2020

Age at Diagnosis and Lung Cancer Presentation

Marida Gingras
Walden University

Walden University

College of Health Professions

This is to certify that the doctoral dissertation by

Marida Gingras

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. Ji Shen, Committee Chairperson, Public Health Faculty
Dr. German Gonzalez, Committee Member, Public Health Faculty
Dr. Chinaro Kennedy, University Reviewer, Public Health Faculty

Chief Academic Officer and Provost
Sue Subocz, Ph.D.

Walden University
2020

Abstract

Age at Diagnosis and Lung Cancer Presentation

by

Marida Gingras

MPH, Walden University, 2016

BSPH, University of Cincinnati, 2014

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Public Health

Walden University

November 2020

Abstract

Lung cancer is the leading cause of cancer-related mortality in the United States because of its lack of symptoms until late stages. It is noted that a 5-year relative survival rate can be improved by earlier cancer detection. The currently recommended age of screening by the U.S. Preventive Services Task Force may not be optimal, and the recommendations are only for smokers. The purpose of this cross-sectional study was to examine the association between the stages of presentation of lung cancer and age at diagnosis using quantitative research. Using the Surveillance, Epidemiology, and End Results (SEER–18) database, 63,107 records were examined of patients diagnosed with lung cancer between 2010–2015 was examined. Chi-squared and logistic regression were used to perform the data analyses. The results of both the chi-squared test and logistic regression showed a significant association between (1) age at diagnosis and the stages presentation of lung cancer after controlling for demographic risk factors; (2) the stages of presentation of lung cancer and the demographic risk factors after controlling age; and (3) the stages of lung cancer, age at diagnosis, and the demographic risk factors of lung cancer. The significant risk of people diagnosed with lung cancer was associated with age and demographic factors: gender, race/ethnicity, and geographical region. This study may help provide additional information for lung cancer screening with the most effective method in adults starting at age 45, which may have a significant positive social change.

Age at Diagnosis and Lung Cancer Presentation

by

Marida Gingras

Master of Public Health, Walden University, 2016

Bachelor of Science in Public Health, University of Cincinnati, 2014

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Public Health

Walden University

November 2020

Dedication

I would like to dedicate my work to those who are currently suffering from lung cancer or coronavirus (COVID–19), and their family members; to those who are frontline healthcare workers nationally and globally; health professionals (CDC, NIOSH, NIH, and WHO employees); the National Cancer Institute for allowing me to use their SEER data; my chair, Dr. Ji Shen, my committee member, Dr. German Gonzalez, my URR, Dr. Chinaro Kennedy, and my colleague, Ms. Susan Afanuh, who supported me during my deployment to help in the response to COVID–19 and my academic journey; and my family and friends who supported me throughout my dissertation journey.

Acknowledgments

I have always thought I would be able to finish my dissertation regardless of the time it takes. As Dr. Ronald E. McNair stated, whether you reach your goals in life depends entirely on how well you prepare for them and how badly you want them. “You’re an eagle, stretch your wings and fly to the sky” ~ Ronald E. McNair. Without the support, hard work, and encouragement of my chair, Dr. Ji Shen, committee member Dr. German Gonzalez, University Research Reviewer, Dr. Chinaro Kennedy, I could not have successfully completed this dissertation. Thank you to all of them and my friends and family for their patience through this long journey.

Table of Contents

List of Tables

List of Figures

Chapter 1: Foundation of the Study

Because many studies have found that the incidence of cancer increase with age (Chen et al., 2019; White et al., 2014), studies that evaluate a combination of factors provide a better tool to measure age-related factors that contribute to lung cancer development based on the stages of lung cancer. However, resources are insufficient for attributing age-associated factors to lung cancer. Although lung cancer can be considered age-related because the incidence of cancer increases with age, it can also be assumed that there is a potential link between age and health impairment based on the length and amount of exposure to carcinogens (WHO, 2019). Lung cancer ranks first in cancer mortality and second in cancer morbidity among all top ten cancers as cited in https://seer.cancer.gov/statfacts/html/all.html (NCI, n. d.). According to the data from the SEER program, 228,150 new cases of lung and bronchus cancer were reported in 2019 in the United States, and an estimated 142,670 (62%) died of lung or bronchus cancer (Siegel et al., 2019). According to data from the National Cancer Institute (NCI, n. d.), at least 93% of people who died from lung cancer were aged 55 or older (NCI, 2018). The risk factors for lung cancer include smoking, age, gender, race/ethnicity, and geographical region (Bakulski et al., 2019; Chu et al., 2018; Fos, 2011; NTP, 2016; White et al., 2014; Wagner, Cameron-Smith, Wessner & Franzke, 2016). In the United States, the U.S. Preventive Services Task Force (USPSTF) now recommends that high-risk individuals who are between 45 and 80 years of age and have a history of smoking 30 packs per year (or those who are current smokers) should have yearly lung cancer screening. (Ryan, 2018; USPSFT, 2018). Although cigarette smoking causes lung cancer,

age is a risk marker for lung cancer regardless of smoking status and may be an important determinant for lung cancer screening. This study therefore investigates the relationship between age at diagnosis and stage of lung cancer presentation to determine whether a recommendation for lung cancer screening based only on age is necessary. In addition, this study can help predict the morbidity and mortality of lung cancer.

Background of the Problem

Types of Lung Cancer

There are two main types of broadly classified lung cancers: non-small-cell-lung cancer (NSCLS) constitutes about 70%–85% cases, and small-cell lung cancer constitutes about 15%–30% of cases (AJCC, 2010; Jin et al., 2018; Lozano et al., 2018). However, most reported cases of lung cancer are NSCLC, which consists of five different subtypes: (i) adenocarcinoma, (ii) squamous cell carcinoma, (iii) adenosquamous carcinoma, (iv) large cell neuroendocrine carcinoma, and (v) large cell carcinoma (Gingras, M., 2018; Zamay et al., 2018). In NSCLC, adenocarcinoma is the most common type seen in both smokers and non-smokers (ACS, 2019; Zamay et al., 2018). Adenocarcinoma arises from glandular cells of the bronchial mucosa and expresses several protein makers. The diagnosis of adenocarcinoma is often based on the identification of molecular markers of mutations in epidermal growth factor receptor, ERCC–1 (DNA excision repair protein), RRM–1, KRAS, TS, and EML4–Alk (Zamay et al., 2018). Squamous cell carcinoma arises from modified bronchial epithelia cells, and one of the most characteristic features of squamous cell cancer is high levels of fragmented cytokeratin CK–19 subunit (CYRFA21–1). The level of CYRFA21–1 increases during the malignization process of

normal epithelia cells and is highly expressed in the serum of patients with a metastatic form of squamous cell carcinoma. Adenosquamous carcinoma contains two types of cells: (i) squamous cells (thin, flat cells that line certain organs) and (ii) gland-like cells (Zamay et al., 2018). Large-cell neuroendocrine carcinoma is a malignant epithelia tumor comprised of large poloyonal cells that do not show any evidence of histological differentiation (Zamay et al., 2018). The tumor arises from neuroendocrine cells of the respiratory tract lining layer or smooth muscle cells of its wall. A large-cell carcinoma is a heterogeneous group of undifferentiated malignant neoplasms that lack the cytological and architectural features of cell carcinoma and glandular or squamous differential (Zamay et al., 2018). Large-cell carcinoma is categorized as a subtype of NSCLC that originates from epithelial cells of the lung (Zamay et al., 2018).

Association Between Smoking and Lung Cancer

Smoking is associated more with squamous cell carcinoma and small-cell cancers than with adenocarcinomas (Stram et al., 2019). An association between exposure to cigarette smoke and the development of lung cancer is well recognized; more than 80% of lung cancer cases are caused by cigarette smoke (Bakulski et al., 2019; Gingras, M., 2018; Ni et al., 2018). According to the National Toxicology Program, cigarettes contain at least 69 carcinogens that promote cancer in humans (National Toxicology Program, 2016; Pu, Xu, Zhang, Yuan, Hu, Huang, & Wang, 2017). Since smoking affects DNA methylation throughout the genome (Bakulski et al., 2019), the longer a person smokes cigarette, the higher the exposure to carcinogens (including carbon monoxide, hydrocarbons, ammonia, cadmium, and other substances) that elevate the risk of lung

cancer. However, age is a major risk factor for lung cancer, and the death rate of lung cancer is higher among middle-aged and older populations (NCI, n. d.). The current understanding of age-related factors that contribute to lung cancer is insufficient. Although some researchers may not agree that age (as an absolute number) is a risk factor for cancer, age-related factors such as a change in the biological materials of DNA can contribute to cancer (Adams et al., 2015).

Association Between Age and Cancer Risk

Many epidemiology studies have found that age increases the risk of cancer, and that human physiological function declines and cell mutations increase with age (Bakulski et al., 2019; Kathuria et al., 2014; Swanton et al., 2015; Wagner et al., 2016; Yokoyama et al., 2019). Molecular studies have found that multiple alterations and damage within molecular pathways increase as age increase (Wagner et al., 2016; Xia & Han, 2018; Chen et al., 2017; Yang, D., Yang, K., & Yang, M., 2018). Several cancer studies have found that at least 90% of carcinogens are mutagens that increase with age, which leads to earlier death (Adams et al., 2015; Smith et al., 2009; Swanton et al., 2015; Weiss, 2002; Yancik et al., 2005). Thus, age-related factors could be the most profound risk factor for almost all non-communicable diseases, including cancer (Wagner et al., 2016). Because physiological function declines as age increases (Adams et al., 2015; Wanger et al., 2016; Xia et al., 2018), a single test that measures age-related risk factors could predict the future onset of lung cancer (Adams et al. 2015). Although many studies found that age-related factors increase the overall risk of cancer, there is still a lack of understanding of age-related factors that contribute to lung cancer.

Other Risk Factors for Cancer

Nevertheless, lung cancer is not caused by just a single factor, but a combination of internal and external (also known as genetic and environmental) risk factors (Swanton, McGranahan, Starrett & Harris, 2015; Wagner et al., 2016; Shao, Liang, Long, & Jiang, 2017). Cancer development starts at the cellular level and takes approximately 20–40 years to develop after cell mutations, based on multiple risk factors (Adams et al., 2015; Wagner et al., 2016). The epidemiology of lung cancer studies often includes variables besides age such as cigarette smoke, gender, race/ethnicity, genetic factors, and geographical regions to find the association with the health problem. Studies focusing on gender differences show that more men develop and die from lung cancer than women (Dorak & Karpuzoglu, 2012; Ryan et al., 2018). According to the World Health Organization (WHO), there is an association between genetics and differences in gender (WHO, 2019). For example, multicenter studies confirmed low testosterone exerts in 84% of lung cancer cases (Hyde et al., 2012; Wagner et al., 2016). In a global study, Ryska et al (2018) found the highest age-standardized incidence rates of non-small lung cancer in men around the world.

In race/ethnicity studies, African Americans in the United States had the highest rates of lung cancer and were disproportionately affected by lung cancer in terms of incidence and survival (David et al., 2016; Ryan et al., 2018). In the exploration of genetic study for lung cancer risk, African-ancestry populations show differences in disease allele frequency, linkage disequilibrium patterns, and phenotype prevalence (David et al., 2016). The percentage of African American men diagnosed with lung

cancer each year is approximately 32% higher than among White men, even though African American men have similar rates of smoking and they initiate smoking later in life on average (David et al., 2016; Ryan, 2018). The higher risk of lung cancer could be because African Americans are more susceptible to the effects of carcinogens from chemical exposure through employment (Ryan, 2018). Geographical region is another possible risk factor for lung cancer; it can affect incidence, survival rates, stage of diagnosis, surgical treatment, and availability of screening centers. Although many risk factors for lung cancer are known, the morbidity and mortality of lung cancer is still the leading cause of public health burdens.

Problem Statement

The problem is that the currently recommended age (55–80 years) for lung cancer screening by the United States Preventive Services Task Force (USPSTF) may not be optimized to effectively catch early stage lung cancer. Although chest X-rays and imaging screening using low-dose computed tomography (LDCT) have decreased the risk of lung cancer, the rates of mortality and morbidity of lung cancer remain stable and have not significantly decreased over the past decades (Patnaik et al., 2017). By the time symptoms appear in imaging screening, lung cancer has already metastasized (Zamay et al., 2017). Survival rates are negatively affected when individuals are not aware of their disease because of the lack of signs and symptoms in early stages (Srivastava, 2012). To enhance screening methods for the early detection of cancer, studies need to identify how age contributes to lung cancer. This dissertation assesses the association between age at diagnosis and stages of presentation of lung cancer by examining the SEER Database. As

many previous studies have found, age increases the risk of lymph node and distant metastasis (Adams et al., 2015; Chen et al., 2019; Yokoyama et al., 2019). This study hypothesizes that stages of presentation of lung cancer differ between patients diagnosed at different ages. The types of stages of lung cancer within age subgroups are assessed.

Purpose of the Study

The purpose of this investigation is to examine the association between the age at which lung cancer is diagnosed and the stage of presentation of lung cancer using quantitative research. SEER data is used to explore age groups at diagnosis [(45–49), (50–54), (55–59), (60–64), (65–74), and (75–84)] and stages of presentation of lung cancer. In addition, the effects of gender, race/ethnicity (African American, White, Other), geographic location, health behaviors, and gene-environment interaction are explored. The presentation of lung cancer includes stage (I, II, III, IV). The results of this study may help to provide a recommendation for effective annual lung cancer screening of adults aged 45–80 who are both smokers and non-smokers (Soo et al., 2018).

Research Questions and Hypotheses

RQ1: Is there a significant association between age at diagnosis and the stages of presentation of lung cancer after controlling for demographic factors: gender, race/ethnicity, and geographic region?

H_o1: There is no significant association between age at diagnosis and the stage of presentation of lung cancer after controlling for demographic factors: gender, race/ethnicity, and geographic region.

H_a1: There is a significant association between age at diagnosis and the stage of presentation of lung cancer after controlling for demographic factors: gender, race/ethnicity, and geographic region.

RQ2: Is there a significant association between the stage of lung cancer and demographic factors gender, race/ethnicity, and geographic regions after controlling age?

H_o2: There is no significant association between the stage of lung cancer and demographic factors after controlling for age.

H_a2: There is a significant association between the stage of lung cancer and demographic factors after controlling for age.

RQ3: Is there any significant relationship between the stage of lung cancer and age and demographic factors?

H_o3: There is no significant relationship between the stage of lung cancer and age and demographic factors.

H_a3: There is a significant relationship between the stage of lung cancer and age and demographic factors.

Theoretical Framework

The theoretical framework for this study was the Cancer Control Continuum (CCC), which is an original framework of the NCI (NCI, n. d.). The goal of using this CCC approach was to reduce the risk of lung cancer through early detection. Yabroff et al. (2019) examined ways to reduce the cancer burden of the nation by accessing healthcare throughout CCC and by making screening methods more effective. Yabroff et

al. stated that although having access to health care was the most effective way to reduce health burdens caused by cancer, implementing early cancer screening would ensure early recognition and a timely response to cancer. Most cases of lung cancer are found in a late stage or stage IV (Zamay et al. 2017), and the survival rate for advanced stages of lung cancer is approximately 15% (AJCC, 2017). Therefore, the recommendation should be updated for vulnerable populations to receive annual preventive screening beginning at age 45. If cancer could be caught early, no Americans would suffer from stage IV lung cancer—the most advanced stage of lung cancer when cancer has spread to the other part of lungs or distant organs, is more difficult to treat, and when the survival rate is generally lower. Therefore, this study used CCC framework to provide valuable information about how screening age should be updated for early detection of lung cancer by prioritizing early detection to increase survivorship.

The three principles of the CCC framework are to view plans, progress, and priorities in terms of cancer etiology, prevention, and detection. Based on these three CCC principles, the various stages of cancer are described with respect to etiology, prevention, and early detection. The principles helped us identify research gaps about age-related factors that contribute to lung cancer. Here, the association between stages of lung cancer and age groups was investigated to increase the knowledge of how age can be a risk factor for developing lung cancer.

The second focus of the CCC framework is prevention through screening. For high-risk populations (those who are both smokers and older), the USPSTF recommends yearly lung cancer screening with low-dose CT and chest X-ray (USPSTF, 2018). The

main purpose of the USPSTF is to protect the health of all Americans by making evidence-based recommendations about preventive services such as screenings (USPSTF, 2015). Although imaging screening using LDCT and chest X-ray decreases the risk of lung cancer, the rate of mortality and morbidity of lung cancer has not significantly decreased over the past decades (NCI, 2018; Patnaik et al., 2017). Because of the exposure to low-dose radiation using LDCT and chest X-rays, UPSTF recommends discontinuing preventive screening once a person has not smoked for 15 years or develops a health problem that substantially limits life expectancy (USPSTF, 2015).

The third focus of the CCC framework is detection. Although early detection has been a challenge in the research community, additional information is provided in this study about the timing of cancer progression from stage I to stage IV based on the age of lung cancer patients. There is limited literature on how cancer progresses from stage I to stage IV, and how long it takes for cancer to develop after cell mutations. A few studies show that cancer development starts at the cellular level and takes approximately 20–40 years to develop after cell mutations (Adams et al., 2015; Wagner et al., 2016). While researchers are still investigating the connection between mutations and the time frame of cancer development, finding an effective method for early detection is crucial for saving lives. In previous research studies, the CCC has been a useful framework for early detection and prevention of cancer (Yabroff, et al., 2019).

Effectiveness in screening is another critically important factor for preventing and reducing the public health burden, since the symptoms of lung cancer do not appear in the early stages. Through early screening, cancer detection can happen before tumors appear

in the lungs, increase in size, and metastasize to nearby organs. This early detection may prevent and reduce the rate of mortality and morbidity caused by lung cancer. The evaluation of the association between age at diagnosis and demographic factors such as gender and race/ethnicity, health behaviors, and gene-environment interactions will help us understand where cancers begin and how to detect them at an early stage.

