconomic status and health insurance, which are important factors for lung cancer disease outcome (Elchoufaniz, et al, 2013). Thus, one limitation to this study is that I was not able to test the effects of these factors on patients' stage at diagnosis and treatment they received. In addition, the sample size was limited by the number of AFBB that were diagnosed with NSCLC from 2004-2011. The low sample size for AFBB was further affected by the number of cases with missing dependent variable data which I eliminated from the study. I ignored the probability sample that I initially calculated for this study of the size discrepancy between the two groups, AFBB (n= 135) and AMBB (n = 22293). Meanwhile, from the total AMBB population, I performed an approximate random sampling to obtain the study sample size and thus preserving the representativeness of the AMBB selected study sample. Thus, it made the study more valid given the assured representativeness of the samples selected from the total AMBB (n = 22293) group. The lack of adequate number of sample for each group of early stage (I and II) NSCLC patients (AFBB = 1 and AMBB = 7) as shown on Appendix A that received neither surgery nor radiation as asked in RQ 4 was a hindrance in answering the research question.

Conclusion

Using the chi-square tests and the logistic regression analysis, I found no significant difference in the proportions of AFBB and AMBB patients at early and late stages of NSCLC diagnosis. However, the proportions of AFBB patients who underwent surgery differed significantly from the corresponding proportion of AMBB patients ($p < 0.05$; see Table 6). I also found a significant difference in the proportion of AFBB NSCLC (stages I, II, III and IV) patients treated with radiation as compared to the corresponding proportion of AMBB patients ($p < 0.05$). But, the difference between AFBB and AMBB patients that received neither surgery nor radiation could not be calculated because of low sample counts for the subgroups.

The logistic regression analysis for predicting treatment receipt showed that AFBB patients had a slightly higher probability to receive surgery compared to AMBB. This study found that the AFBB patients were younger than AMBB as shown on Table 2. The logistic regression analysis (OR [AFBB/AMBB]) = 1.94; 95% CI 1.02-3.68) means that the AFBB patients had a higher probability of having had surgery as shown on Table 13, where the stage at diagnosis and the country of birth (USA or Africa) were the best predictors of a patient to receive surgical therapy. Since the AFBB patients are younger, their diagnosis would be earlier. With the AFBB patients being younger implies that the AMBB patients were older. This analysis showed that the AMBB patients were more likely to have received radiation treatment compared with AFBB patients where the Stage at diagnosis and Country of birth (USA or Africa) with respective p-values of 0.00 and 0.03 were the best predictors for a patient to receive radiation therapy as shown on Table

14. Literature search revealed that there are no existing previous studies that compared these two populations with respect to NSCLC stage at diagnosis and the treatment received. Thus it is difficult to make a comparison of this study outcome to previous study literature in that same context. But, in the context of treating the disease as it relates to the age and the disease stage at diagnosis for the general population of NSCLC patients, this study outcome is similar to results obtained by a number of studies. In one study that analyzed SEER lung cancer in general by comparing the older patient population to the younger ones the result showed that elderly patient population received surgical treatment only at half the rate at which younger patient population did (Owonikoko, Ragin, Belani, Oton, Gooding, Taioli, et al., 2007). In another study that analyzed data from a regional cancer registry, 80.2% of younger patients had surgery for NSCLC, compared with 55.8% of the elderly; while radiation treatment was administered more often in the elderly patients (30.5%) compared with 14.0% of the younger patients (Hillner, McDonald, Desch, Smith, Penberthy, & Retchin, 1998).

These findings are all consistent with expectations based upon NSCLC treatment guidelines or recommendations (National Cancer Institute, 2015; National Comprehensive Cancer Network, 2015). Researchers analyzing SEER data have found that stage at diagnosis is similar for all age groups, from the young through the elderly (Owonikokoet al., 2007). Their findings validate the finding in this study that even though AFBB and AMBB patient populations are significantly different in their ages, the stage at diagnosis of the diseases is not significantly different between them.

Recommendations

Even though I did not find any significant differences among AFBB and AMBB NSCLC patients for the stage at diagnosis, I recommend that further lung cancer studies involving comparisons of these two groups be performed with a much larger sample of AFBB patients. This would provide better insights to whether any within race differences do exist between the two groups with a bigger sample size, essential for efficient planning the prevention and management of the disease.