The CCC framework may reduce the challenge of preventive screening studies by explaining the association between well known risk factors and lung cancer at the molecular level. The association between cigarette smoking and lung cancer is well known, and approximately 87% of deaths among smokers are due to lung cancer (Bakulski et al., 2019). However, the literature is limited on how the carcinogens in cigarette smoke increase the risk of mutation and how mutations increase with age. According to the National Toxicology Program, a cigarette contains at least 69 carcinogens that promote cancer in humans (National Toxicology Program, 2016). Any substance that causes cancer is known as a carcinogen, and exposure to carcinogens may arise from ingestion, physical touch, or inhalation (NCI, n. d.). However, harmful effects from exposure to carcinogens are based on the concentration and duration of the exposure (NCI, n. d.). Gene-environment interaction (changes in DNA and mutations) that can contribute to lung cancer is understudied. Many studies have identified biomarkers found in the blood that can be used as a sign of a normal or abnormal progression of cancer (Srivastava, 2012).

Taking advantage of modern technology to use public health information may help achieve a higher level of confidence for interpreting the biological process.

Technology has changed our understanding of cancer development. The CCC framework helps us collaborate with others to determine where more resources may be needed. Using the underlying framework of CCC, this study investigates how risk factors related to age, health behavior, gender, race/ethnicity, geographical location, and gene-environment-interaction can contribute to the development of lung cancer. The CCC framework can improve our understanding of the epidemiology of lung cancer based on contributing risk factors and help guide recommendations for screening. In addition, this framework can provide information for future blood-based biomarkers screening.

Nature of the Study

The nature of this research was a quantitative study using a cross-sectional design to examine data from age, stage, and demographic factors. Specifically, the differences in age groups and stages of presentation of lung cancer were analyzed. This quantitative method includes stages of lung cancer analyses on each age groups and demographic factors using chi-squared test. A multinomial regression analysis was also performed to explore the effect of one dependent with four subgroups and the four independent variables (age at diagnosis, gender, race/ethnicity, geographic region). The chi-squared test to determine whether there was any association between stages of lung cancer and age groups, whether there was any association between stages of lung cancer and demographic risk factors after controlling age, and whether there was any association between stages of lung cancer, age, and demographic factors. The association between age at diagnosis and the stage of presentation of lung cancer after controlling for demographic factor was identified by a chi-squared test. The association between stages

of presentation of lung cancer and demographic risk factors of lung cancer after controlling for age at diagnosis was identified by chi-squared test. Logistic regression was used to obtain odd ratios (OR) and their respective 95% confidence intervals and displayed the strong association of the stages of presentation of lung cancer, age at diagnosis, and demographic risk factors.

Definition of Terms

The Dependent variable:

1) Stages of Lung Cancer: Stages of lung cancer were based on the collaborate stage, which was cancer fields in the SEER program that include size of the tumor, extension, and swelling or malignancy of lymph nodes (which are small ball-shaped immune system organs distributed throughout the body) (NCI, n. d.). The stages of lung cancer were based on clinical (cTNM) (TNM = tumor, node, metasteses) and pathological (pTNM) staging. Clinical staging is based on the evidence acquired before treatment, including physical examination, imaging studies, laboratory test, and staging procedures such as bronchoscopy or esophagoscopy with ultrasound directed biopsies (AJCC, 2010). The presentation of lung cancer stages is defined by where the cancer cells are located, the size of the lung cancer tumor, and where cancer has spread. The staging of cancer helps provide information about lung cancer prognosis; however, it does not predict how long a patient will live (ALA, 2020). However, patients with stage 0 were excluded from this study. The lower the lung cancer stage, the less the cancer has spread (AJCC, 2010). The

types of lung cancer depend on where the malignant tumor (uncontrolled growth and damage of healthy cells) arises from different cells such as bronchial epithelium, bronchioles, alveoli, or bronchial mucous glands. As many factors such as smoking, family history, occupational exposure, and genetic background contribute to developing lung cancer, different types of lung cancer can occur based on individual risk. Because of unknown causes and the diversity in the molecular-biological features of lung cancer, identifying the genetic profile that can predispose individuals to a high risk of lung cancer will help us better understand and may lead to improved screening options.

i. Stage I: Lung cancer included cancer cells in the lungs and also cancer cells that have spread to any lymph nodes. If the lung cancer is in the lungs but has not spread to lymph nodes, it is described as stage I. The classification of stage I lung cancer included stage IA as (T1a–N0–M0), (T1b–N0–M0), and stage IB as (T2a–N0–M0). T1a was a tumor that was (~ 2 cm in size) and TIb was (> 2–3 cm in size).

ii. Stage II: The classification of stage II lung cancer included stage IIA as (T2b–N0–M0) T2b was (> 5 – 7 cm in size), (T1a–N1–M0), (T1b–N1–M0), (T2a–N1–M0), stage IIB as (T2b–N1–M0) and (T3–N0–M0).

iii. Stage III: The classification of lung cancer included stage IIIA as (T1a–N2–M0), (T1b–N2–M0), (T2a–N2–M0), (T2b–2–M0), (T3–N2–M0), (T3–N2–M0), (T4–N0–M0), and (T4–N1–M0), stage IIIB as (T1a–N3–

M0), (T1b–N3–M0), (T2a–N3–M0), (T2b–N3–M0), (T3–N3–M0), and (T4–N3–M0) (> 7 cm in size).

iv. Stage IV: The classification of lung cancer included stage IV as (AnyT–AnyN–M1a) and (AnyT–AnyN–M1b). Stage IV lung cancer is the most advanced stage of lung cancer. It indicates that the cancer has spread to both lungs and metastasized to distant organs. Because most cases of lung cancer are asymptomatic, and current imaging screening methods detect only visible and irreversible changes in lung, a late stage of lung cancer was the most diagnosed case among all stages of lung cancer (Zamay et al. (2017).

The Independent Variables:

1) Age at diagnosis: defined as the measurement of the age of the patient at their last birthday. Age at diagnosis were groups in 5 year-interval (45–49), (50–54), (55–59), (60–64), (65–74), (75–79), and (80–84).
2) Gender: defined by the chromosomal genotypes and sexual phylotype present at birth (NCI, n. d).
3) Race/ethnicity: defined by specific physical, hereditary, and cultural traditions or origins.
4) Geographical region: based on residency in a SEER geographic catchment area at the time of diagnosis, metropolitan areas included (counties in metropolitan area of greater than 1 million, counties in metropolitan area of 250 to 1 million, counties in metropolitan area of less than 250 thousand) and

non-metropolitan areas included (non-metropolitan counties adjacent to a metropolitan area, and non-metropolitan counties not adjacent to a metropolitan areas). Registries collected state, county, zip code, and address derived from the census tract. A version of the census tract depended on the year of diagnosis.

SEER: The SEER program database included large amounts of data, and these data were representative of the U.S. population. SEER database was supported by the Surveillance Research Program (SRP) in NCI's Division of Cancer Control and Population Sciences (DCCPS). The SRP not only provides data but also disseminates reliable population-based statistics. All the available lung cancer cases from SEER were used, and all patients diagnosed with lung cancer between 2010–2015 were included in the analysis.

The SEER program is a population-based cancer registry covering approximately 36.7% of the U.S. population (based on the 2010 Census; Gingras, M., 2018; NCI, n. d.). The information provided in SEER is maintained by the NCI and represents an effort to reduce the public health burdens of the U.S. population. Some differences may exist between the population of patients recorded in the SEER database and the U.S. general population; for example, SEER has a higher likelihood of foreign-born persons compared with the 2010 U.S. census, and a higher proportion of races/ethnicities other than White American and African American populations. To reduce the limitation of knowledge about age-related factors that contribute to lung cancer, data from the National Cancer Instigate of SEER program were extracted for all cases of lung cancer diagnosed between 2010–2015.

Assumption, Delimitation, and Limitation

Assumptions

Based on the literature reviews, this project assumes that the following groups have a higher risk of being diagnosed with more advanced stages of lung cancer: older age groups, men, African American men, and people who live in Kentucky. The results also assume that recommending preventive screening beginning at age 45 for nonsmokers more frequently catches early stages of lung cancers that may reduce the rate of the morbidity and mortality of lung cancer. These assumptions are based on the correlations between age at diagnosis, the stages of presentation of lung cancer, and demographic factors gender, race/ethnicity, and geographical regions, which are evaluated by multinomial analysis using a cross-sectional study. The chi-squared analyses were used to analyze contributing factors of independent variables (age at diagnosis) and (demographic factors) and dependent variables (stages of presentation of lung cancer: stage: I, II, III, and IV). The variables included in this assumption were all based on the previous studies and how long it took for cancer to develop and mutate as age increases (Adams et al., 2015).

According to previous studies, it takes approximately 20–40 years to develop cancer after cell mutation (Adams et al., 2015). Several studies also found that 90% of cancers are caused by mutations, and mutations increase with age (Adams et al., 2015; Swanton et al., 2015; Wagner et al., 2016). Because many studies have found that the incidence of cancer increases with age (Chen et al., 2019; White et al., 2014), studies that evaluate a combination of factors provide a better tool to measure age-related factors that

contribute to lung cancer development. A combination of both age-related markers and cancer stage-related markers can accurately predict the future onset of cancer (Bürkle et al., 2015). To describe how to evaluate age-related lung cancer, this study specified those age groups of lung cancer by displaying the data of lung cancer in stages. The analysis of this study included the correlation between age and the stages of presentation of lung cancer, which were all based on TMN stages. This study provided reasonable justification because it used only patients who were diagnosed with lung cancer to ensure that it made a sound assumption based on the result. The assumption determined a relationship between age and stages of lung cancer, helped better understanding of age-related factors contributed to lung cancer, and supported existing studies. The results of this study also supported the future use of biomarkers of aging to promote effective preventive screening.

Delimitations

The scope of this dissertation was to discover how age differences affect lung cancer by assessing the association between age of diagnosis, gender, race/ethnicity, and stages of presentation of lung cancer, health behaviors, and geographic location. As the human immune system responds to carcinogens differently based on a person's age, the potential for early treatment response and metastatic patterns may also differ among age groups (Zhuang et al., 2017). Many research studies have explored the different prognosis outcomes among younger and older age groups (Becker et al., 2015; Chen et al., 2019). However, resources were insufficient for attributing age-associated factors to lung cancer. Although lung cancer can be considered age-related because the incidence of

cancer increases with age, it can also be assumed that there is a potential link between the age of an individual and health impairment based on the length and amount of exposure to carcinogens (WHO, 2019).

Limitations

The SEER data was broadly representative of the U.S. population, although there were some differences: patients recorded in the SEER database were more likely to be foreign-born compared with the U.S. Census data. To reduce biases, statistical analyses were performed based on the target population (lung cancer patients) of the United States to compare stages of presentation of lung cancer with age at diagnosis. A large proportion of differences in race/ethnicity and age group may increase bias in the statistical analysis. However, to reduce the issue of internal validity, this study addressed specific risk factors such as stages of presentation of lung cancer among the middle and older age groups of the population. Better knowledge of age-related factors is still needed to improve the early detection and outcome of cancer.

Significance of the Study

This project is unique because it demonstrated how age-related risk factors can contribute to lung cancer and how age at diagnosis can help predict the morbidity and mortality of lung cancer. As technological innovations have expanded in health screening over the past few decades, age-related factors have provided information for next-generation research studies to find potential markers for early detection, diagnosis, monitoring, and therapies (Fenizia, Pasquale, Roma, Bergantino, Iannaccone, & Normanno, 2018; Liu, Zhou, & Cao, 2016). Almost all studies of cancer epidemiology

have included age as one of the variables in cancer research (White et al., 2014). Yet age-related factors that contribute to cancer are understudied. Although age can be defined by completed units of time or a time-dependent variable (White et al., 2015), physiological systems declined as time progresses (Bürkle et al., 2015).

Because the incidence of most cancers increases with age, cancer can be considered an age-related disease (Bürkle et al., 2017; Wagner et al., 2016; White et al., 2014). Many cancer studies have found that 90% of cancer development begins at the cell level (Adams et al., 2015; Hammond, 2015; Swanton et al., 2015; Wagner et al., 2016). Mutations/damage in cells increase with aging, which is a sign of the pathological process of cancer and which can be identified as an abnormal biological process in the bloodstream (Bürkle et al., 2017). The results from this dissertation added more information to existing studies about the association between age-related factors and lung cancer. Therefore, age-related factors can be used for detecting lung cancer before it appears in imaging. The results from this study provided valuable information that will have positive social implications for the scientific community and contribute to future research in public health. As public health is multi-faceted for the surveillance of vulnerable populations, age-related factors may aid in the diagnosis of asymptomatic patients who suffer from lung cancer. Insights from this study should aide in efficient and effective future screening studies for early detection of lung cancer. Moreover, it should lead to better health outcomes and reduce the public health burden.

Social Change Implications

The results of the statistical analyses provided information about the most likely age of diagnosis and the optimal age range for preventive screening. Demonstrating how the optimal age at diagnosis contributes to lung cancer can decrease the morbidity and mortality of lung cancer and benefit the public health system. Depending on the rate of decrease, the lowered medical cost, and the health benefits to patients, earlier screening may be a large benefit to society. This study provided statistical data analysis of existing information that can draw conclusions about whether the risk of being diagnosed with lung cancer in 45 to 49-year-old patients is higher than in other older age groups. The results of our data analyses provided a new screening framework to lower the age of lung cancer screening, which will result in reduced cost and improved outcomes for lung cancer treatment.

Chapter 2: Literature Review

Introduction

Lung cancer affects more people than any other disease and can develop without any symptoms. Lung cancer has a large impact on American public health, and many people are not aware of lung cancer until they have symptoms. As the function of physiology declines with increasing age, the function of the healthy lung also declines. Chronological age is a unit number of times that measures one's life starting from the day one is born. Age alone cannot be a risk factor for cancer (White, 2014). However, age-related factors that contribute to lung cancer can be measured through physiological functional capacity and genetic integrity of adult tissue stem cells that decline as age increases (Adams et al., 2015).

Lung cancer can be considered an age-related cancer, because many research studies have shown that the incidence of lung cancer increases with age (Adams et al., 2015; Bai, 2018; White et al., 2014). Changes in DNA and cell mutations affect physiological function that gauge to physical age and are factors that can predict the future onset of ill health (Bai, 2018; Popović et al., 2018; White et al., 2014; Xie et al., 2018). Although a time-dependent physiological functional decline is not preventable as age increases, it is possible to intervene and delay the process of aging and reduce the incidence of age-related lung cancer (Wang, 2018). As age-related factors change biological materials at the cellular level, assessing factors that contribute to lung cancer will help elucidate the epidemiology of lung cancer. Many epidemiological studies of diseases and cancers included diseases and cancers that mainly occur in older people. The

average age of people diagnosed with lung cancer is about 65 (Wagner et al., 2016; White, 2014). Yet, age-related factors that contribute to lung cancer have been understudied (Wagner et al., 2016; White, 2014).

To overcome the lack of understanding of the age-related factors that contribute to lung cancer, this literature review focuses on variables that are related to lung cancer development (such as age at diagnosis, cell mutations, stages of tumors, nodes, and metastasis) to assess how human biological age can help explain the epidemiology of lung cancer. Several biological studies of cancer studies found that (1) cancer cell impairment increases with age, (2) accumulation of carcinogens increases with age, and (3) 90% of cancers are caused by cells mutations (ACA, 2015; Adams et al., 2015; Hammond, 2015; Adams et al., 2015; Swanton et al., 2015; WHO, 2019). On the basis of this evidence, age has a major impact on cell mutations that lead to lung cancer (Adams et al., 2015; Wager et al., 2016). As age increases and exposure and degenerative processes accumulate, assessing age-related factors that contribute to lung cancer will help us understand the epidemiology of lung cancer (Wagner et al., 2016). The number of adults aged 65 or older is projected to reach 98 million by 2060 (Healthy People 2020, 2019). As the population of older adults increases, morbidity and mortality from cancer will also increase (Healthy People 2020, 2019). Thus, identification of age-related factors contributing to lung carcinogenesis not only brings valuable information to the research community, but also explains why older populations are more vulnerable to lung cancer. It will also help support future preventive screening for early detection of lung cancer.

This chapter reviews findings from previous studies on the association between age-related factors and lung carcinogenesis. The results of this study add value to existing risk assessment standards and support future biomarker screening studies for early detection of cancers, as lung cancer symptoms appear only at an advanced stage. In addition, the findings from this dissertation underline the needs for further investigation of age-related factors and highlight knowledge gaps about causal variants and genetic factors responsible for the underlying demographic factors associations. The study also proposes short-term solutions to ensure further progress through a cross-sectional model.

Literature Search Strategy

The literature search used two libraries databases: the Walden University library database and the Stephen T. Tucker CDC library. Two search engines (PubMed and Scopus) were used to review scholarly articles that are relevant to this current study of age-related factors, using keywords with Microsoft Office 10. The keywords included *lung cancer*, *stages*, *age*, *age at diagnosis*, *5-year survival*, and *factors that contribute to lung cancer* within the 5 years between 2015 and 2020. To elucidate the lack of understanding about age-related factors that contribute to lung cancer, a literature review is included to provide a summary of each finding. The statistical analyses answer the three research questions of this dissertation: (1) Is there a significant association between age at diagnosis and the stages of presentation of lung cancer? (2) Is there a significant association between the stage of lung cancer and demographic factors? (3) Is there any significant relationship between the stage of lung cancer, age at diagnosis, and demographic factors? The results from this dissertation provide additional information to

existing published research studies. Age-associated factors that contribute to lung cancer are understudied in the research community. However, in order to find the age-associated factors that contribute to lung cancer, this study uses age at diagnosis of lung cancer and the stages of lung cancer. Since there is little current research available to identify age as a marker for early detection of lung cancer, this study includes information about the proliferation of lung cancer, stages of lung cancer, factors that affect stages, and how age can be used as a potential marker for early detection of lung cancer.

Lung Cancer

Cancer is an abnormal proliferation of the cells in human tissues and organs and a type of disease caused by uncontrolled cell division (Toh et al., 2017). The leading cause of cancer deaths in the United States is lung cancer (Chu et al., 2018; Gingras, M., 2018; Mayo Clinic, 2019). Lung cancer is a type of cancer that starts when cells in the body begin to grow out of control in the lungs (ACS, 2019). The lung is the primary organ of the respiratory system, and the primary etiology of lung cancer is exposure to tobacco smoke (Bakulski et al., 2019; Chu et al., 2018; Dong et al., 2019; Ryan, 2018). Lung cancer is usually diagnosed at an advanced stage, and approximately 70% of patients are diagnosed with NCLS (Gyoba et al., 2016; Lozano et al., 2018). The overall survival rate for patients is approximately 15% at an advanced stage (AJCC, 2020). The two most common NSCLC are adenocarcinomas (~50%) and squamous cell carcinoma (~40%) (AJCC, 2010; Lozano et al., 2018). Patients with NSCLC are often diagnosed with an advanced stage of disease (Jin et al., 2018; Lozano et al., 2018). Adenocarcinomas start

in the cells that secrete substances such as mucus and occur mainly in both current and former smokers.

Cancer Staging

The cancer staging is based on three factors: tumor (T), node (N), and metastases (M). The staging system is maintained by the American Joint Committee on Cancer (AJCC) and the Union for International Cancer Control (UICC; ALA, 2020). The seventh edition of the TNM classification of lung cancer was published and implemented at the beginning of 2010, and the stages of lung cancer in this study were based on the seventh edition of the TNM classification. The primary sites of lung cancer are defined as carcinomas of the lung that arise from the alveolar, trachea, bronchi, visceral plural, the alveolar lining cells of the pulmonary parenchyma, or from the mucosa of the tracheobronchial tree (AJCC, 2010). The trachea (also known as windpipe) lies in the middle mediastinum, divides into the right and left main bronchi (the airways), and extends into the right and left lungs (AJCC, 2010). The bronchi then subdivide into the lobar bronchi in the upper, middle, and lower lobes on the right and upper and lower lobes on the left. The lungs are covered in membranes called the visceral pleura. The mediastinum contains structures in between the lungs, including the heart, thymus, great vessels, lymph nodes, and esophagus. From the great vessels of the primary site, the cancer cells travel through the aorta, superior vena cava, inferior vena cava, main pulmonary artery, and intrapericardial segments of the truck of the right and left pulmonary artery to the lymph nodes, and metastasize to different organs (AJCC, 2010).

The stages of lung cancer can be based on clinical (cTNM) and pathological (pTNM) staging. Clinical staging is a system that is based on cTNM, which is staging based on the evidence acquired before treatment, including physical examination, imaging studies, laboratory test, and staging procedures such as bronchoscopy or esophagoscopy with ultrasound directed biopsies (AJCC, 2010). Pathologic staging is another staging system that uses the evidence acquired before tested, supplemented or modified by the additional evidence acquired during and after surgery. The pathologic staging provides additional precise data used for estimating prognosis and calculating end results. According to the seventh edition of the TNM classification, approximately 50% of all NSCLC are localized or locally advanced at the time of diagnosis, and the rest of NSCLC are metastasized. In contrast, 80% of all SCLC are metastasized, and 20% SCLC are initially localized to the hemithorax and are usually locally advanced tumors (AJCC, 2010).

The staging describes the severity of individual cancer based on the extent of the original (primary) tumor size as well as the spread of cancer in the body (AJCC, 2010). The TNM staging system has traditionally been used for NSCLC, although it is supposed to be applied also to SCLC (AJCC, 2010). Understanding the stage of lung cancer based on the age at diagnosis will not only help health professionals to develop a prognosis and design a treatment plan for individual patients but will also inform vulnerable populations to get preventive screening as early as possible.

According to the seventh edition of lung cancer classification, occult carcinoma stage (primary) tumors are tumors that cannot be identified and classified as Tx–N0–M0.

The letter T stands for tumor size and location of the original primary tumor. For example Tx (primary tumor cannot be evaluated), T0 (no evidence of primary tumor), Tis (carcinoma in situ—early cancer that has not spread to neighboring tissue), and T1–T4 (size and extent of the primary tumor) (AJCC, 2010). The letter T category describes tumor size, how deep the tumor has grown into the organ, and the extent of tumor growth into nearby tissues. For example, the two letters TX means the tumor cannot be measured, T0 means there is no evidence of a primary tumor, and Tis means *in-situ* or pre-cancerous cells are growing only in the most superficial layer of tissue without growing into deeper tissues. The numbers after T, N, and M (such as T1, T2, T3, T4:N1, N2, N3; and M1a, M1b) describe the tumor size and the amount of spread into nearby structures (ACS, 2019). The higher the number, the larger the tumor size, and the greater the extent of node and metastasis into nearby tissues. The stage T1 is ≤ 3 cm surrounded by lung/visceral pleura that has not involved the main bronchus. T1 has been subclassified into T1 (≤ 2 cm) and T1b (>2–3 cm) in size. T2 has been subclassified into T2 (> 3– ≤ 5 cm) in size and involves the main bronchus without carina, regardless of the distance from carina, visceral pleural invasion, atelectasis or post abstractive pneumonitis extending to hilum (ACS, 2019). The T2a is >3–5 cm in size, T2b is >5–7 cm in size. T3 is >7 cm in size and is the largest tumor that involves chest wall, pericardium, phrenic nerve, or satellite nodules in the same lobe (AJCC, 2010). T4 has multiple tumor nodules in the same lung, but a different lobe has been reclassified from M1 to T4.

The tumor cells of each histological type release certain protein biomarkers into the bloodstream, which play an essential role in carcinogenesis (Zamay et al., 2017).

Tumor biomarkers are divided into (1) genetic, epigenetic, proteomic, metabolic, DNA and RNAs circulating in blood plasma, (2) exosomal microRNAs, (3) synthesis profile and level of miRNAs, (4) protein biomarkers, (5) circulating tumor cells, and (6) immune, stromal, and endothelial cells (Zamay et al., 2017). Biological materials such as tumor tissues, blood, exhaled breath condensate, sputum, and urine are generally used for non-invasive detection of LC biomarkers (Zamay et al., 2017).

Lung cancer includes cancer cells in the lungs and also cancer cells that have spread to any lymph nodes. If the lung cancer is in the lungs and has not spread to lymph nodes, it can describe as stage 1. The lymph node is abbreviated as the letter (N), which can be considered as noncancerous (benign) or cancerous (malignant). When cancer cells break away from a tumor, they travel to other areas through bloodstream or the lymph system, and surviving cancer cells may end up in lymph nodes (ACS, n. d.). Normal lymph nodes are tiny and can be hard to find. However, when those lymph nodes are infected, inflamed, or cancerous, they can get large (ACS, n. d.). If there are a lot of cancer cells in a node, it can be seen easily as a large mass (ACS, n. d.). According to the Mayo Clinic, lymph nodes that are 30 millimeters or larger are more likely to be cancerous than smaller lymph nodes (Mayo Clinic, 2019). The risk of cancer diagnosis with a large-mass lymph node can depend on the age of an individual, because the mutation increases in cancer development as age increases. The chance that a lymph node becomes lung cancer is less than one percent for people younger than 35 years (Mayo Clinic, 2019). Cancer can appear in the lymph nodes in two ways: it can either start there at the organ or it can spread there from somewhere else (ACS, 2019). The letter NX

means cancer cells in the nearby lymph nodes cannot be evaluated, and N0 means nearby lymph nodes do not contain cancer cells. The numerical numbers after the letter N (N1, N2, and N3) describe the size, location, and number of nearby lymph nodes affected by cancer. Stage N1 means ipsilateral peribronchial or hilar nodes and intrapulmonary nodes (ACS, 2019). Stage N2 means ipsilateral mediastinal and subcarinal nodes.

Stage N3 is contralateral mediastinal or hilar. The letter N stands for nodes that tell whether cancer has spread to the nearby lymph nodes (AJCC, 2010): for example, N0 (no regional lymph node involvement—no cancer found in the lymph nodes), and N1–N3 (involvement of regional lymph nodes—number and extent of spread). It is important to know whether the lung cancer has spread to the lymph nodes around the lung to diagnose the stages of cancer. Regional lymph nodes are the sites where lymph nodes extend from the supraclavicular region to the diaphragm. During the past three decades, two different lymph maps have been used to describe the regional lymph nodes potentially involved in lung cancers (AJCC, 2010). The first lymph map was proposed by the late professor Dr. Tsuguo Naruke to Japanese officials and is used primarily in the Japan Lung Cancer Society (AJCC, 2010). The second lymph map was proposed by the Mountain-Dresler modification of the American Thoracic Society lymph node map and is used primarily in North American and Europe (AJCC, 2010). However, some locations of lymph nodes differ between the two maps, especially in nodes located in the paratracheal, tracheobronchial angle, and subcarinal areas (AJCC, 2010). To reconcile the discrepancies between the two maps, the International Association for the Study of Lung Cancer (IASLS) proposed a new lymph node map that provides more detailed

terminology for the anatomical boundaries of lymph node stations (AJCC, 2010). The latest new lymph node map proposed by IASLS is now the means of describing regional lymph node involvement for lung cancers (AJCC, 2010). However, the use of lymph node zones for N staging remains investigational and needs to be confirmed by future prospective studies.

The letter M category describes whether cancer has metastasized to distant parts of the body (ACS, 2019). According to the AJCC 7th edition, metastases are found in most pleural effusions and are due to the size of the tumor (AJCC, 2010). Lung metastasis is a cancerous growth in the lung that has spread from its primary site to other places in the body (ALA, 2020; Schueller & Herold, 2003). For example, stage M0 indicates no distant metastatic, i.e., cancer has not spread to other parts of the body. The stage M1 indicates that distant metastases were found and subdivided into M1a and M1b. M1a has been reclassified as malignant pleural, pericardial effusions, and separate tumor nodules in the contralateral lungs. In addition, nodules that are found in the contralateral lung are also classified as M1a. M1a includes malignant pleural effusion with a median overall survival of eight months in 448 patients and contralateral lung nodes that had an overall survival of ten months in 362 patients. M1b classified as distant metastases in extrathoracic organs (AJCC, 2010), and median overall survival was 6 months in 4343 patients.

Factors That Can Affect the Stage of Cancer

The values of T, N, M are not the only criteria that can determine the stage of lung cancer, but other factors such as grade, cell type, tumor location, tumor markers level, performance status, patient age, and gender (AJCC, 2010).

Grade

In general, for most cancers, the grade is measured by the abnormality based on the appearance of the cancer cells (AJCC, 2010). The cancer grade is usually assigned a number on a scale of 1–3, with the larger number being the highest grade or undifferentiated cancer. Low-grade (well differentiated) cancers are characterized by cells that look largely the same as cells from normal tissue. High-grade or poorly differentiated cancers are characterized by cancer cells which look very different from normal cells.

Cell Type

Some cancers can be made up of different types of cells and it can affect treatment and outlook (ACS, 2019). Therefore, it can be used as a factor of staging. There are eight types of lung cells: [(i) alveolar type 1 cells [8%], (ii) alveolar type 2 cells [16%], (iii) capillary endothelial cells [30%], (iv) pneumocytes type 1 (v) pneumocytes type 2 (vi) alveolar macrophage, (vii) alveolar ducts, and (viii) bronchioles] (Crapo et al., 1982). Although some molecular abnormalities of cells are used to stratify patients for treatment, none of these cells are being used for lung cancer staging (AJCC, 2010). The tumor location is another factor of staging because it affects the prognosis of cancer. Lungs are two spongy organs located on each side of the chest that take in oxygen during

inhalation and release carbon dioxide during exhalation (Mayo Clinic, 2019). The right lung is shorter and wider than the left lung. The right lung consists of three lobes (upper, middle, and lower) and the left lung consists of two lobes (upper and lower). For example, the stage of lung cancer can depend on the lobe—whether the cancer is in the upper, the middle or lower third of the lungs, or overlapping lesion of the lung.

Smoking

Lung cancer is caused by both internal and external risk factors (NCI, n. d.). Internal factors are things that cannot be controlled such as age, sex, and family history. However, external factors such as cigarette smoking and exposure to carcinogens can be controlled. Having several risk factors can make a person more likely to develop lung cancer, such as an older person who continues to smoke cigarettes. The longer a person smokes, the greater the risk of lung cancer (ACS, 2019). Although age cannot be controlled, age-related risk factors can be controlled, such as reducing an avoidable exposure to carcinogens such as changing a lifestyle by cessation of smoking to delay the development of cancer. In 2019, the estimated number of new cases of lung and bronchus cancer were 228,150 and of these new cases, 62.5% were predicted to die (NCI, n.d.). The number of new cases and the percentage of predicted deaths have not significantly decreased over the past years, and lung cancer is still lethal both nationally and internationally.

The main function of the lungs is to deliver and convert air to oxygen from the airway through pulmonary ventilation to the bloodstream (Mayo Clinic, 2019). However, inhalation of carcinogens from cigarette smoke that travels through the pulmonary

ventilation to the bloodstream attack normal healthy cells and change their chemical compounds that lead to cell mutations (Bakulski, Dou, Lin, Long & Colacino, 2019). Pulmonary ventilation provides air to the alveoli for gas exchange and delivers oxygen to the bloodstream during inhalation and eliminates carbon dioxide from the lungs during exhalation (Mayo Clinic, 2019). However, when the lungs receive carcinogens through contaminated air from the pulmonary ventilation, the alveolar-capillary membrane carries out carcinogen-contaminated oxygen molecules to the bloodstream and returns carcinogen-contaminated carbon dioxide molecules to the lungs. Elderly populations are vulnerable to lung cancer, and it is a public health problem, especially when the aging population is growing and life expectancy is increasing in the United States (Battle, 2007).

According to the Centers for Disease Control and Prevention (CDC), people who smoke cigarettes are 15–30 times more likely to get lung cancer or die from lung cancer than those who do not smoke (CDC, 2019). Cigarettes contain at least 69 carcinogens that promote cancer in humans (Kathuria, Gesthalter, Spira, Brody, & Steiling, 2014; Izzotti et al., 2016; National Toxicology Program, 2016; Pu, Xu, Zhang, Yuan, Hu, Huang, & Wang, 2017). Nicotine, one of the carcinogens in cigarettes, is addictive to the brain (Battel, 2007). Because of the unpleasant side effects of smoking cessation, many people are unwilling to stop smoking. However, the good news is that since alternative methods are available to replace cigarette smoking and may reduce the desire to smoke cigarette. These alternative methods such as the nicotine patch and e-cigarettes are available to stop smoking, but studies have found that nicotine patch and e-cigarettes contain harmful

chemicals such as formaldehyde and may serve as delivery agents that deposit deeply in the lungs and are known to cause lung damage (Salamanca et al., 2018). Formaldehyde is absorbed from the nose to the upper part of the lungs. Repeated exposure to formaldehyde vapors at 40 ppm, 6 hours/day, 5 days/week for up to 13 weeks produced 80% mortality in an animal study with mice (ATSDR, 1999). Thus, cessation of cigarette smoking is the only effective way to reduce the rate of mortality and morbidity caused by lung cancer (Akushevich, Kravchenko, Yashkin & Yashin, 2018). Cigarette smoking is one major risk factor for lung cancer, and cigarettes contains at least 250 known harmful substances. Of these substances, at least 69 of them are carcinogens that are linked to lung cancer (National Toxicology Program, 2016). The longer an individual smoke cigarette, the more carcinogens accumulate in the lungs. Carcinogen-contaminated oxygen is then released into the blood stream (Bakulski et al., 2019; Dong et al., 2019).

Age

Despite medical advances, screening for early detection of lung cancer is failing because of a lack of knowledge about how age-related factors contribute to lung carcinogenesis and insufficient knowledge of how fast cancer grows and spreads. For example, lung cancer is already at a late stage when cancer tissue on a computed tomography (CT) scan is found (Zamay et al. (2017). The quantification of cancer cells' growth rate can be determined by doubling time (DT) (Mehrara, Forssell-Aronsson, Ahlman, Bernhardt, 2007). The cancer cells' growth rate is based on doubling-time (Mehrara, Forssell-Aronsson, Ahlman, Bernhardt, 2007). The rate of growth in the size of the lung nodules is much faster for malignant than for benign nodules (Harris et al.,

2012). Aria et al. (1994) reported that rapidly growing tumors tend to have a poorer prognosis than slowly growing tumors. Although the elderly population is more sensitive to carcinogens because of the lower physiological reserve capacity and slower immune system response, it is unclear whether elderly populations are more likely than younger populations to have rapidly growing tumor doubling time (Fougère et al., 2018). These cells get instruction from a gene in the DNA of a molecule to make a protein that is essential for life and used by the cell to perform certain functions: whether to grow or to survive and perform a different job to help lung function. These cells contain genes that are a portion of DNA (deoxyribonucleic acid). DNA carries genes (genetic information), and changes in key genes cause the cells to be in disorder, such as developing cancer (Crapo et al., 1982; Shao et al., 2018). Increasing knowledge in the association of how different age-related factors contribute to the growth of different types of lung cancer cells will help us understand the biology of lung cancer.

Age Differences in Cancer

Studies found lung cancer is the leading cause of cancer death between ages 60 and 79, and 80%–85% of them were caused by NSCLC (ACS, 2019; Jin et al., 2018; Fougère et al., 2018). Age could be the primary risk factor for major human pathology, as aging is the time-dependent physiological functional decline that affects most living organisms. It affects mutations and is the most profound risk factor for many non-communicable diseases such as cancer (Xia et al., 2017). In order to determine the association between age and molecular biomarkers, the American Federal of Aging Research (AFAR) proposed criteria for biomarkers of aging (1) it must predict the rate of

aging, (2) it must monitor a basic process that underlines the aging process, not the effects of the disease, (3) it must be able to be tested repeatedly without harming the person, and (4) it must be something that works in humans and in laboratory animals (AFAR, n. d.). The molecular biological materials should predict the rate of aging and must monitor a basic process that underlies the aging process. According to the National Cancer Institute, approximately 82% of new lung cancer diagnoses are in people aged 55 to 84. The average age at the time of diagnosis is approximately 70 years old (NCI, n. d.). Although the health effects of carcinogens affect all age groups, the elderly population is more sensitive to exposure to carcinogens (Fougère et al., 2018). As aging causes a biological system to decline, assessing age-related lung cancer helps explain that aging is one of the risk factors that influence the development of early cellular epigenetic alterations involved in carcinogenesis (Jin et al., 2018; Xia & Han, 2018). Cancer occurs when cells have multiple mutations; 90% of cancers are caused by mutations; and the incidence of cancer increases as age increases (Griffith et al., 2000, ACA, 2015; Hammond, 2015; Adams et al., 2015; Swanton et al., 2015; WHO, 2019). The aging population is growing, and more than 70% of cancer-related deaths occur in the elderly population (Fougère et al., 2018; McClelland et al., 2016). Increasing knowledge of the causation of lung cancer, assessing factors affecting the biological initiation, and progression of the disease will bring positive social changes to our society.