Other covariates that include profession, socioeconomic status, and health insurance status were not available in the SEER database for all the years, 2004-2011 for both groups. Since all the patients in the SEER database are de-identified for confidentiality purposes, it was not possible to do any datalink to obtain information on these unavailable covariates from institutions that diagnosed them with NSCLC. I could not measure the effects of these variables on predicting the probability for the receipt of surgery or radiation therapy as a treatment choice. Therefore, I recommend that researchers conduct further studies with a larger AFBB sample. Future researchers may consider conducting original research in partnership with diagnostic institutions instead of using a secondary data source such as the SEERS dataset. In doing so, they may be able include these covariates to test their effects in predicting the probability of a patient receiving surgery as treatment.

Conclusively, while there was no significant difference between AFBB and AMBB for the stage at diagnosis of the NSCLC, there were significant differences in each type of treatment receipt. The two subgroups involved in this study are both of the same race

and are resident in the same country and so race and current country of residence are not likely causes for this observed treatment differences. Given that there were no significant gender differences before the comparisons were done and that I controlled for age by doing the approximate random sampling to obtain the study sample size, it means that gender and age differences are not likely cause for the observed significant difference in treatment receipt. While other likely cause for such significant difference are the same and others have been controlled for, the country of birth was not controlled for in terms of their early childhood experiences and so might have been the reason why such a difference in treatment receipt is observed. Therefore, the ecological theory can be used to explain the outcome of this study.

References

Adler, A. (1912). *Primary malignant growths of the lung and bronchi: A pathological and clinical study*. New York, NY: Longmans, Green. Retrieved from https://archive.org/details/primarymalignant00adle

Bello, B., Fadahun, O., Kielkowski, D., & Nelson, G. (2011). Trends in lung cancer mortality in South Africa: 1995-2006. *BMC Public Health*, *11,* 209. Retrieved from http://bmcpublichealth.biomedcentral.com/articles/Table 11.1186/1471-2458-11-209

Bonjour, S., Adair-Rohani, H., Wolf, J., Bruce, N. G., Mehta, S. Prüss-Ustün, A. & Smith, K. R. (2013) Solid fuel use for household cooking: Country and regional estimates for 1980–20Table 11. *Environmental Health Perspectives, 121*(7), 784-790. Retrieved from http://ehp.niehs.nih.gov/1205987/

Bradley, C. J., Dahman, B., & Given, C. W. (2008). Treatment and survival differences in older Medicare patients with lung cancer as compared with those who are dually eligible for Medicare and Medicaid. *Journal of Clinical Oncology, 26*(31), 5067-5073. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2652096/

Brenner, D. R., Boffetta, P., Duell, E. J., Bickebdller, H., Rosenberger, A., McCormack, V., …Hung, R. J. (2012). Previous lung diseases and lung cancer risk: A pooled analysis from the International Lung Cancer Consortium. *American Journal of Epidemiology, 176*(7), 573-585. Retrieved from http://aje.oxfordjournals.org/content/176/7/573.full.pdf

Brown, D. W., Anda, R. F., Felitti, V. J., Edwards, V. J., Malacher, A. M., Croft, J. B., & Giles, W. H. (20Table 11). Adverse childhood experiences are associated with the risk of lung cancer: A prospective cohort study. *BMC Public Health, Table 11,* 20. Retrieved from http://www.biomedcentral.com/1471-2458/Table 11/20

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press. Retrieved from http://site.ebrary.com/lib/waldenu/docDetail.action?docID=Table 11331287

Capps, R., McCabe, K., & Fix, M. (2011). New streams: Black African migration to the United States. Retrieved from http://www.migrationpolicy.org/pubs/africanmigrationus.pdf

Centers for Disease Control and Prevention. (20Table 11). Vital signs: Current cigarette smoking among adults aged ≥18 years, United States, 2000-2009. *Morbidity and Mortality Weekly Report, 59*(35), 1135–1140. Retrieved from http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5935a3.htm

Centers for Disease Control and Prevention. (2014). Epi info user guide: StatCalc. Retrieved from http://wwwn.cdc.gov/epiinfo/user-guide/StatCalc/How-to-Opening-StatCalc.html

Chavan, S., Goodman, M., Jemal, A., & Fedewa, S. A. (2014). Receipt of surgical treatment in us women with early stage breast cancer: does place of birth matter? *Ethnicity & Disease*, *24*(1), 1Table 11-115.