Screening

Because of the age threshold of lung cancer begins at 65–74, the current lung cancer screening targets individuals beginning at age 55 (Annangi et al., 2019). As the

incidence of cancers and diseases increases with age (White, 2014; Yang et al., 2018), age could be a valuable tool to measure physiological age, assess healthy aging, and predict risk factor of cancers and diseases (AFAR, n. d.; McClelland et al., 2017). Our cells become less efficient with age at performing normal functions (Wagner et al., 2016), and age-associated factors such as the decline of immune function may lead to increased cancer incidence (Chen et al., 2019). For example, the accumulation of degradative processes that are reinforced by multiple alterations and damage within molecular pathways can weaken the immune system and make a person less able to defend against infection and disease (Wagner et al., 2016). Yang et al. (2018) found and association between circular RNA (cRNA) in aging and the cRNA in diseases such as cancer.

Biomarkers

Based on the AFAR criteria, Xia et al. (2018) investigated biomarkers of aging using telomeres, DNA repair, epigenetic modifications, transcriptome profiles, non-coding RNAs, metabolism, protein metabolism, lipid metabolism, oxidation stress, and mitochondria (Xia et al., 2017). Xia et al. (2017) found that telomeres are ribonucleoprotein complexes at the end of chromosomes and become shorter after each replication. The enzymes are responsible for its replication and shortness of telomeres. Thus, the length of telomeres in leukocytes has been associated with aging and life span as well as age-related diseases such as cardiovascular diseases, cancer, and neurological disorder (Xia et al., 2017). Aging has implications for the link between DNA damage and repair because of the accumulation of senescent cells or genomic rearrangements (Xia et al., 2017). The age-related changes in DNA methylation patterns are measured by the

epigenetic clock among the best-studied aging biomarkers (Xia et al., 2017). Researchers found that cell-to-cell expression variation (measured by single-cell RNA seq of high-dimensional flow cytometry sorted T cells) is associated with aging and disease susceptibility (Xia et al., 2017).

Metabolism

MicroRNAs (miRNAs) are small non-coding RNAs that regulate a broad range of biological process, including metabolism and aging (Xia et al., 2017). Among the miRNAs, circulating miRNA (c-miRNA) are stable in plasma. The miR–34a was the first miRNA with an altered expression pattern during mouse aging. The expression has been found to correlate with age-related hearing loss in mice and humans (Xia et al., 2019). Thus, the association between age and DNA methylation can be extended to study age-related diseases. RNA-seq, non-coding RNAs are a broad range of biological processes including metabolism and aging (Xia et al., 2017). In protein metabolism, protein carbamylation is one of the non-enzymatic post-translational modifications that occur throughout the whole life span of an organism. The tissue accumulation of carbamylation in proteins increases as age increases and is believed to be a hallmark of molecular aging age-related disease (Xia et al., 2017). In lipid metabolism, triglycerides are found to increase monotonously with age, and phophosphingolipids in serum sample are found as a putative marker of healthy aging (Xia et al., 2017).

Oxidative Stress

Oxidative stress and mitochondrial dysfunction have been a class of aging marker as the products of oxidative damage to proteins include 0-tyrosine, 3-chlorotrosin, and 3-

nitrotyrosin. Oxidative stress is an imbalance of free radicals and mainly produced in mitochondria. Dysfunctional mitochondria can contribute to aging independently of reactive oxygen species. As age is a biological process of life that spans from birth to death (Xia et al., 2017); physiological functional can decline as age increases. Yang et al. (2018) stated that the specific cRNAs were identified in many cancers, including lung cancer (NSCLC), and these cRNAs are associated with age as well as diseases (Yang et al., 2018). However, individuals of the same age may not age at the same rate (Xia et al., 2017). For example, some people who reach 85 years old are in good physical and mental health while others may have difficulties (AFAR, n. d.).

DNA Methylation

Although the link between the age of individual and cancer is complex, researchers found some age-associated epigenetic modifications such as the decrease in DNA methylation and an increase in promoter-specific CpG islands methylation are also features of cancer (Fougère et al., 2018). In order to determine whether there is any influence of age on the cell biological methylation profile altered during tumorigenesis, Fougère et al. (2018) investigated the association between P16 gene expression (protein inhibitors that are often silenced during carcinogenesis) and promoter-specific CpG islands methylation. Fougère et al. (2018) found that P16 significantly increases with age, and DNA methylation significantly decreases with age after exposure to PM (Fougère et al., 2018).

In the DNA methylation study, only three CpG sites could predict age with a mean absolute deviation from the chronological age of less than 5 years (Xia & Han,

2018). The elderly population experiences a complex relationship with the environment, since they are more vulnerable to carcinogens because of a lower physiological reserve capacity, a slower response to the immune system, and less ability to tolerate stress (Fougère et al., 2018). The degenerative pathologies such as cancer are directly caused by genetic modifications that are also found in the biology of aging (Fougère et al., 2017). In a DNA methylation study, Bakulski et al. (2019) found that exposure to cigarette smoke is associated with altered DNA methylation throughout the genome. The abnormal growths of cells can transform into cancer cells in any part of the human body and result in malignant tumor formation in the organs (Shao et al., 2018). Cell proliferation is another factor that contributes to carcinogenesis. It is an essential process of normal tissue development that results in an increasing number of cells. It is defined by the balance between cell divisions and cell loss through cell death or differentiation (Yokoyama et al., 2019). However, aberrations in cell proliferation can give rise to malignant transformation and cancer pathology (Jone & Baylin, 2007; Yokoyama et al., 2019). The main difference between a mutagen and carcinogen is that a mutagen causes a heritable change in the genetic information, whereas a carcinogen causes or promotes cancer in humans (Griffiths et al., 2002).

Biomarkers

Millions of human cells in a human body are linked to age as the properties of chemistry change with aging (Zagryazhskaya & Zhivotovsky, 2014). Yet, previous biology research studies found aging-related biomarkers in the gene, molecule, and protein that can also be found in biological materials such as telomeres, proteomics,

cytokines, etc. (Bai, 2018; Hayashi, 2017; Xia & Han, 2018). More than 90% of tumor cells were associated with telomerase activity. Shortening in telomere is caused by a mutation that is linked to an increased risk of aging-related diseases and morality (Hayashi, 2017, Shao et al., 2018). As numerous degenerative pathologies such as cancers and diseases are directly caused by mutations in cells, DNA methylation and cell proliferation (Fougère et al., 2018) could be more representative of an individual's health status than chronological age (Bai, 2018).

According to the American Federation for Aging Research, in order to use the age of an individual as a predictor of ill health, the markers have to meet four criteria: (1) can predict the rate of aging, (2) can monitor the basic process that underlies the aging process, (3) can be tested repeatedly without harming the person, and (4) can work with human and laboratory animals (Bai, 2018). As age is one of the essential variables in many public health research studies, studies on aging can be reproduced and may be well suited for prospective studies involving larger cohorts, which have a potential major advantage for public health. Using age-related biomarkers in research studies has advantages because they are stable, reliable, and can be measured in a variety of biological specimens (Sun et al., 2018).

Although people of the same age may not age at the same rate (White, 2014), aging markers can be a valuable tool to measure physiological age to assess how the biology of aging affects lung carcinogenesis. However, not all human organs react to exposure to carcinogens the same way, since different organs have different functions (Popović et al., 2018). For example, lungs are exposed to carcinogens through the

inhalation of carcinogen-contaminated air, while skin is exposed to a radiated carcinogen through direct contact with the sun. Thus, age-related risk factors that contribute to lung cancer may be a valuable tool for measuring the dose of exposure to carcinogens over a period that an individual is exposed to carcinogens.

Mutagenesis

In general, carcinogenesis needs at least six or seven mutagenic events (over a period of 20–40 years), which can be described in three different steps: initiation, promotion, progression (Battle, 2009; Weiss, 2004). As carcinogenesis can take years to develop into cancer, cancer prevention can begin as early as in utero (Battle, 2009). In the investigation of molecular biomarker screening for early detection of lung cancer using positive predictive value (PPV), researchers found that circulating miRNA profiles of lung cancer are stable in blood and strongly affected by age, gender, smoking status (Ameling et al., 2015; Atwater & Massion, 2016; Sun et al., 2018; Zagryazhskaya & Zhivotovsky, 2014). For example, Zagryazhskaya & Zhivotovsky (2014) found that miRNAs play an important role in cancer cell development and altered in lung cancer cells during aging.

Dong, Zhang, He, Lai, Alolga, Shen…Christiani (2019) performed integrative analyses of clinical information, DNA methylation, and gene expression data using 825 lung cancer patients with early-stage cancer (stage 1 and II) from five cohorts. They focused on developing prognostic models with trans-omic biomarkers for early-stage lung adenocarcinoma (LUAD). They used the iCluster plus machine learning approach with a joint latent variable model for fusing clinical variables that includes two age

groups: age < 65 and age ≥ 65 (based on the definition of elderly using United Nation standard) and trans-omic biomarkers (12 DNA methylation and 7 gene expression probes)., Dong et al. (2019) found that trans-omic biomarkers have different prediction ability, significant heterogeneity, and diverse effects on early-stage lung adenocarcinoma prognosis. The miR–346 expression was upregulated in NSCLC patients with elder age, bigger tumor sizes, smokers, positive lymph node metastasis, and advanced stage. Each of these variables is assumed to be a predictor of overall survival in NSCLC patients (Dong et al., 2019). In the lung cancer cohort study of Haung et al. (2019), the median age at lung cancer diagnosis was 69.8 (range between (53.6–82.0) using circulating markers of cellular immune activation as biomarkers for immune regulation in a pre-diagnostic blood sample (Haung et al., 2019).

Studies found age-associated increases in the initiation and progression of the mutant stem and progenitor clones. This age-associated increase in the initiation and progression of the mutant stem and progenitor highlights the roles of stem cell quiescence, replication-associated DNA damage, telomere shortening, epigenetic alterations, and metabolic challenges as determinants of stem cell mutations and clonal dominance in aging (Adams et al., 2015). A gene mutation is a permanent alteration in the DNA sequence that can further be classified into hereditary mutations and acquired (somatic) mutations that can occur at some time during the life of individuals (Jin et al., 2018). Cancer is thought to include gene mutations, and mutation is an inevitable consequence of normal aging that depends in part on lifestyle (Yokoyama et al., 2019). A decrease in genome integrity and impaired organ maintenance increases the risk of cancer

development (Adams et al., 2015). In the study of age-related remodeling of oesophageal epithelia by mutated cancer drivers, Yokoyama et al. (2019) found that somatic mutations were detected in 96% of physiologically normal oesophageal epithelia (PNE) tissues. Accumulated errors in signaling the cell, and abnormalities that can cause the cell to stop its normal function are associated with cancer (Jin et al., 2018; Mayo, 2017).

Chemical Carcinogens

This section focuses on one of the three main cancer-causing agents that can cause mutations to the cell. When this cancer-causing agent (chemical carcinogen) enters our bodies through ingestion or inhalation, it often results in the activation of a compound to reactive state (Battle, 2009). The reactive molecules are capable of damaging DNA and can lead to mutations that contribute to carcinogenesis (Battle, 2009). The good news is that somatic mutations that are caused by lifestyle behavior, occupational exposure, and environmental exposure cannot be passed on to the next generation (Genetics Home Reference, 2019).

Mutations in the cell genome lead to proteins changes in the body that regulate cell growth and are caused by carcinogens (Adams et al., 2015; Yokoyama et al., 2019). Studies have found that at least 90% of carcinogens are mutagens, and these cell mutations increase as age increases (Adams et al., 2015; Enge et al., 2018; Smith et al., 2009; Swanton et al., 2015; Weiss, 2002; Yancik et al., 2005). According to Enge et al. (2018), as organisms age, cells accumulate genetic and epigenetic error that leads to cancer cell development. Mutations can be subtyped into gene mutations, constitutional mutations, somatic mutations, chromosomal aberrations, structural abnormalities, and

numerical abnormalities (Jones & Baylin, 2002; Shao et al., 2018). Mutations can increase in number and size with aging and ultimately replace almost the entire epithelium in the elderly patients (Yokoyama et al., 2019). Although cancer affects anyone and almost any part of the body, elderly individuals with high risk exhibited a higher mutation density (NCI, n. d., Yokoyama et al., 2019).

Gender Differences

According to The Cancer Atlas, lung cancer is the leading cause of cancer death among men in over half of the world (The Cancer Atlas, 2018). Research studies show the evidence of gender differences: lung cancer affects more men than women, and women patients have better survival rates at every stage of the disease (Xiao et al., 2016). In contrast, a join-point analysis study Siroglavić et al. (2017) found an increased incidence rate in women and a decreased incidence rate in men in Croatia. The gender differences in mutation burden might be the reason why there is a clinical gender difference in the outcome of lung cancer cases (Xiao et al., 2016). The highest death rates and shortest survival times in most cancers are found in African American men (McClelland et al., 2017). In the study of Genome-wide association (GWAS), Bossé et al. (2019) identified genetic factors robustly associated with lung cancer and stated that some genetic risk loci have refined to more homogeneous subgroups of lung cancer patients such as histological subtypes, smoking status, gender, and ethnicity.

Racial and Ethnicity Differences

Cancer outcomes vary among different racial and ethnic groups. Racial and ethnic disparities exist in cancer incidence and survival (Stram et al., 2019; Robbine et al.,

2015). For example, in the United States, African Americans have a higher rate of mortality than Whites for most common cancers and cancer overall (Stram et al., 2019; Robbins et al., 2015). Stram et al. (2019) stated that the differences were more evident at relatively low levels of smoking intensity (fewer than 20 cigarettes per day) than at higher intensity. In the overall risk study of lung cancer and lung cancer subtypes, Stram et al. (2019) found that African American and Native Hawaiians men had higher incidences of lung cancer than Japanese Americans and Whites (Stram et al., 2019). White men (40%), Japanese American men (54%), and Latino men (70%) had lower excess relative risk (ERR) of lung cancer for the same quantity of cigarettes smoked (Stram et al., 2019). The risk of NSCLC with adenocarcinoma and squamous cell carcinomas was highest in African Americans, while the risk of small cell lung cancer was highest in Native Hawaiians (Stram et al., 2019). The potential reasons why African Americans are more susceptible to NSCLC include that they are less likely to receive interventional care (McClelland et al., 2017), are more likely to live in working-class communities (Ryan, 2018), are at increased risk for exposure to environmental hazards (Ryan, 2018), and have genetic factors that make them more susceptible to the effects of carcinogens of chemical exposure (Ryan, 2018). Several studies show that African Americans are typically diagnosed with lung cancer at earlier ages compared with Whites and other races (Robbin et al., 2015; Ryan, 2018). Using SEER Medicaid and Medicare data, McClelland et al. (2016) found that African Americans are 42% less likely than Whites to receive radiation therapy, which could be because African American men fear exposure to low-dose radiation.

Geographic Differences

There are striking geographic differences in the rate of morbidity and mortality caused by lung cancer in different geographic regions. The national diversity reflects both the presence of local risk factors for lung cancer and the extent to which effective lung cancer control measures have been implemented. Much of the observed variation in recorded lung cancer cases in different registry populations can be attributed to lifestyle, occupational exposure, and environmental factors. The American Lung Association collected incidence, survival rates, stages of diagnosis, surgical treatment, and availability of screening centers. According to the state data of the American Lung Association (ALA, 2020), the national incidence rate of lung cancer is 59.6 per 100,000, and the 5-year survival rate is 21.7%. In Kentucky, the rate of new lung cancer cases is 92.6 per 100,000 which is significantly higher than the national rate of 59.6 per 100,000. The 5-year survival rate in Kentucky is 17.6%, which is significantly lower than the national rate of 21.7% (ALA, 2020). In Louisiana, the rate of new lung cancer cases is 67.9 per 100,000 which is significantly higher than the national rate of 59.6 per 100,000. The 5-year survival rate is 17.0% which is lower than the national rate of 21.7% (ALA, 2020). In Michigan, the rate of new lung cancer cases is 64.5 per 100,000 which is significantly higher than the national rate of 59.6 per 100,000. The 5-year survival rate is 23.2% which is higher than the national rate of 21.7% (ALA, 2020). In Georgia, the rate of new lung cancer cases is 64.5 per 100,000, which is significantly higher than the national rate of 59.6 per 100,000. The 5-year survival rate is 19.3%, which is lower than the national rate of 21.7% (ALA, 2020). In Iowa, the rate of new lung cancer cases is 63.5 per 100,000

which is significantly higher than the national rate of 59.6 per 100,000. The 5-year survival rate is 19.1% which is lower than the national rate of 21.7% (ALA, 2020).

In Connecticut, the rate of new lung cancer cases is 60.2 per 100,000 which is not significantly different from the national rate of 59.6 per 100,000. The 5-year survival rate is 26.4%, which is significantly higher than the national rate of 21.7% (ALA, 2020). In Utah, the rate of new lung cancer cases is 59.6 per 100,000 which is significantly lower than the national rate of 59.6 per 100,000. The 5-year survival rate is 21.4% which is not significantly different than the national rate of 21.7% (ALA, 2020). In New Jersey, the rate of new lung cancer cases is 56.6 per 100,000 which is significantly lower than the national rate of 59.6 per 100,000. The 5-year survival rate is 25.0% which is higher than the national rate of 21.7% (ALA, 2020).

In Alaska, the rate of new lung cancer cases is 56.3 per 100,000, which is lower than the national rate of 59.6 per 100,000. The 5-year survival rate is 17.6%, which is significantly lower than the national rate of 21.7%. In Hawaii, the rate of new lung cancer cases is 46.0 per 100,000, which is lower than the national rate of 59.6 per 100,000. The survival rate is 18.7%, which is lower than the national rate of 21.7%. In New Mexico, the rate of new lung cancer cases is 39.9 per 100,000 which is significantly lower than the national rate of 59.6 per 100,000. The 5-year survival rate for New Mexico is 19.6% which is lower than the national rate of 21.7% (ALA, 2020). In California, the rate of new lung cancer cases is 42.3 per 100,000, which is significantly lower than the national rate of 59.6 per 100,000. The survival rate of 21.7% not significantly different than the national rate of 21.7% (ALA, 2020).