Doku, D., Raisamo S., Wiium, N. (2012). The role of tobacco promoting and restraining factors in smoking intentions among Ghanaian youths. *BMC Public Health, 12,*

662. Retrieved from http://bmcpublichealth.biomedcentral.com/articles/Table 11.1186/1471-2458-12-662

Doll R., & Hill, A. B. (1950). Smoking and carcinoma of the lung preliminary report. *British Medical Journal, 4682*, 739-748. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2038856/pdf/brmedj03566-0003.pdf

Doll, R., & Hill, A. B. (1954). The mortality of doctors in relation to their smoking habits; a preliminary report. *British Medical Journal, 1* (4877), 1451–1455. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2085438/pdf/brmedj03396-0011.pdf

Doll, R., Hill, A. B. (1964). Mortality in Relation to Smoking: Ten Years' Observations of British Doctors. *British Medical Journal, 1*(5395), 1399–14Table 11. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1814562/pdf/brmedj02630-0017.pdf

Doll, R., Peto, R. (1976). Mortality in relation to smoking: 20 years' observations on male British doctors. *British Medical Journal, 2*(6051), 1525–1536. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1690096/pdf/brmedj00549-0007.pdf

Doll, R., Peto, R., Wheatley, K., Gray, R., & Sutherland, I. (1994). Mortality in relation to smoking: 40 years' observations on male British doctors. *BMJ, 309*(6959),

901–911. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2541142/pdf/bmj00460-0017.pdf

Elchoufaniz, S. E., Efird, J. T., O'Neal, W. T., Davies, S. W., Landrine, H., & Biswas, T. (2013). The relation of race and type of health insurance to long-term risk of mortality among lung cancer patients in rural eastern North Carolina. *N C Medical Journal, 74* (6), 464-469. Retrieved from http://www.ncmedicaljournal.com/archives/?74601

Etzel, C. J., Kachroo, S., Liu, M., D'Amelio, A., Dong, Q., Cote, M. L., (2008). Development and Validation of a Lung Cancer Risk Prediction Model for African-Americans. *Cancer Prevention Research, 1* (4), 255-265. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2854402/

Ferlay, J., Shin, H.-R., Bray, F., Forman, D., Mathers, C. & Parkin, D. M. (20Table 11). Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International Journal of Cancer, 127*, 2893–2917. Retrieved from http://onlinelibrary.wiley.com/doi/Table 11.Table 1102/ijc.25516/pdf

Fosgate, G. T. (2009). Practical Sample Size Calculations for Surveillance and Diagnostic Investigations. *Journal of Veterinary Diagnostic Investigation, 21* (1), pp. 3-14. Retrieved from http://vdi.sagepub.com/content/21/1/3.full.pdf+html

Hardy, D., Xia, R., Liu, C., Cormier, J. N., Nurgalieva, Z., & Du, X. L. (2009). Racial disparities and survival for nonsmall-cell lung cancer in a large cohort of black

and white elderly patients. Retrieved from http://onlinelibrary.wiley.com/doi/Table 11.Table 1102/cncr.24521/pdf

Hill, A. B. (1965). "The Environment and Disease: Association or Causation?". *Proceedings of the Royal Society of Medicine, 58* (5), 295–300. Retrieved from PMC1898525http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1898525/pdf/procrsmed00196-00Table 11.pdf

Hillner, B. E., McDonald, M. K., Desch, C. E., Smith, T. J., Penberthy, L. T., Retchin, S. M. (1998). A comparison of patterns of care of nonsmall cell lung carcinoma patients in a younger and Medigap commercially insured cohort. *Cancer. 83*(9),1930–1937. Retrieved from http://onlinelibrary.wiley.com/doi/Table 11.Table 1102/(SICI)Table 1197-0142(19981Table 111)83:9%3C1930::AID-CNCR8%3E3.0.CO;2-X/full

Hou, W., Fu, J., Ge, Y., Du, J., & Hua, S. (2013). Incidence and risk of lung cancer in HIV-infected patients. *Journal of Cancer Research and Clinical Oncology, 139* (11),1781-1794. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/23892408

IBM Corporation. (2014). IBM SPSS software. Retrieved from http://www-01.ibm.com/software/analytics/spss/