Carcinogenesis

Carcinogenesis, also known as cancer development, is a multistage process that can be prevented from progressing to subsequent stages (Battle, 2009; Jin et al., 2018). A cancer stem cell is small and has the capacity to generate the different cell types that constitute the whole tumor (Toh et al., 2017) that can invade adjoining parts of the body (Adams et al., 2015; Griffith et al., 2000, ACA, 2015; Hammond, 2015; Swanton et al., 2015; WHO, 2019). Normal cells can divide in the process of mitosis, repair themselves, differentiate, or die under the control of the molecular mechanism (Tyson & Novak, 2014). However, when epigenetic changes occur, such as DNA methylation and histone modifications, the normal cell may transform into cancer stem cells (Toh et al., 2017). The main difference between a mutagen and carcinogen is that the mutagen causes a heritable change in the genetic information, whereas a carcinogen causes or promotes cancer in humans. Carcinogenesis is the mechanism by which tumors occur because of mutagenic events. In general, carcinogenesis needs at least six or seven mutagenic events over a period of 20–40 years, which can be described in four different steps:

1. Initiation: cells change genetic material
2. Promotion: promoters increase the proliferation of cells
3. Malignant transformation: cells acquire the properties of cancer
4. Tumor progression: the last phase of tumor development (Roberti et al., 2019).

Jia et al. (2016) found that the expression level of miRNA in the plasma of (NSCLC) and (SCLC) was lower than in the control group and that the occurrence and

prognosis of lung cancer may be related to the expression quantity of miRNA (Jia et al., 2016). These biomarkers are up-regulated amongst lung cancer patients. As cited in Gingras (2018), although differences in biomarkers of miRNA–486 and miRNA–499 expression can be analyzed (Jia, Zang, Feng, Li, Zhang, Li, Wang, Wei, & Huo 2016), Jia et al. (2016) found no statistical significance between gender, age, history of smoking, pathologic types, differentiation types, and primary tumor extent and the expression of miRNA–486.

Cancer causes mutations in the cell genome, leading to the changes in the proteins that regulate cell growth. These changes are caused by changes to the DNA mutations in the cells (Shao et al., 2017). During cancer development, five biological capabilities are acquired: (1) mutations, (2) conquering the barriers of cell death, (3) sustaining proliferative signaling, (4) resistant to cell death, and (5) activation of the mechanism for invasion and metastatic to the other part of the body (Shao et al., 2017). Homeostasis maintains a balance between cell proliferation and cell death (Shao et al., 2017). However, when destruction occurs in homeostasis, it leads to the uncontrolled growth of cells and causes the loss of a cell's ability to undergo apoptosis, resulting in a genetic error to the DNA cycle that could develop the process of cancer (Shao et al., 2017). A gene mutation affects the cell in many ways, and some mutations stop a protein from being made. Others may change the protein that is made so that it will no longer work the way it should, which leads to cancer (American Cancer Society, 2018; Shao et al., 2017).

Cancer is a genetic disease and is caused by mutations to genes in the deoxyribonucleic acid (DNA) (NCI, 2017) Each of those individual genes signal the cell

activities to perform, to grow, or to divide (Mayo, 2017). The three main classes of agents causing cancer are chemicals, radiation, and viruses (Choudhuri, Chanderbhan, & Mattia, 2018). At least one of these exposures causes the normal cells to become abnormal, which leads to the development of cancer (Choudhuri, Chanderbhan, & Mattia, 2018). Cancer arises from almost anywhere in the human body from the accumulation of genetic mutations, or when the orderly process breaks down to cause the old or damaged cells to survive instead of forming a new cell. These extra old and damaged cells can replicate without stopping and may form tumors (NCI, n.d.). As cells can react to their direct microenvironment, cancer can also develop in complex tissue environments, which they depend on for sustained growth, invasion, and metastasis (Quail & Joyce, 2013). In the past few decades, the impetus for studies of biological materials has grown stronger in public health after the findings of markers in cancer cells. Since carcinogenesis begins at the cell levels, the biomarkers could measure genetic risk, which is caused mostly by cell mutations.

Atwater and Massion (2016) and Chu, Lazare, and Sullivan (2018) assessed whether molecular biomarkers can be used for screening programs to reduce the costs of screening and to reduce the number of individuals harmed by screening. To assess the risk, Atwater and Massion (2016) investigated biomarkers such as serum-based inflammation, circulating miRNA, and a strong positive predictive value with great specificity or negative predictive value with greater sensitivity. Atwater and Mission (2016) found that serum-based and circulating miRNA profiles are associated with inflammation marker levels and are stable in the blood. They can be used as biomarkers

of disease and may be a benefit for future study of predictive biomarkers of disease (Atwater & Masson, 2016). Chu, Lazare, and Sullivan (2018), Jia et al. (2016), Pu et al. (2017), Shen et al. (2011), Yang et al. (2015), Xing et al. (2018), Zagryazhskaya and Zhivotovsky (2014) investigated how to improve the accuracy of lung cancer screening to decrease over-diagnosis and morbidity by using blood and serum-based biological materials (Gingras, 2018). These researchers found that miRNA biomarkers are promising markers for early detection of lung cancer (Chu et al., 2018; Jia et al., 2016; Pu et al., 2017; Shen et al., 2011; Yang et al., 2015; Xing et al., 2018; Zagryazhskaya & Zhivotovsky, 2014). The results from Chu et al. (2018) and Jia et al. (2016) showed that miRNA and serum-based biomarkers can be a promising marker for the early detection of lung cancer (Chu et al., 2018; Jia et al., 2016). But as of now, Chu et al. (2018) found no high-quality clinical evidence supporting the implication of these biomarkers.

While innovation of technology increases in our society, many researchers are using technologies to help them understand the process of biological materials and how it relates to cancer development. Eggert, Palavanzadeh, and Blanton (2017) examined early detection of lung cancer using cell-free DNA, miRNA, circulating tumor cells (CTCs), LDCT, and spiral CT. The study identified expired malignant or circulating cells, and metabolites associated with specific lung cancer types. As cited in Gingras (2018), Eggert et al. (2017) stated that the diagnosis of solitary pulmonary nodules can be elucidated by the miRNA sputum via real time-polymerase chain reaction before screening with LDCT, which could detect lung cancer 1–4 years earlier than using LDCT and chest x-ray methods (Eggert et al., 2017). However, despite ongoing research and laboratory work,

there is still a lack of information and methodology regarding the gene pathways and metabolites produced, including byproducts of cancer metabolism (Eggert et al., 2017). Effective screening options are needed to provide a non-invasive approach to early detection of lung cancer using biomarkers, including circulating epithelial (CEpC), circulating tumor cell (CTCs), metabolomics, methylation markers, serum cytokine level, and neutrophil microRNA (miRNA).

Since a high rate of false-positive CT scans are found in lung cancer detection, Gyoba, Shan, Wilson, and Bédard (2016) examined miRNA biomarkers in blood, plasma, serum, and sputum for early detection of lung cancer. It was discovered that biological fluids such as whole blood, plasma, serum, and sputum can contain miRNA levels that can serve as biomarkers for detection and diagnosis of lung cancer. According to Gyoba et al. (2016), the promise of miRNA being a biomarker for the early detection of lung cancer is high because (1) it is easier and more effective to detect circulating miRNAs and other biological fluids in small quantities compared with healthy patients, and (2) miRNA levels are altered in lung cancer patients (Gyoba et al., 2016). The dysregulation of miRNA may have different pathological and physiological conditions, such as cancer patients having higher miRNA and abnormal expression (increased or decreased) in miRNA levels (Gyoba eta l. 2016). Sensitivity as a percentage is used to calculate the probability that the biomarkers are positives in cancer cases. False-positive values are calculated as specificity in percentage, which is the probability that the biomarkers are desired (Gyoba et al., 2016). After comparing sensitivity and specificity values of

different miRNAs in sputum, Gyoba et al. (2016) defined 80% or higher as a good screening and diagnostic test using biomarkers (Gingras, M. 2018).

Volume Doubling Times

Stages of lung cancer can also be based on the volume doubling time (VDT) of a nodule, which is a key parameter in lung cancer screening that can be defined as the number of days in which the nodule doubles in volume (Kanashiki et at., 2012; Honda et al., 2009; Usuda et al., 1994). Kanashiki et at., (2012), Honda et al. (2009) and Usuda et al. (1994) studied VDT of lung cancer based on annual chest radiograph screening and compared it with computed tomography (CT) screening for patients with lung cancer. Kanashiki et at., (2012) stated that VDT helps to differentiate between benign and malignant pulmonary nodules (Kanashiki et at., 2012). Shorter VDT may reflect greater histological tumor aggressiveness that may have poor prognosis (Kanashiki et at., 2012). Honda et al. (2009) investigated the difference in doubling time between squamous cell carcinoma (SCC) and adenocarcinoma of solid pulmonary cancer. Honda et al. (2009) found the median doubling time of SCC lung cancers to be less than that of adenocarcinoma. Usudu et al. (1994) used univariate analyses for survival rates and multivariate analyses for significant factors affecting survival using the Cox proportional hazard model.

Univariate analyses showed a significant difference in survival in relation to DT, age, gender, method of tumor detection, smoking history, symptoms, therapy, cell type, primary tumor (T) factor, regional node (N) factor, distant metastasis (M) factor, and stage. Usuda et al. (1994) found that DT was an independent and a significant prognostic

factor for lung cancer patients. The minimum size of lung cancer that can be detected on a chest X-ray has been reported to be 6 mm, the minimum DT was 30 days, and the maximum DT was 1,077 days (Usudu et al. (1994). The mean DT in patients with adenocarcinoma was significantly longer than in patients with squamous cell carcinoma and undifferentiated carcinoma (Usudu et al. (1994). The mean DT in patients with a T1 lung cancer was significantly longer in patients with T2, T3, or T4 lung cancer. The survival rate in men was significantly lower than that of women, and a better survival rate was associated with long DT, not smoking, N0, resection, and absence of symptoms (Usuda et al., 1994).

Knowledge Limitation

Although 80%–85% of patients with lung cancer have a history of smoking, surprisingly only 10%–15% of smokers actually develop lung cancer. The screening method reduces lung cancer mortality with a high rate of false positives. Therefore, the higher likelihood of over-diagnosis has raised the question of how to best implement lung cancer screening (Kathuria et al., 2014; Gyoba et al., 2016). Early detection with age-related factors that contribute to lung cancer may improve the diagnosis of lung cancer in early stages. Kathuria et al. (2014) found opportunities and challenges related to lung cancer screening using biomarkers and found that it could be a promising method. Developing highly sensitive and specific age-related biomarkers may improve mortality rates.

Although there is a strong relationship between lung cancer and tobacco smoke, tobacco use must be the first target for preventing risk factors for lung cancer (de Groot et

al., 2018). The most important step is early screening to detect and prevent lung cancer for high-risk populations. It is possible that biomarker screening in blood and urine could detect lung cancer, as tumors may cause a high rate of circulating miRNA (Robles & Harris, 2016). However, miRNA screening has not yet been used to detect the risk of lung cancer (Robles & Harris, 2016). New treatment options will be required if more lung cancer patients can be identified at a younger age and early stage of disease. Low-dose CT can identify a large number of nodules, however, fewer than 5% are finally diagnosed as lung cancer. By using biomarker screening of MSC–miRNA signatures, false positives can be reduced (Robles & Harris, 2016). Therefore, biomarker screening may improve the efficacy of lung cancer screening.

Theoretical Foundation

The theoretical framework of this study is the CCC, which is a framework of the National Cancer Institute (NCI, n. d.). To operationalize the use of the CCC theoretical construction in this study, three research questions examined detection as it related to the second tenet: prevention through screening of this study. Despite LDCT and X-ray detection of lung cancer, these imaging screening found lung cancer at a late stage or stage IV. While researchers are still improving the effectiveness of lung cancer screening using biological markers, early detection through age recommendation was examined for annual preventive screening through CCC. The original purpose of this CCC framework was to view plans, progress, and priorities that will help highlight knowledge gaps about causal variants and demographics factors of lung cancer. Using these three items (etiology, prevention, and detection), the risk factors for cancer at the molecular level

were also examined to prevent and detect cancers as early as possible (NCI, n. d.). The first focus of the CCC framework was the etiology of lung cancer, which included demographic risk factors (i.e. age, gender, race/ethnicity), health behavior risk factors (i.e. cigarette smoking), and the risk factor of gene-environment interactions (i.e. caused changes DNA methylation and mutations in cells) that reflected the state of health of the patient with NSCLC. Based on the etiology of lung cancer, this study investigated why older White American men were more vulnerable to lung cancer than other age groups. Also included were how mutations increase as age increase, how gender can be susceptible to lung cancer, and why African Americans have a higher risk of cancer than other racial or ethnic groups.

The second focus of the CCC framework was prevention through screening. The purpose of the Task Force is to improve the health of all Americans by making an evidence-based recommendation about preventive screenings as a part of health promotion (USPSTF, 2015). Although imaging screening using LDCT and chest X-ray has decreased the risk of lung cancer, the rates of mortality and morbidity of lung cancer remain stable and have not significantly decreased over the past decades (NCI, 2018; Patnaik et al., 2017). The U.S. Task Force recommends that LDCT screening should be discontinued once a person has not smoked for 15 years or develops a health problem that substantially limits life expectancy (USPSTF, 2015).

The third focus of the CCC framework is detection. There is limited literature on how long it takes for cancer to develop after cell mutations. Cancer development starts at the cellular level and takes approximately 20–40 years after cell mutations (Adams et al.,

2015; Wagner et al., 2016). Effectiveness in screening is crucial to preventing lung cancer. Moreover, the evaluation of demographic risk factors (age, gender, race/ethnicity), environmental risk factors (cigarette smoking and other substances), gene-environmental interaction (changes in DNA and mutations in cells), and geographical risk factor may increase our knowledge of how differences in cancer cells grow based on these risk factors. The evaluation of demographic data, health behaviors, and gene-environmental interactions helped us understand how cancers begin and how to detect it at an early stage.

This framework can be a foundation of how researchers should continue to develop appropriate strategies to prepare for the evaluation of biomarkers for early detection of lung cancer. The CCC framework may reduce the challenge of preventive screening studies by explaining the association of well-known risk factors of lung cancer at a molecular level. However, there is limited literature on how carcinogens in cigarette smoke increase the risk of mutation and how an increase in age increases the risk of mutations (Bakulski et al., 2019).

The study of gene-environment interaction (changes in DNA and mutations) that can contribute to lung cancer is understudied. Since technology has changed our understanding of cancer development at a molecular level, the framework of CCC can be used with existing findings for the early detection of lung cancer. Based on the underlying framework of CCC, both internal and external risk factors' role in lung cancer development were investigated: age, gender, race/ethnicity, and gene-environment interaction. The findings from this dissertation improved our understanding of the

epidemiology of external and internal risk factors that contribute to the development of lung cancer. In addition, the results provided information for future clinical study using a blood-based test for the early detection of lung cancer.

Transition and Summary

The main point of this study was to improve our understanding of how age-related risk factors that contribute to lung cancer help us understand the epidemiology of lung cancer. The investigation of TNM stages and the age of diagnosis presumed that the observed variation reflected the influence of cigarette smoke, age, gender, race/ethnicity, and environment as a genetic factor in lung cancer patterns. The finding helped not only minimizing the burden of lung cancer but bridged a gap in our understanding of the implication of epigenetics. Despite an improvement in imaging screening, the images that are produced by chest X-rays and LDCT screening have some drawbacks for effective screening (USPTF, 2018). Because of the lack of effective screening, lung cancer is still extremely lethal in the United States. To make age-related factors that contribute to cancer recognizable in the public health field, this dissertation has increased the knowledge of how the histology of lung cancer and TNM stages differ in age groups of population using SEER data. Chapter 2 (recent literature reviews) details how biomarkers of aging can be related to cancer, as an accumulation of carcinogens increases with aging. It is hoped that this research will bridge a gap in the lack of understanding of how age-related factor influence the epidemiology of lung cancer.

Chapter 3: Methodology

Introduction

RQ1: Is there a significant association between age at diagnosis and the stages of presentation of lung cancer after controlling for demographic factors: gender, race/ethnicity, and geographic region?

H_o1: There is no significant association between age at diagnosis and the stage of presentation of lung cancer after controlling for demographic factors: gender, race/ethnicity, and geographic region.

H_a1: There is a significant association between age at diagnosis and the stage of presentation of lung cancer after controlling for demographic factors: gender, race/ethnicity, and geographic region.

A chi-squared test of independence was performed to examine the association between the stages of lung cancer and age groups after controlling for demographic risk.

RQ2: Is there a significant association between the stage of lung cancer and demographic factors gender, race/ethnicity, and geographic regions after controlling for age?

H_o2: There is no significant association between the stage of lung cancer and demographic factors.

H_a2: There is a significant association between the stage of lung cancer and demographic factors.

The chi-squared test was used to examine the association between stages of lung cancer and demographic risk factors after controlling for age.

RQ3: Is there any significant relationship between the stage of lung cancer, age and demographic factors?

H_03: There is no significant relationship between the stage of lung cancer, age, and demographic factors.

H_a3: There is a significant relationship between the stage of lung cancer, age and demographic factors.

Multinomial regression analysis was performed the find the association between stages of lung cancer, age, and demographic risk factors.

Research Design and Rational

This study used data from the National Cancer Institute Surveillance, Epidemiology, and End Results (SEER) program (November 2010 submission). The data was on lung cancer patients recorded during 2010–2015 in the SEER database. The SEER program provides U.S. cancer statistics in an effort to reduce the cancer burdens in the U.S. population. NCI's Division of Cancer Control and Population Sciences (DCCPS) support the SRP (NCI, n. d.). The data included diagnosis in cancer cases from 2010–2015 from state registries that were based on the 2010 U.S. Census data (NCI, n. d.). The two major types of cancer registries included population-based registries and hospital-based registries, including special cancer registries. The population-based registries recorded all cases from a particular geographical area such as metropolitan and non-metropolitan areas with an emphasis on the use of the data for epidemiology. Population-

based registries were designed to determine cancer patterns among various populations, monitor cancer trends over time, guide planning, and evaluate cancer control efforts to help prioritize health resource allocations, and advance clinical, epidemiological, and health services research (NCI, n. d.).

As lung cancer data were collected on the whole study population at a single point in time, a cross-sectional study was used to examine the relationship between lung cancer, age at diagnosis, and demographic factors. This cross-sectional study provided information about the frequency of lung cancer in a population during 2010–2015. To observe the correlations between independent and dependent variables for this study, age at diagnosis, the stages of presentation of lung cancer, and demographic factors (gender, race/ethnicity, health behaviors, and geographical regions) were investigated. Although a cross-sectional study cannot demonstrate the cause-and-effect of independent and dependent variables, the result produced inferences about possible relationships to support further research.

Comparison of the age groups of patients [(45–49), (50–54), (55–59), (60–64), (65–69), (70–74), (75–79), and (80–84)] who were diagnosed with lung cancer in different stages were analyzed. To increase the accuracy of the result, the variables gender, race/ethnicity, and geographical region were included in this study using X^2 tests, odd ratio, and multinomial analysis. As bias in sample size can distort the result of the outcome, power analysis for the sample size was performed using G*Power (G*power 3.1 manual, 2017) in order to reduce the effect of bias from the sample size. The G*power analysis showed that the estimated size would be 3,670. The purpose of a cross-

sectional study was to determine whether there was any correlation between independent and dependent variables. However, this type of study can be limited in its ability to draw valid conclusions about any association or possible causality, because risk factors and outcomes are measured simultaneously.

To find consistent results, this quantitative research method included chi-squared and multinomial analyses to elucidate the association between age at diagnosis, stages of presentation of lung cancer, and demographic risk factors. As younger ages at diagnosis for lung cancer have been reported for several cancer types (Robbins et al. 2014), the result of this study may help provide a recommendation for annual screening for lung cancer with the most effective method in adults aged 45 years and older who are both smokers and non-smokers. Reducing the age limitation for preventive screening may overcome the lack of effectiveness in lung cancer screening for early detection of lung cancer. Increasing knowledge of how age-related factors contribute to lung cancer may reduce the rate of morbidity and mortality caused by lung cancer. The hypothesis of this study was to investigate how age at diagnosis may predict outcomes in a patient who is in the early stages of lung cancer, as age and mutations are the most important predictors of prognosis in lung cancer patients (Wang et al., 2019; White et al., 2015). Although there may not be any cause-and-effect between age at diagnosis and the stages of presentation of lung cancer, older patients still may have a bad a prognosis with the worst outcome. Thus, hypotheses based on age-related factors that contribute to lung cancer may encourage future early detection of lung cancer using biological material.

Data Collection

This study was conducted using publicly available data obtained by signing a Surveillance, Epidemiology, and End Results (SEER) Research Data Agreement for secondary data analysis. Data on lung cancer specific mortality and patients who were diagnosed between 2010–2015 were extracted from SEER–18 Registries (2010–2017 dataset). The SEER program collected cancer data by identifying people with cancer who were diagnosed with cancer or received cancer care in hospitals, outpatient clinics, radiology department, doctors' offices, laboratories, surgical cancers, or from other providers (such as pharmacists) who diagnose or treat cancer patients (NCI, n. d. a). After removing identifying information, data were placed into five categories: stage of lung cancer, age at diagnosis, gender, race/ethnicity, and geographical region. Data were collected on the whole study population at a single point in time to examine the relationship between age at diagnosis and the stages of presentation of lung cancer (NCI, n. d.). Cross-sectional studies showed the association between and exposure and an outcome in a population at a given point in time. Thus, it was useful to inform and plan for health screenings.

The study population comprised lung cancer patients with the International Classification of Disease for Oncology, 7th Edition. SEER collects cancer incidence data from population-based cancer registries covering approximately 34.6 percent of the U.S. population. This study collected data on patient demographics, age at diagnosis, stage at diagnosis, geographical regions, and deaths attributable to lung cancer. The geographical regions of SEER–18 registries include (1) Alaska Native Tumor Registry, (2)

Connecticut Registry, (3) Detroit Registry, (4) Atlanta Registry, (5) Greater Georgia Registry, (6) Rural Georgia Registry, (7) San Francisco-Oakland Registry, (8) San Jose-Monterey Registry, (9), Greater California Registry, (10) Hawaii Registry, (11) Iowa Registry, (12) Kentucky Registry, (13) Los Angeles Registry, (14) Louisiana Registry, (15) New Mexico Registry, (16) New Jersey Registry, (17) Seattle-Puget Sound Registry, and (18) Utah Registry. Patient demographics information identified the current patient, age, gender, race/ethnicity, and geographical regions. Characteristics of lung cancer include (1) stage of cancer, (2) size of the tumor, (3) any spread of cancer into nearby tissue, (3) any spread of cancer to nearby lymph nodes, and (4) any spread of cancer to other parts of the body. To find the association between the age and presentation of lung cancer, age at diagnosis, mortality cases caused by lung cancer, and the metastatic patterns diagnosed with lung cancer were used.

Methodology

A cross-sectional study was appropriate for inferential analyses and for generating hypotheses. Using descriptive statistics, the study assessed the frequency and distribution of lung cancer among these age groups: [(45–49), (50–54), (55–59), (60–64), (65–74), (75-79), and (80–84)] based on the stages of lung cancer (stage: I, II, III, IV). A total of 63,107 patients who were diagnosed with lung cancer during (2010–2015) or had lung cancer as a cause-specific mortality met the eligibility criteria. All the lung cancer statistical analyses were performed using the Statistical Package for Social Science (SPSS, Inc., Chicago, IL, USA) software (version 25.0) for Windows. The inferential data were expressed as chi-squared, OR, and confidence intervals to describe the sample

under study. The incidence of lung cancer and age-related factors are important to assess the burden of disease in a specified population and in planning and allocating health resources. The available data on the incidence of lung cancer was obtained for the period 2000–2015 from the National Cancer Institute's SEER Registries database using SEER*Stat (version 8.3.6).

The demographic variables included age at diagnosis [(45–49), (50–54), (55–59), (60–64), (65–74), (75–79, and (80–84)], gender (women and men), race/ethnicity (African American and White American), and geographical regions [(metropolitan counties: counties in metropolitan areas with a population greater than 1 million, counties in metropolitan areas with a population between 250,000 and 1 million, counties in metropolitan areas with a population of less than 250,000) and non-metropolitan counties: counties that adjacent to a metropolitan area and not adjacent to a metropolitan area)] and the presentation of lung cancer stage (I, II, III, IV). To reduce potential bias in a cross-sectional study, non-responses and data on un-staged lung cancer were excluded from the study. The exclusion criteria were (1) patients without a pathological diagnosis, (2) cases from 2016 and 2017 because TNM stages were not identified (3) patients without any TNM information, and (4) patients with non-cancer specific death (missing data or unknown or un-staged). The primary endpoint of the study was calculated from the date of diagnosis to the data of cancer-specific death or the last follow-up. The study was approved by Walden IRB.

A request was made to have access to SEER data (using the link https://seer.cancer.gov/seertrack/data/request/) after receiving Walden IRB approval. The

SEER program processed the data usage request after receiving a signed agreement and providing a personalized SEER Research Data Agreement. SEER data involves no more than a minimal risk to the participants and meets other standards: data do not include vulnerable populations such as prisoners, children, veterans, or cognitively impaired persons. However, the data include terminally ill patients of lung cancer without their identity.

Data Analysis Plan

RQ1: Is there a significant association between age at diagnosis and the stages of presentation of lung cancer after controlling for demographic factors: gender, race/ethnicity, and geographic region?

H_o1: There is no significant association between age at diagnosis and the stage of presentation of lung cancer after controlling for demographic factors: gender, race/ethnicity, and geographic region.

H_a1: There is a significant association between age at diagnosis and the stage of presentation of lung cancer after controlling for demographic factors: gender, race/ethnicity, and geographic region.

The chi-squared test was used to examine the association between age at diagnosis and stages of presentation of lung cancer after controlling for demographic risk factors.

RQ2: Is there a significant association between the stage of lung cancer and demographic factors (gender, race/ethnicity, and geographic regions) after controlling for age at diagnosis?

H_o2: There is no significant association between the stage of lung cancer and demographic factors after controlling for age at diagnosis.

H_a2: There is a significant association between the stage of lung cancer and demographic factors after controlling for age at diagnosis.

The chi-squared test was used to examine the association between the stages of presentation of lung cancer and demographic risk factors after controlling for age.

RQ3: Is there any significant relationship between the stage of lung cancer, age at diagnosis, and demographic factors?

H_o3: There is no significant relationship between the stage of lung cancer, age at diagnosis, and demographic factors.

H_a3: There is a significant relationship between the stage of lung cancer, age at diagnosis, and demographic factors.

The multinomial logistic regression test was used to examine the association between the stages of presentation of lung cancer, age at diagnosis, and demographic risk factors.

Ethical Consideration

The ethical issue faced with this study was compliance with the Health Insurance Portability and Accountability Act (HIPPA) regulations. SEER data is abstracted from medical records at healthcare facilities, including hospitals, physicians' offices, and

pathology laboratories, and follows the North American Association of Central Cancer Registries (NSSCCR) data standards. SEER data contain information about geographic location at the county level as well as dates of receiving health care services and data on diagnosis. Because of these variables, the data from the SEER program are considered a limited data set by HIPAA requirements. To protect human subjects and the stakeholders, an Internal Review Board (IRB) approval from the Walden IRB was obtained. The Walden IRB's ethics review focuses on the protection of the data stakeholders: anyone who contributed significantly to the creation of the SEER data as well as human participants in the data. This dissertation includes ethical considerations of the IRB approval criteria required for all students to address analysis of data on living persons. Based on these criteria from section (46.111(b) of 45 CFR 46), this dissertation is exempt from the full IRB review for the use of anonymized data.

Threats to Validity

A trial study was performed to evaluate whether using the SEER data would be feasible and to inform the best way to conduct the future full-scale project. All the available lung cancer cases are stratified by age groups of the patients. The four main components in this study included (1) process—whether the eligibility criteria would be feasible; (2) resources—how much time the main study would take to complete; (3) management—whether there would be a problem with available data; and (4) all the data needed for the main study to test before data analysis. The trial study helped clarify the barriers to and challenges of identifying the strengths and limitations of a planned larger-scale study and testing the design, methodology, and feasibility of the main study. The

threats to external validity are any factors within a study that reduce the generalizability of the results (Laerd Dissertation, n. d.). The external validity is the extent to which findings can be generalized to the whole population, and it can distort the result of the main study. The main threats to external validity in this dissertation included (1) biases with all available lung cancer cases, (2) constructs, methods, and confounding, and (3) the real world versus the experimental world. Since the goal is for this quantitative study to be generalized to the whole population, using all the available lung cancer cases across populations can be the most significant threat to external validity.

On the other hand, internal validity is the extent to which only age at diagnosis (independent variable) caused the changes in the stage of presentation of lung cancer (dependent variable). This was a potential threat to internal validity. The threats to internal validity included the effects of history, maturation, main testing, mortality, statistical regression, and instrumentation. To eliminate the threats to internal validity, only lung cancer patients who were diagnosed with lung cancer during the years 2010–2015 were included. The data included demographic information such as age at diagnosis, gender, race/ethnicity, geographic regions, and stages of presentation of lung cancer during 2010–2015. The standardized instrument included the measurement of stages of presentation of lung cancer in a subgroup of stages: I, II, III, and IV. All the available lung cancer cases in SEER program were used. All patients diagnosed with lung cancer between 2010–2016 were selected and included in the analysis.

Summary

This research used a cross-sectional study design, which is a type of observational study that analyzes data from a population at a specific time. For the best outcome of this study, only lung cancer was included. The results explained the association between age at diagnosis and the stages of presentation of lung cancer based on the representative population at a specific point in time. This study investigated participants who were usually separated by age in groups, gender, race/ethnicity, and the stages of presentation of lung cancer. Therefore, a cross-sectional study design was useful to determine the burden of disease or the health needs of a population. To increase the accuracy of the result, four tests were performed: normal distribution, chi-squared test, and multinomial analyses on categorical variables. Before data analyses were conducted, the study tested for normal distribution of the data. Then, chi-squared test and multinomial analyses were continued. The chi-squared analysis described one characteristic—age at diagnosis—for each stage of lung cancer.

Chapter 4: Results

Introduction

*RQ*1: Is there a significant association between age at diagnosis and the stages of presentation of lung cancer after controlling for demographic factors?

H_o1: There is no significant association between age at diagnosis and the stage of presentation of lung cancer after controlling for demographic factors.

H_a1: There is a significant association between age at diagnosis and the stage of presentation of lung cancer after controlling for demographic factors.

The chi-squared test was used to examine the association between age at diagnosis and stages of presentation of lung cancer after controlling for demographic risk factors. The results of the association between stage I compared with other stages (II, III, and IV), stage II compared with other stages (II, III, and IV), and IV compared with other stages (II, III, and IV) and age groups at diagnosed were statistically significant - stage I [X^2 (7, N=63,107) = 695.94], stage II [X^2 (7, N=63,107) = 193.35], and stage IV [X^2 (7, N=63,107) = 609.80], at $P < 0.05$ respectively (Table 3). Therefore, the null hypothesis was rejected.

RQ2: Is there a significant association between the stage of lung cancer and demographic factors gender, race/ethnicity, and geographic regions after controlling for age at diagnosis?

H_o2: There is no significant association between the stage of lung cancer and demographic factors after controlling for age at diagnosis.

H_a2: There is a significant association between the stage of lung cancer and demographic factors after controlling for age at diagnosis.

The chi-squared test was used to examine the association between stages of lung cancer and demographic risk factors after controlling for age at diagnosis. The chi-squared test analysis displayed a statistically significant relationship between stages of lung cancer and gender X^2 (3, N=63,107) =38.93; race, X^2 (3, N=63,107) = 113.44; and geographical regions X^2 (3, N=63,107)=20.94, $P < .01$ after controlling for age at diagnosis (Table 4). Therefore, the null hypothesis was rejected.

RQ3: Is there any significant relationship between the stage of lung cancer, age at diagnosis, and demographic factors?

H_o3: There is no significant relationship between the stage of lung cancer, age at diagnosis, and demographic factors.

H_a3: There is a significant relationship between the stage of lung cancer, age at diagnosis, and demographic factors.

The multinomial logistic regression analysis was used to examine the association between stages of lung cancer, age at diagnosis, and demographic risk factors.

The purpose of these research questions was to identify any possible associations between the age of an individual and the stages of presentation of lung cancer. Lung cancer patient records were obtained from the special SEER database from 2000 to 2017. The inclusion criteria were definitive NSCLC and SCLC diagnosis by pathology. The exclusion criteria used for my data collection were as follows: (1) patients who were

younger than age 45 years because there were a small number of people diagnosed with lung cancer were younger than 45 years old, (2) patient without any TNM information, and (3) mortality cases that were not caused by lung cancer. SEER*Stat Version 8.3.6.1 (2019; National Cancer Institute Statistics Branch, Bethesda, MD; www.seer.Cancer.gov/seerstat) was used to identify all patients with lung and bronchus cancer during (2010–2015) based on the International Joint Committee on Cancer (AJCC) 7th edition. The demographics of patients included gender, age at diagnosis, race/ethnicity, and geographic region.

The incidence of lung cancer was extracted from the CS-Derived American Joint Committee on Cancer (AJCC), 7th edition (2010–2015) with the stages cancer, stages of tumor (T), stages of node (N), and metastasis (M). The proportion of lung cancer was classified by 5-year intervals [(45–49), (50–54), (55–59), (60–69), (70–74), (75–79), (80–84)]. The primary endpoint of the study was mortality cases that are attributable to lung cancer, and the cancer cases were confirmed using direct visualization without microscopic confirmation, positive histology, radiography without microscopic confirmation, positive microscopic confirmation, method not specified, and cause-specific death. The secondary endpoint for this study was analyzed by the chi-squared test and Post Hoc test. Quantitative analyses were conducted using the chi-squared test on categorical variables and multinomial logistic regression test. Differences in patients' demographics, cancer characterizations, and cancer outcomes among subgroups are summarized in Table 3. The study was approved by institutional ethics committee of Walden University (No. 06–29–20–0563966).

Statistical Analysis:

All statically analyses were performed using the SEER*Stat and the IBM Statistical Package for Social Science (SPSS) Statistics (SPSS, Inc, Chicago, IL, USA) software, version 25. The frequency analysis was used to describe the sample under study, and statistics data were expressed as chi-squared tests to find the relationship between the age at diagnosis and the stages of lung cancer. After testing the normal distribution of the data set, baseline characteristics (demographic and clinical staging data of each patient) were analyzed using the (X^2) test. The association between age at diagnosis and gender, race/ethnicity, stages of lung cancer, and geographic region were individually evaluated by (X^2) test (Tables 5). All risk factors were tested with X^2, OR, 95% CI test, and $P < 0.05$. The four main components in this study included (1) process—whether the eligibility of criteria were feasible; (2) resources—how much time the main study would take to complete; (3) management—whether there would be a problem with available data; and (4) all the data needed for the main study. The external threat included extremely large number lung cancer cases during the years 2010–2015. The stages of presentation of lung cancer were normally distributed, and a P-value of ≤ 0.05 was considered statistically significant. To reduce the sample size, lung cancer cases from 2010–2015 were used for the main study. All statistical analyses and the bar graphs of disease specific mortality (DSM) were performed using SPSS 25.0.

Results

Descriptive Statistic Results

In this quantitative, cross-sectional study, a total of 63,107 cases (N=63,107) of lung cancer from 2010–2015 met the eligibility criteria. The distribution of age groups, stages of presentation of lung cancer, gender, race/ethnicity, and geographical regions are summarized in Table 2.