Howington, J. A, Blum, M. G., Chang, A. C., Balekian, A. A. & Murthy, S. C. (2013). Treatment of stage I and II non-small cell lung cancer: Diagnosis and Management of Lung Cancer. American College of Chest Physicians evidence-based clinical practice guidelines. Chest, 143(5), e278s–e313s. Retrieved from

http://journal.publications.chestnet.org/data/Journals/CHEST/926876/chest_143_5_suppl_e278S.pdf

International Agency for Research on Cancer. (2011). Cancer Incidence in Five Continents Volumes I to IX: Cancer registry list. Retrieved from http://ci5.iarc.fr/CI5i-ix/registry_list_by_registry.htm

International Agency for Research on Cancer. (2013a). Cancer Incidence in Five Continents, Vol. X: Publications. Retrieved from http://ci5.iarc.fr/CI5-X/ci5-X.htm

International Agency for Research on Cancer. (2013b). Cancer Incidence in Five Continents, Vol. X: Data sources and methods. Retrieved from http://globocan.iarc.fr/Pages/DataSource_and_methods.aspx

Jemal, A., Center, M. M., & Ward, E. (2009). The convergence of lung cancer rates between blacks and whites under the age of 40, United States. *Cancer Epidemiology Biomarkers Prevention*, *18*(12), 3349-3352, Retrieved from http://cebp.aacrjournals.org/content/18/12/3349.long

Jemal, A., Thun, M., Yu, X. Q., Hartman, A. M., Cokkinides, V. Center, M. M., & Ward, E. R., (2011). Changes in smoking prevalence among U.S. adults by state and region: Estimates from the Tobacco Use Supplement to the Current Population Survey, 1992-2007. *BMC Public Health*, *11*, 512. Retrieved from http://www.biomedcentral.com/1471-2458/11/512

Jemal, A., Ma, J., Rosenberg, P. S., Siegel, R., & Anderson, W. F. (2012). Increasing lung cancer death rates among young women in Southern and Midwestern States. *Journal of Clinical Oncology, 30*(22), 2739-2744. Retrieved from http://jco.ascopubs.org/content/early/2012/06/20/JCO.2012.42.6098.full.pdf

Kenneway, N. M., & Kenneway, E. L. (1936). A Study of the Incidence of Cancer of the Lung and Larynx. *J. Hyg*., *6*, 236-267. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2170982/pdf/jhyg00231-01Table 11.pdf

Lathan, C. S., Okechukwu, C., Drake, B. F., & Bennett, G. G. (20Table 11). Racial Differences in the Perception of Lung Cancer. 2005 Health Information National Trends Survey (HINTS). Cancer. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2854860/pdf/nihms163775.pdf.

Link, B. G., & Phelan, J. (1995). Social Conditions as Fundamental Causes of Disease. *Journal Of Health & Social Behavior, 36*, 80-94. Retrieved from http://hss.ucsf.edu/PDF/pop-health-classics/Link%201995.pdf

Liu, Y., Steenland, K., Rong, Y., Hnizdo, X., Zhang, H., Shi, T., --- & Chen, W. (2013). Exposure-Response Analysis and Risk Assessment for Lung Cancer in Relationship to Silica Exposure: A 44-Year Cohort Study of 34,018 Workers. *American Journal of Epidemiology. 178* (9), 1424-1433. Retrieved from http://www.medscape.com/viewarticle/813584

Mirsadraee, S., Oswal, D., Alizadeh, Y., Caulo, A., van-Beek, E. (2012). The 7th lung cancer TNM classification and staging system: Review of the changes and implications. *World Journal of Radiology, 4*(4),128-134. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3351680/

Missaoui, N., Hmissa, S., Landolsi, H., Korbi, S., Joma, W., Anjorin, A., ... & Mokni, M. (2011). Lung cancer in central Tunisia: epidemiology and clinicopathological features. *Asian Pacific Journal of Cancer Prevention*, *12*, 2305-2309.retrieved from http://apjcpcontrol.net/paper_file/issue_abs/Volume12_No9/2305-09%20c8.13%20Missaoui.pdf

Moldoveanu B, Otmishi P, Jani P, Walker, J., Sarmiento, S., Guardiola, S., Saad, M., & Yu, J. (2009). Inflammatory mechanisms in the lung. *Journal of Inflammatory Research*, *2*, 1-11. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3218724/

Msyamboza, K. P., Ngwira, B., Dzowela, T., Mvula, C., Kathyola, D., Harries, A. D., Bowie, C. (2011). The Burden of Selected Chronic Non-Communicable Diseases and Their Risk Factors in Malawi: Nationwide STEPS Survey. *PLoS One*, *6* (5), e20316. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3Table 110352/?tool=pubmed

Msyamboza, K. P., Dzamalala, C., Mdokwe, C., Kamiza, S., Lemerani, M., Dzowela, T., & Kathyola, D. (2012). Burden of cancer in Malawi; common types, incidence and trends: National population-based cancer registry. *BMC research notes, 5*(1), 149. Retrieved from http://www.biomedcentral.com/1756-0500/5/149/

National Cancer Institute. (2013a). SEER Registries. Retrieved from http://seer.cancer.gov/registries/

National Cancer Institute. (2013b). SEER Program Overview. Retrieved from http://seer.cancer.gov/about/overview.html

National Cancer Institute. (2013c). SEER Data Management System. Retrieved from http://seer.cancer.gov/seerdms/

National Cancer Institute. (2013d). SEER program coding and staging manual 2013. Retrieved from http://seer.cancer.gov/manuals/2013/SPCSM_2013_maindoc.pdf

National Cancer Institute. (2013e). SEER Program Coding and Staging Manual 2013 Appendix B: country and state codes. Retrieved from http://seer.cancer.gov/manuals/2013/SPCSM_2013_AppendixB.pdf

National Cancer Institute. (2013f). SEER cancer statistics review. Retrieved from http://seer.cancer.gov/csr/1975_20Table 11/results_merged/topic_race_ethnicity.pdf

National Cancer Institute. (2013). SEER Program Coding and Staging Manual 2013 Appendix B: country and state codes. Retrieved from http://seer.cancer.gov/manuals/2013/SPCSM_2013_AppendixB.pdf

National Cancer Institute (2014a). SEER training module: Process of cancer data collection retrieved from http://training.seer.cancer.gov/registration/data/collection.html).

National Cancer Institute. (2014b). SEER training module: Data Collection. Retrieved from http://training.seer.cancer.gov/registration/data/

National Cancer Institute. (2014c). SEER training module: Confidentiality. Retrieved from http://training.seer.cancer.gov/registration/data/confidentiality.html

National Cancer Institute. (2014d). SEER training module: How to Abstract Cancer Registry Information, retrieved from http://training.seer.cancer.gov/abstracting/intro/how/

National Cancer Institute. (2014e). SEER training module: Biographical information. Retrieved from http://training.seer.cancer.gov/abstracting/intro/how/biological.html

National Cancer Institute. (2014f). Non-small-cell lung cancer stages and treatment. Retrieved from http://training.seer.cancer.gov/lung/

National Cancer Institute (2015). Non-Small Cell Lung Cancer Treatment: Treatment Options by Stage. Retrieved from

http://www.cancer.gov/types/lung/patient/non-small-cell-lung-treatment-pdq#section/_205

National Comprehensive Cancer Network (NCCN). (2015). NCCN Guidelines Version 7.2015: Non-Small Cell Lung Cancer. Principles of surgical therapy. Retrieved from http://www.nccn.org/professionals/physician_gls/pdf/nscl.pdf

Owonikoko, T. K, Ragin, C. C, Belani, C. P, Oton, A. B , Gooding, W. E, Taioli, E., et al. (2007). Lung cancer in elderly patients: an analysis of the surveillance, epidemiology, and end results database. *Journal of Clinical Oncology, 25*, 5570-5577. Retrieved from http://jco.ascopubs.org/content/25/35/5570.long

Owusu-Dabo, E., Lewis, S., McNeill, A., Gilmore, A., & Britton, J. (2009). Smoking uptake and prevalence in Ghana. *Tobacco Control, 18*, 365-370. Retrieve d from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2745559/?tool=pubmed

Pampel, F. (2008). Tobacco use in sub-Sahara Africa: Estimates from the demographic health surveys. *Social Science & Medicine. 66*, 1772e-1783e. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2679748/

Pope, C. A., Burnett, R. T., Krewski, D., Jarrett, M., Shi, Y., Calle, E. E. & Thun, M. (2009). Cardiovascular mortality and exposure to airborne fine particulate matter and cigarette smoke: shape of the exposure-response relationship. *Circulation 120*, 941–948. Retrieved from http://circ.ahajournals.org/content/120/11/941.full