Inferential Analyses Results

The inferential analyses used chi-squared, odd ratio, and multinomial analysis. Research question 1 investigated the association between stages and age groups at diagnosis after controlling for demographic factors. The chi-squared test of research question 1 showed a statistically significant association between the stages of presentation of lung cancer and the age at diagnosis: stage I, [X^2 (7, N=63,107) = 696.94], stage II [X^2 (7, N=63,107) = 191.35], and stage IV [X^2 (7, 63,107) = 608.80], at $P < 0.05$ respectively (Table 3). The chi-squared test of research question 2 showed a statistically significant relationship between stages of lung cancer and demographic risk factors after controlling for age at diagnosis—gender: X^2 (3, N=63,107) =38.93; race: X^2 (9, N=63,107) = 113.44, and geographical regions: X^2 (3, N=63,107)=20.94, $P < .01$ respectively (Table 4). The results of multinomial analyses displayed the risk of people diagnosed with stage I lung cancer in age group (45–49) years old was 75.4% lower than stage IV, stage II lung cancer in age group (45–49) years old was 66.8% lower than stage IV, and stage III lung cancer in age group (45–49) was 26.4% lower than stage IV, $P < 0.05$ (Table 5).

Table 2

Descriptive Statistics of Sex, Race, Age Groups, and Stages of Presentation of Lung Cancer

	Frequency	**Percent**
Sex		
Women	28,035	44.4
Men	35,075	55.6
Total	63,107	100.0
Race/Ethnicity		
White	55,049	87.2
Black	8,058	12.8
Total	63,107	100.0
Age		
45–49	1,536	2.4
50–54	3,831	6.1
55–59	6,550	10.4
60–64	8,967	14.2
65–70	11,548	18.3
70–74	11,942	18.9
75–79	10,734	17.0
80–84	7,999	12.7
Total	63,107	100.0
Geographical Regions		
Metropolitan counties	5,2835	83.7
Nonmetropolitan counties	10,272	16.3
Total	63,107	100.0
Stage		
Stage I	8,319	13.2
Stage II	5,943	9.4
Stage III	15,441	24.5
Stage IV	33,404	52.9
Total	63,107	100.0

Note: SEER–18 Registries Data.

RQ1: Is there a significant association between age at diagnosis and the stages of presentation of lung cancer after controlling for demographic factors gender, race/ethnicity, and geographic region?

H_o1: There is no significant association between age at diagnosis and the stage of presentation of lung cancer after controlling for demographic factors gender, race/ethnicity, and geographic region.

H_a1: There is a significant association between age at diagnosis and the stage of presentation of lung cancer after controlling for demographic factors: gender, race/ethnicity, and geographic region.

A chi-squared test of independence was performed to examine the relationship between stags of lung cancer [stage I (13.2%), stage II (9.4%), stage III (24.5%), and stage IV(52.9%) and age groups [(45–49), (50–54), (55–59), (60–64), (65–69), (70–74), (75–79), and (80–84)] at diagnosis. The results were statically significant for stage I (X^2 (7, N=63,107) = 695.94), stage II (X^2 (7, N=63,107) = 191.35), and stage IV (X^2 (7, N=63,107) = 609.80), at $P < 0.01$ respectively. However, the result for stage III (X^2 (7, N=63,107) = 695.94) was not statistically significant (Table 2). Therefore, the null hypothesis was rejected for stage I, II, and IV, and the alternative hypothesis was accepted. However, compared with other stages (stage I, II, and IV), the result of stage III and age groups was not statistically significant at X^2 (7, N=63,107) = 11.84 $P > .05$ (Table 3, Figure 1).

Table 3

Chi-squared Tests Results of the Association Between Age at Diagnosis and Stages of Presentation of Lung Cancer

	Age Groups											
Stage	45─49	50─54	55─59	60─64	65─69	70─74	75─79	80─84	Total	X^2	*DF*	*P*-value
Stage I	88	275	512	926	1,560	1,771	1,790	1,397	8,319			
Other	1,448	3,556	6,035	8,041	9,988	10,171	8,944	6,602	54,788	695.94	7	0.00*
Total	1,536	3,831	6,550	8,967	11,548	11,942	10,734	7,999	63,107			
Stage II	94	251	475	734	1,075	1,201	1,192	921	5,943			
Other	1,442	3,580	6,075	8,233	10,473	10,741	9,542	7,078	57,164	191.35	7	0.00*
Total	1,536	3,831	6,550	8,967	11,548	11,942	10,734	7,999	63,107			
Stage III	357	977	1,650	2,182	2,822	2,941	2,649	1,863	15,441			
Other	1,179	2,854	4,900	6,785	8,726	9,001	8,085	6,136	47,666	11.84	7	0.11
Total	1,536	3,831	6,550	8,967	11,548	11,942	10,734	7,999	63,107			
Stage IV	997	2,328	3,913	5,125	6,091	6,029	5,103	3,818	33,404			
Other	539	1,503	2,637	3,842	5,457	5,913	5,631	4,181	2,9703	609.80	7	0.00*
Total	1,536	3,831	6,550	8,967	11,548	11,942	10,734	7,999	63,107			

Note: SEER─18 Registries Data. X^2 = chi-squared test, * indicates $P < 0.05$.

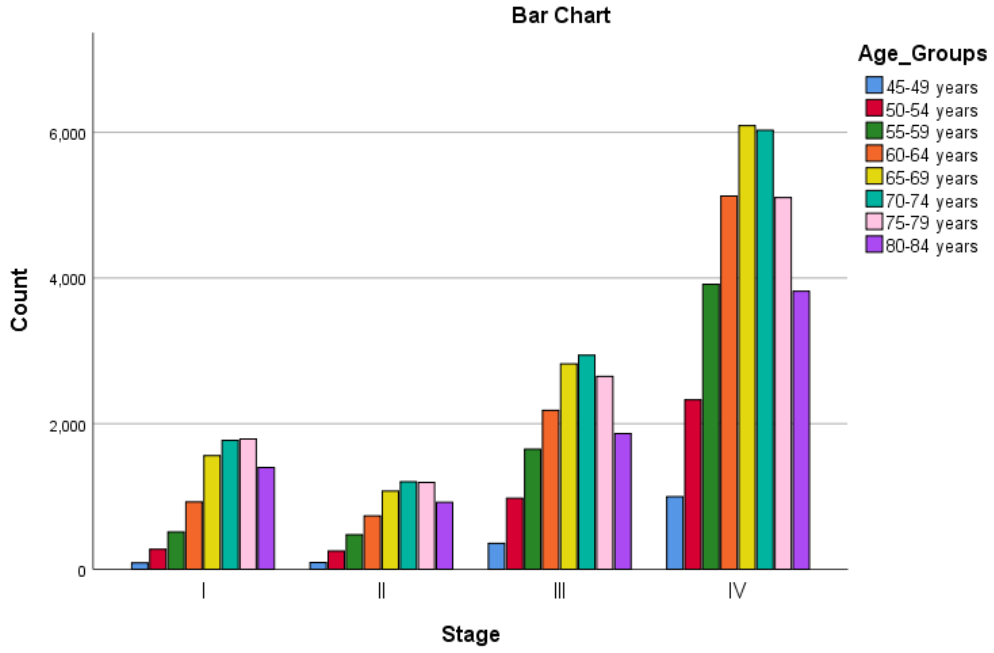

Figure 1. The Association Between Stages of Lung Cancer and Age groups.

Research Question #2: Is there a significant association between the stage of lung cancer and demographic factors after controlling for age?

H_o2: There is no significant association between the stage of lung cancer and demographic factors after controlling for age.

H_a2: There is a significant association between the stage of lung cancer and demographic factors after controlling for age.

A chi-squared test of independence was performed to examine the association between the stage of lung cancer and respective demographic factors using SPSS. The results were statically significant for sex: X^2 (3, N=63,107) = 38.93; race: X^2 (3, N=63,107) = 113.44; and geographical regions: X^2 (3, N=63,107) = 20.94, at $P < 0.01$

respectively (Table 4, Figures 2, 3, & 4) after controlling for age. Therefore, the null hypothesis was rejected, and the alternative hypothesis was accepted.

Table 4

Chi-squared Test Results of Stages of Presentation of Lung Cancer and Demographic Risk Factors

	Stage I	Stage II	Stage III	Stage IV	Total	X^2	DF	*P*-value
Sex								
Women	3,957	2,587	6,773	14,718	28,035	38.93	3	.00*
Men	4,362	3,356	8,668	18,686	35,072			
Total	8,319	5,943	15,441	33,404	63,107			
Race								
White	7,509	5,277	13,468	28,795	55,049			
Black	810	666	1,973	4,609	8,058	113.44	3	.00*
Total	8,319	5,943	15,441	33,404	63,107			
Geographical Regions								
Metropolitan Counties	6,922	4,875	12,900	28,138	52,835			
Non-Metropolitan Counties	1,397	1,068	2,541	5,266	10,272	20.94	3	.00*
Total	8,319	5,943	15,441	33,404	63,107			

Note: Source: SEER−18 Registries Data. X^2 = Chi-square, DF = degree of freedom, * indicates $P < 0.01$.

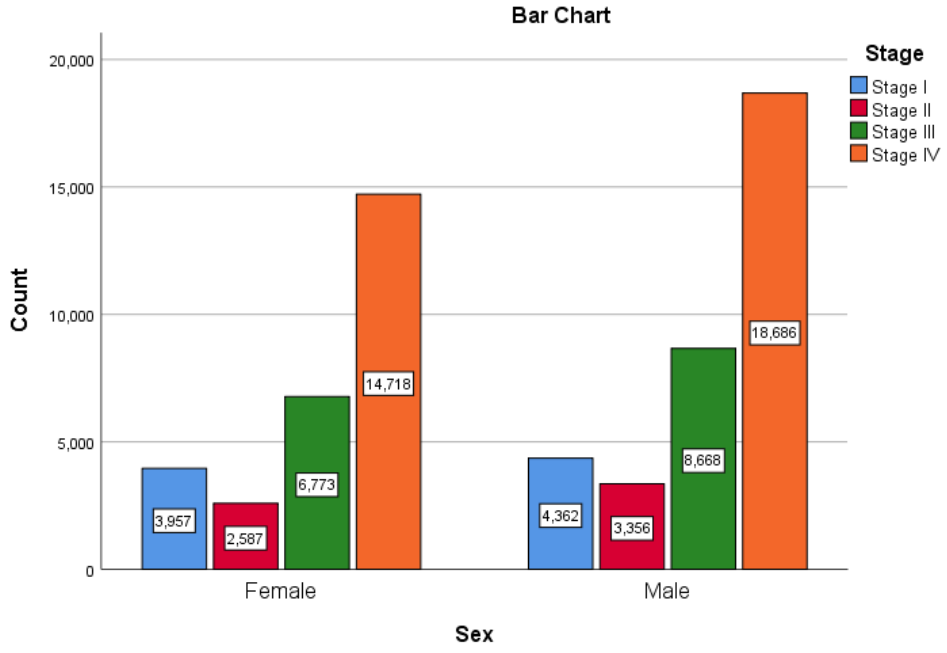

Figure 2. The Association Between Gender and Stages of Lung Cancer

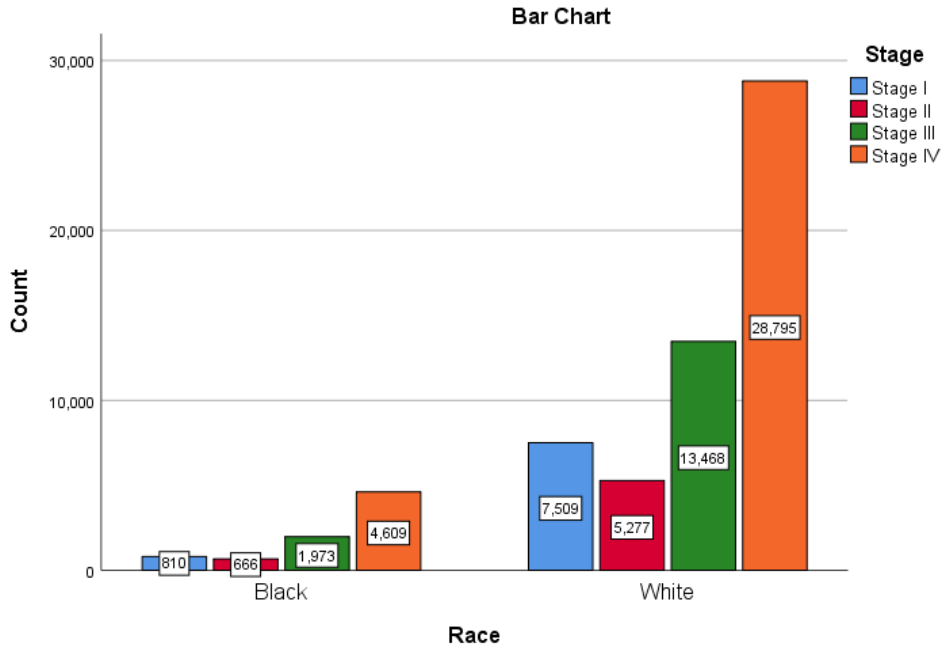

Figure 3. The Association Between Race and Stages of Lung Cancer

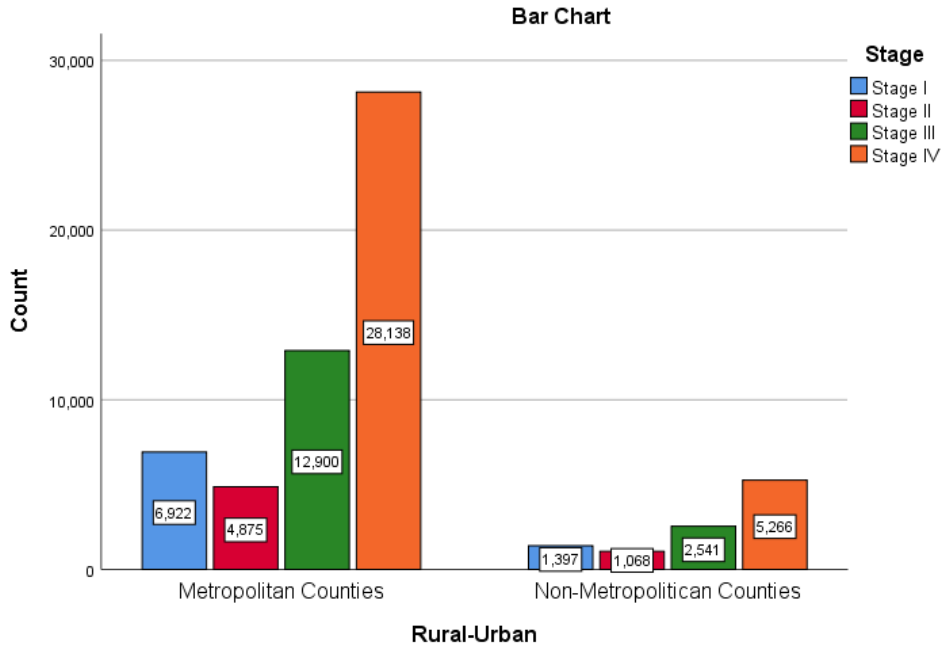

Figure 4. The Association Between Geographic Regions and Stages of Lung Cancer

Research Question #3: Is there any significant relationship between the stage of lung cancer, age at diagnosis, and demographic factors?

H_o3: There is no significant relationship between the stage of lung cancer, age at diagnosis, and demographic factors.

H_a3: There is a significant relationship between the stage of lung cancer, age at diagnosis, and demographic factors.

The multinomial logistic regression analysis was performed to the find the association between stages of lung cancer, age at diagnosis, and demographic risk factors. The risk of women diagnosed with stage I lung cancer was 14.1% [OR:1.141, 95% CI: 1.087–1.198] higher than men diagnosed with stage IV lung cancer in non-metropolitan

counties, $P < 0.05$ (Table 5). However, the risk for women diagnosed with stage II and stage III lung cancer compared with the risk for men diagnosed with stage IV lung cancer in metropolitan areas was not statistically significant at [OR:.975, 95% CI: .922–1.032] and stage III lung cancer is [OR:.991, 95% CI:.954 – 1.030], $P > 0.05$, respectively (Table 5). The risk for African Americans diagnosed with stage I lung cancer was 23.9% [OR:.761, 95% CI:.702–.824], stage II lung cancer was 13.5% [OR:.865, 95% CI:.798–.944], and stage III lung cancer was 6.1% [OR:.939, 95% CI:.887–.995] lower than the risk for White Americans diagnosed with stage IV lung cancer in non-metropolitan areas at $P < .05$. The risk for people diagnosed with stage I, II, and III lung cancer in metropolitan counties (metropolitan areas > 1 million population, counties in metropolitan areas of 250,000 to 1 million population, and metropolitan areas of less than 250,000 population) was 9.8%, 15.9%, and 5.2% lower than people diagnosed with stage IV lung cancer in non-metropolitan counties, respectively (Table 5).

Table 5

Multinomial Regression Analysis of the Association between Stages of Lung Cancer and Demographic Risk Factors

Variable		B	Std. Error	Wald	*Exp*(B)	95% CI LL	95% CI UL	*P*-value
Stage I								
Age Groups	45–49	-1.404	.116	147.400	.246	.196	.308	.00*
	50–54	-1.103	.071	239.710	.332	.289	.382	.00*
	55–59	-1.003	.057	313.883	.367	.328	.410	.00*
	60–64	-.687	.048	208.310	.503	.458	.552	.00*
	65–69	-.346	.042	66.968	.707	.651	.768	.00*
	70–74	-.213	.041	26.571	.808	.745	.876	.00*
	75–79	-.037	.042	.803	.963	.888	1.045	.37
Sex	Women	.132	.025	28.23	1.141	1.087	1.198	.00*
Race	Black	-.273	.041	44.96	.761	.702	.824	.00*
Metropolitan Counties		-.103	.033	9.463	.902	.845	.963	.02*
Stage II								
Age Groups	45–49	-.935	.114	67.109	.393	.314	.491	.00*
	50–54	-.797	.076	109.593	.451	.388	.523	.00*
	55–59	-.683	.061	124.978	.505	.448	.569	.00*
	60–64	-.521	.054	92.809	.594	.534	.660	.00*
	65–69	-.316	.050	40.622	.729	.662	.804	.00*
	70–74	-.194	.048	16.040	.823	.749	.906	.00*
	75–79	-.034	.049	.485	.967	.878	1.064	.49
Sex	Women	-.025	.029	.758	.975	.922	1.032	.39
Race	Black	-.145	.045	10.622	.865	.793	.944	.00*
Metropolitan Counties		-.173	.037	21.57	.841	.782	.905	.00*
Stage III								
Age Groups	45–49	-.306	.068	20.277	.736	.645	.841	.00*
	50–54	-.146	.048	9.355	.864	.787	.949	.00*
	55–59	-.143	.041	12.199	.867	.800	.939	.00*
	60–64	-.135	.038	12.414	.874	.811	.942	.00*
	65–69	-.052	.035	2.029	.950	.884	1.02	.15
	70–74	.000	.036	.000	1.00	.931	1.073	.99
	75–79	.062	.037	2.788	1.064	.989	1.144	.09
Sex	Women	-.009	.020	.205	.991	.954	1.030	.65
Race	Black	-.062	.029	4.601	.939	.887	.995	.03*
Metropolitan Counties		-.053	.027	3.996	.948	.900	.999	.05*

Note: Reference categories = Stage IV, Age (80-84), men, White, nonmetropolitan counties. CI=confidence interval, LL = lower limit, UL = upper limit. *p*-value set at 0.05, * indicates *p*-value < 0.05.