Pope, C. A., Burnett, R. T., Turner, M. C., Cohen, A., Krewski, D., Jerrett, M., & Thun, M. J. (2011). Lung cancer and cardiovascular disease mortality associated with ambient air pollution and cigarette smoke: shape of the exposure–response relationships. *Environmental Health Perspective, 119*, 1616–1621. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3226505/

Raaschou-Nielsen, O., Andersen, Z. J., Hvidberg, M., Jensen, S. S., Ketzel, M., Sorensen, M., & Tjonneland, A. (2011). Lung cancer incidence and long-term exposure to

air pollution from traffic. *Environmental health perspectives*, *119*(6), 860. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3114823/

Raz, D. J., Gomez, S. L., & Chang, E. T. (2008). Epidemiology of nonsmall cell lung cancer in Asian Americans: incidence patterns among six subgroups by nativity. *Journal of Thoracic Oncology*. *3*, 1391–1397. Retrieved from http://journals.lww.com/jto/Fulltext/2008/12000/Epidemiology_of_Non_small_Cell_Lung_Cancer_in.6.aspx#

Samet, J. M., Avila-Tang, E., Boffetta, P., Hannan, L. M., Olivo-Marston, S., Thun, M. J., & Rudin, C. M. (2009). Lung cancer in never smokers: clinical epidemiology and environmental risk factors. *Clinical Cancer Research, 15(*18*),* 5626-5645. Retrieved from http://clincancerres.aacrjournals.org/content/15/18/5626.full

Shebl, F. M., Engels, E. A., Goedert, J. J., & Chaturvedi, A. K (20Table 11). Pulmonary infections and risk of lung cancer among persons with AIDS. *Journal of Acquired Immune Deficiency Syndrome*, *55*(3), 375-379. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2955766/pdf/nihms-220467.pdf

Sigel, K., Wisnivesky, J. Gordon, K.., Dubrow, R., Justice, A., Brown, S. T., Goulet, J.,--- & Crothers, K..(2012). HIV as an independent risk factor for incident lung cancer. *AIDS*, *26*(8),Table 1117-Table 1125. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC35802Table 11/

Sitas, F., Parkin, M., Chirenje, Z., Stein, L., Mqoqi, N., & Wabinga, H., (2006). Cancer. In Jamison, D. T., Feachem, R. G., Makgoba, M. W., Bos, E.R., Baingana, F. K.,

Hofman, K. J., & Rogo, K. O. (eds.), *Disease and Mortality in Sub-Saharan Africa* (2nd ed.). Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK2293/

Smith, K. R., McCracken, J. P., Thompson, L., Edwards, R., Shields, K. N., Canuz, E., & Bruce, N. (20Table 11). Personal child and mother carbon monoxide exposures and kitchen levels: methods and results from a randomized trial of wood fired chimney cook stoves in Guatemala. *Journal of Exposure Science and Environmental Epidemiology, 20*, 406–416. Retrieved from http://ehs.sph.berkeley.edu/krsmith/publications/20Table 11/jeseecorespire.pdf

Thun, M. J., Brian, D. Carter, B. D., Feskanich, D., Freedman, N. D., Prentice, R., ---, & Gapstur, S. M. (2013). 50-Year Trends in Smoking-Related Mortality in the United States. *New England Journal of Medicine, 368*, 351-364. Retrieved from http://smokingcessationleadership.ucsf.edu/thun_et_al_nejm_2013.pdf

Turner, M. C., Krewski, D., Pope, C. A. Chen, Y., Gupstur, S. M., & Thun, M. (2011). Long-term ambient fine particulate matter air pollution and lung cancer in a large cohort of never-smokers. *American Journal of Respiratory Critical Care Medicine, 184* (12), 1374-1381. Retrieved from http://cleanaircarolina.org/wp-content/uploads/2011/06/2011_lung-cancer-nonsmokers_research.pdf

United Nations. (2013). Composition of macro geographical (continental) regions, geographical sub-regions, and selected economic and other groupings. Retrieved from http://millenniumindicators.un.org/unsd/methods/m49/m49regin.htm

United Nations. (2013a). Standard country or area codes and geographical regions for statistical use. Retrieved from http://unstats.un.org/unsd/methods/m49/m49.htm

United Nations. (2013b). Composition of macro geographical (continental) regions, geographical sub-regions, and selected economic and other groupings. Retrieved from http://millenniumindicators.un.org/unsd/methods/m49/m49regin.htm

United States Census Bureau (2012). 2000 National Population Projections: Detailed Data Files: Total Population by Age, Sex, Race, Hispanic Origin, and Nativity. Retrieved from http://www.census.gov/population/projections/data/national/np-d2.html

United States Department of Health and Human Services. (20Table 11, November). Healthy People 2020.
Retrieved from http://www.healthypeople.gov/2020/TopicsObjectives2020/pdfs/HP2020_brochure_with_LHI_508.pdf

United States Department of Health and Human Services. (20Table 11). The Mission, Vision, and Goals of Healthy People 2020: Overarching Goals. Retrieved from http://www.healthypeople.gov/2020/TopicsObjectives2020/pdfs/HP2020_brochure_with_LHI_508.pdf

United States Department of Health and Human Services. (2013). HHS Action Plan to Reduce Racial and Ethnic Health Disparities. Retrieved from http://minorityhealth.hhs.gov/npa/files/Plans/HHS/HHS_Plan_complete.pdf

Waldron, H. A. (1983). A brief history of scrotal cancer. British Journal of Industrial Medicine, 40(4), 390–401. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMCTable 1109212/?page=2

Wynder, E. L. & Graham, E. A. (1950). Tobacco Smoking as a Possible Etiologic Factor in Bronchiogenic Carcinoma A Study of Six Hundred and Eighty-Four Proved Cases. *JAMA, 143*, 329-336. Retrieved from http://www.epidemiology.ch/history/PDF%20bg/Wynder%20and%20Graham%201950%20tobacco%20smoking%20as%20a%20possible%20etiologic.pdf

Yawson, A. E., Baddoo, A. Hagan-Seneadza, N. A., Calys-Tagoe, B., Hewlett, S., Dako-Gyeke, P. Mensah, G., --- & Biritwum, R. (2013). Tobacco use in older adults in Ghana: sociodemographic characteristics, health risks and subjective wellbeing. *BMC Public Health, 13*, 979. Retrieved from http://www.biomedcentral.com/content/pdf/1471-2458-13-979.pdf

Youlden, D. R., Cramb, S. M., & Baade, P. D. (2008). The International Epidemiology of Lung Cancer: Geographical Distribution and Secular Trends. *Journal of Thoracic Oncology, 3*(8), 819-831. Retrieved from http://journals.lww.com/jto/Fulltext/2008/08000/The_International_Epidemiology_of_Lung_Cancer_.4.a

Yu, Y., Liao, C., Hsu, W., Chen, H., Liao, W., Muo, C., & Chen, C. (2011). Increased Lung Cancer Risk among Patients with Pulmonary Tuberculosis: A Population Cohort Study. *Journal of Thoracic oncology, 6*(1), 32–37. Retrieved from

http://journals.lww.com/jto/Fulltext/2011/0Table

1100/Increased_Lung_Cancer_Risk_among_Patients_with.8.aspx

Appendix : *Sample and Population Characteristics*

	COUNTRY OF BIRTH		GENDER		STAGE @ DIAGNOSIS		THERAPY			
Number of:	AFBB	AMBB	Males	Females	Early Stage	Late Stage	Surgery without Radiation	Radiation without Surgery	Surgery and Radiation Combined	No Surgery/ No Radiation @ Late Stage
Original SEER Query	135	25139								
Unknown /Missing Variables	16	2846								
AFBB and AMBB Study Population	119	22293								
AFBB Study Sample	119		59	60	28	91	29	38	6	45
AMBB Study Sample		238	127	111	60	178	41	114	6	70

Sample and Population Characteristics (continued)

`	No Surgery/No Radiation @ Early Stage	Single	Married	Divorced	Widowed	30-39 Yrs	40-49 Yrs	50-59 Yrs	60-69 Yrs	>70 Yrs
Original SEER Query										
Unknown/Missing Variables										
AFBB and AMBB Study Population										
AFBB Study Sample	1	30	68	8	13	7	12	38	43	19
AMBB Study Sample	7	57	82	54	45	2	19	37	69	111