The risk for people diagnosed with stage I lung cancer in age group (45–49) years was 75.4% [OR:.246, 95% CI=.196–.308], in age group (50–54) was 66.8% [OR:.332, 95% CI=.289–.382], in age group (55–59) was 63.3% [OR:.367, 95% CI=.328–.410], in age group (60–64) was 49.7% [OR:.503, 95% CI=.458–.552], in age group (65–69) years old was 29.3% [OR:.707, 95% CI=.651–.768], and in age group (70–74) years old was 19.2% [OR:.808, 95% CI=.745–876] lower than people diagnosed with stage IV lung cancer in age group (80–84) years old, and was statistically significant at $P < 0.01$ respectively. However, the risk for people diagnosed with stage I lung cancer in age group (75–79) was not statistically significant at [OR:963, 95% CI=.888-.1.045], $P = 0.37$ (Table 5).

The risk for people diagnosed with stage II lung cancer in age group (45–49) years old was 60.7 % [OR:.393, 95% CI= .314–.491], age group (50–54) years old was 54.9 % [OR:.451, 95% CI= .388–.523], age group (55–59) years old is 49.5 % [OR:.505, 95% CI= .448–.569], age group (60–64) years old is 40.6 % [OR:.594, 95% CI= .534–.660], and age group (65–69) years old was 27.1% [OR:.729, 95% CI= .662–.804], and age group (70–74) years old was 17.7 % [OR:.823, 95% CI= .74–.906] lower than the risk of age group (80–84) years old diagnosed with stage IV lung cancer and was statistically significant, $P < 0.01$ respectively. However, the risk of people diagnosed with stage II lung cancer in age group (75–79) years old was not statistically significant [OR:.486, 95% CI=.878–1.064], $P = 0.48$ (Table 5).

The risk for people diagnosed with stage III lung cancer in age group (45–49) was 26.4%, [OR:.736, 95% CI:.645–.841], age group of (50–54) was 13.6%, [OR:.864, 95% CI:.787–.949], age group (55–59) was 13.3 % [OR:.867 95% CI= .800–.939], age group (60–64) was 12.6% [OR:.874, 95% CI= .811–.942], lower than people with age group (80–84) years old diagnosed with stage IV lung cancer, $P < .01$ respectively. However, the risk for people diagnosed with stage III in age group (65–69), (70–74), and (75–79) was not statistically significant at [OR:.950, 95% CI=.884–1.020, $P = 0.15$]; [OR:.990, 95% CI=.931–1.073, $P = 0.99$]; [OR: 1.064, 95% CI=.989–1.144, $P = .09$], $P > 0.05$ (Table 5).

The risk for women diagnosed with stage I lung cancer was 14.1%, [OR:.1.141, 95% CI = .1.087–1.198] higher than men with stage IV and the association between women and stage I lung cancer was statistically significant at $P < 0.01$. However, the risk for women diagnosed with stage II and III was not statistically significant [OR:.975, CI= .922–1.032, $P = 0.38$] and [OR:.991, 95% CI=.954–1.030, $P = 0.65$] (Table 5). The risk for African Americans diagnosed with stage I lung cancer was 23.9% [OR:.761, 95% CI= .702–.824], stage II lung cancer was 13.5% [OR:.865, 95% CI= .793-.944, and stage III was 6.1% [OR:.948, 95% CI= .900–.999] lower than White Americans diagnosed with stage IV lung cancer at $P < 0.05$ respectively (Table 5). The risk for people living in Metropolitan counties diagnosed with stage I lung cancer was 9.8% [OR:.902, 95% CI= .945-963], stage II lung cancer was 15.9% [OR:.841, 95% C=: .782–.905], and stage III was 5.2% [OR:.948, 95% CI= .900–.999] lower than stage IV lung cancer in non-

Metropolitan counties and the association between Metropolitan counties and stages of lung cancer was statistically significant at $P < 0.05$ (Table 5).

Summary

The results of this study presented the association between the age groups and the stages of presentation of lung cancer. Under the age distribution associated with stages of lung cancer, the risk of stage IV cancer was significantly higher than stage I, II, and III (Figure 2). Multinomial regression analysis indicated the risk of people diagnosed with stage I lung cancer in age groups [(45–49), (50–54), (55–59), (60–64), (65–69), and (70–75)] years old was statistically significant at [OR: .246, 95% CI= .196–.308], [OR: .332, 95% CI: .289–.382], (OR: 367, 95% CI= .328–.410], [OR:503, 95% CI=.458–.552], [OR: .707, 95% CI= .651–.768], and [OR:.808, 95% CI= .745–.876]; stage II lung cancer in age group [(45–49), (50–54), (55–59), (60–64), (65–69), and (70–74)], was statically significant at [OR: .393, 95% CI= .314–.491], [OR: .451, 95% CI: .388–.523], [OR: .505, 95% CI=.448–.569], [OR:.594 95% CI= .534–.660], [OR: .729, 95% CI= .662–.804], and [OR: .823, 95% CI= .749–.906]; stage III lung cancer in age group (45–49), (50–54), (55–59), and (60–64), [OR: .736, 95% CI= .645–.841], [OR: .864, 95% CI= .787–.949], [OR: .867, 95% CI= .800–.939], and [OR: .874, 95% CI= .811–.942] at $P < 0.05$ respectively (Table 5). However, there were no statistically significant association between stage I lung cancer and age group (75–79) [OR:.370 95% CI:.888.1.045 at $P = 0.37$; stage II lung cancer and age group (75–79) [OR: .841, 95% CI= .781–.905], and

stage III lung cancer in age group (65–69), (70–74), and (75–79) were not statistically significant, $P > 0.05$ (Table 5).

Chapter 5: Discussion, Conclusions, and Recommendations

Introduction

The purpose of this dissertation was to provide an age recommendation for preventive screening to detect early stages of lung cancer. The results of this study indicated that each stage of lung cancer was a dependent risk predictor of lung cancer mortality. The nature of this research was a quantitative study using a cross-sectional design to examine data from age, stage, and demographic factors. Specifically, the differences in stages of lung cancer and age groups were analyzed. This quantitative method included stages of lung cancer analyses for age groups, and the chi-squared results indicated that stages I, III, and IV and age groups [(45–49), (50–54), (55–59), (60–64), (65–69), (70–74), and (75–79)] were statistically significant: stage I, X^2 (7, N=63,107) = 695.94, stage II, X^2 (7, N=63,107) = 191.35, and stage IV, X^2 (7, 63,107) = 609.80, at $P < 0.05$ respectively. However, stage III and all age groups (45-84) were not statistically significant, $P = 11$ (Table 3). The results of multinomial analyses indicated low rates of stage I and stage II lung cancer in age group (45–49) years old. Since lung cancer is asymptomatic and screening is not recommended for age groups (45–49) and (50–55), the actual rate of early stage lung cancer may not be accurately ascertained. Therefore, the results of our data analyses support updating the recommended age for annual screening.

Interpretation of the Findings

This study delineated the distinct metastatic features of lung cancer in patients with different age groups, race groups, sex, and geographical regions (Adams et al., 2015). As lung cancer is a malignant lung carcinogenesis characterized by cell mutations, the findings based on mortality rate were consistent with previous population-based studies on ages and different genders and races (Adams et al., 2015; Dorak & Karpuzoglu, 2012; Wang et al., 2015). Historically, social inequalities in the United States have caused African American populations to be more vulnerable to cancers and diseases (David et al., 2016; Toporowski et al., 2012). However, this study found that White Americans (87.2%) were more vulnerable to lung cancer than the African American population (12.8%). This could be due to inequality in health care, and more White Americans may have access to health insurance and were screened for early detection of lung cancer in this SEER–18 registry population (Pickett & Wilkinson, 2015). In this study, patients between ages (45–84) and with stage (IV) lung cancer were more likely to be White than African American. The racial differences observed were likely to be a result of a complex relationship between screening access and etiological factors (age groups) across different racial and ethnic populations. Relevant information was included to explain the different stages of lung cancer across age groups and to support future research studies that focus on biomarkers of lung cancer based on age.

Limitations of the Study

This study had some limitations; for example, surgery and treatment might contribute to differences in stages of lung cancer mortality. Age groups 0–44 and > 84

were excluded to decrease the sample sizes from both ends using lung cancer cases in SEER registries. Differences in lung-cancer specific mortality were identified in different age groups. Future cross-sectional studies focusing on biomarkers are needed to evaluate determinants of outcome.

Recommendations

As age and mutations increase the risk of lung cancer, earlier annual preventive screening is needed to detect earlier stages of lung cancer. Currently, the American Cancer Society recommends yearly lung cancer screening with LDCT for high-risk populations, those who are smokers, and those who have a genetic predisposition. For people at average risk for lung cancer, the American Cancer Society recommends starting regular screening at age 55. This age recommendation can be lowered to reduce the number of new cases of lung cancer and mortality, since cancer can take up to 20 years before it appears in the chest X-ray and LDCT. Since the purpose of cancer screening is to find cancer in people who are asymptomatic, early detection through screening based on age gives the best chance of finding cancer as early as possible.

Summary

The aim of this dissertation was to conduct a population study comprising 63,107 subjects from SEER–18 data to provide an age recommendation for annual preventive screening to implement social change. This dissertation provides information about how age-related factors can contribute to lung cancer and supports a policy recommendation for annual preventive screening to detect early stages of lung cancer for vulnerable populations: everyone over 45 years old, and both smokers and non-smokers. This

research will bridge a gap in the understanding of the association between age-related factors and lung cancer development. This project is unique because it addresses how a simple age variable, which is a unit number of times, can significantly affect cancer prevention. In order to identify a set of age groups, a combination of parameters with appropriate weighting would measure patient age to support current screening methods or future biomarker screening to provide information about age-related factors that contribute to lung cancer. Therefore, this paper included information such as how long it takes for cancer development after exposure to carcinogens that cause mutations. The information provided in this student dissertation will benefit future studies on preventive screening recommendations, help health professionals enhance their understanding of age factors in cancer development, and may help reduce the gap in the lack of effectiveness in preventive screening for early detection.

Implications

The results of this study have substantial implications for social change and suggest age-based recommendation for preventive lung cancer screening for early detection of lung cancer. The results also add more support for the establishment of a comprehensive policy on preventive screening for early detection of lung cancer that aides in alleviating cancer among vulnerable population. Assessing age-associated lung cancer will not only fulfill the goal of providing information to support a recommendation for early diagnosis and treatment of lung cancer, but may help reduce the number of people who die from lung cancer, reduce the burdens of health care costs, and provide better health outcomes. Although current methods for early diagnosis and

treatment reduce the rate of morbidity and mortality caused by lung cancer, lung cancer is still the leading cause of cancer death. This paper concluded by giving information about age stratification to help understand the age-related factors that contribute to lung cancer. The enhancement in knowledge of age-factors related to lung cancer could provide information about the association between age and biomarkers of lung cancer for future molecular biological studies. The statistical analyses in this dissertation indicated that among all age groups, the stage of lung cancer with the fastest progression was stage IV. Because many studies have found that the incidence of cancer rates increase with age (Chen et al., 2019; White et al., 2014), studies that evaluate a combination of factors provide a better tool for measuring age-related factors that contribute to lung cancer development based on the stages of lung cancer. Future studies are needed to focus on biomarkers of lung cancer based on patient age to better understand how age can contribute to lung cancer, and to provide supporting information for age-based recommendation for early detection of lung cancer.

Conclusion

The age at diagnosis and each stage of lung cancer was shown to be statistically significant when chi-squared tests were used. The data also indicated low rates in early stages (stage I and II) of lung cancer in age groups (45–49) and (50–54) years old. However, since the early stage of lung cancer is often asymptomatic and screening is not currently recommended for these age groups, the actual rate of early stage lung cancer may not be able to be determined. Therefore, further research is needed to determine

whether there is a significant difference between the current data and the actual rates of early stage lung cancer.

References

Adams, P. D., Jasper, H., & Rudolph, K. L. (2015). Aging-inducted stem cell mutations as drivers for disease and cancer. *Cell Stem Cell, 16*(6), 601–612. https://doi.org/10.1016/j.stem.2015.05.002

American Cancer Society. (2019). *Lung cancer.* Retrieved from https://www.cancer.org/cancer/lung-cancer/about/what-is.html

American Federation for Aging Research [AFAR], n.d. Retrieved from https://www.afar.org/

American Joint Committee on Cancer [AJCC]. (2010). JCC Cancer staging manual, 7th Edition. Springer-Verlag, New York, N.Y. Retrieved from https://cancerstaging.org/references-tools/deskreferences/Documents/AJCC%207th%20Ed%20Cancer%20Staging%20Manual.pdf

American Lung Association [ALA]. (2020). State of lung cancer, state data. Retrieved from https://www.lung.org/our-initiatives/research/monitoring-trends-in-lung-disease/state-of-lung-cancer/states/

Annangi, S., Nutalapati, S., Foreman, M. G., Pillai, R., & Flenaugh, E. L. (2019). Potential racial disparities using current lung cancer screening. *Journal of Racial and Ethnic Health Disparities, 6*(1), 22–26. https://doi.org/10.1007/s40615-018-0492-z

Atwater, T. & Massion, P. P. (2016). Biomarkers of risk to develop lung cancer in the new screening era. *Annual Translational Medicine, 4*(8), 1–6.

https://doi.org/10.21037/atm.2016.03.46

Bakulski, K. M., Dou, J., Lin, N., & London, S. J. (2019). DNA methylation signature of smoking in lung cancer is enriched for exposure signatures in newborn and adult blood. *Scientific Reports*, *9*(4576), 1–13. https://doi.org/10.1038/s41598-019-40963-2

Bossé, Y., & Amos, C. (2019). A decade of GWAS results in lung cancer. *Cancer Epidemiology, Biomarkers, Prevention, 27*(4), 363–379. https://doi.org/10.1158/1055-9965.EPI-16-0794

Bürkle, A., Moreno-Villanueva, M., Bernhard, J., Blasco, M., Zondag, G., Hoeijmakers, H. J., … Aspinall, J. (2015). Mark-age biomarkers of aging. *Mechanisms of Aging and Development, 151*, 2-12. http://doi.org/10.1016/j.mad,2015.03.006.

Centers for Disease Control and Prevention [CDC]. (n. d.). *Smoking and tobacco use: Fast facts and fact sheets.* https://www.cdc.gov/tobacco/data_statistics/fact_sheets/index.htm

Centers for Medicare and Medicaid Services (2019). *National Health Expenditure: Projected.* Retrieved from https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsProjected.html

Chawinska, E., Tukiendorf, A., & Miszczyk, L. (2014). Interrelation between population density and cancer incidence in the province of Opole, Poland. *Contemporary Oncology, 18*(5), 367–370. https://doi.org/10.5114/wo.2014.44122

Chen, T., Zhou, F., Jiang, W., Mao, R., Zheng, H., Qin, L., & Chen, C. (2016). Age at

diagnosis is a heterogeneous factor for non-small cell lung cancer patients. *Journal of Thoracic Disease, 11*(6), 2251–2266. https://doi.org/10.21037/jtd.2019.06.24

Choudhuri, S., Chanderbhan, R., & Mattia, A. (2018). *Chapter 20 – Carcinogenesis: Mechanisms and models in veterinary toxicology*. In Gupta R, C. (Ed). Academic Press (Third Edition). Cambridge, Massachusetts, United States, [E-reader version], 339–354. https://doi.org/10.1016/B978-0-12-811410-0.00020-9

Chu, G.C.W., Lazare, K., & Sullivan, F. (2018). Serum and bloodbased biomarkers for lung cancer screening: A systematic review. *BioMed Central, 18*(181), 1–6. https://doi.org/10.1186/s12885-018-4024-3

Coe, R. (2002). It's the effect size, stupid: What effect size is and why it is important. Retrieved from https://www.leeds.ac.uk/educol/documents/00002182.htm

Crapo, J. D., Barry, B. E. Gehr, P., Bachofen, M., & Weibel, E. R. (1982). Cell number and cell characteristics of the normal human lung. *American Review of Respiratory Disease, 126*(2), 332–337. https://doi.org/10.1164/arrd.1982.126.2.332

David, S. P., Wang, A., Kapphahn, K., Hedlin, H., Desai, M., Henderson, M,…& Stefanick, M. L. (2016). Gene by environment investigation of incident lung cancer risk in African Americans. *EbioMedicine, 4*, 153–161. https://doi.org/10.1016/j.ebiom.2016.01.002

de Groot, P. M., Wu, C. C., Carter, B. W., & Munden, R., F. (2018). The epidemiology of lung cancer. *Translational Lung Cancer Research, 7*(3), 220–233.

https://doi.org/10.21037/tlcr.2018.05.06

Dorak, M. T., & Karpuzoglu, E. (2012). Gender differences in cancer susceptibility: An inadequately addressed issue. *Frontiers in Genetics, 3*(268), 1–11. https://doi.org/10.3389/fgene.2012.00268

Eggert, J. A., Palavanzadeh, M., & Blanton, A. (2017). Screening and early detection of lung cancer. *Seminars on oncology, 33*(2), 29–140. https://doi.org/10.1016/j.soncn.2017.03.001

Enge., M., Arda, H.E., Mignardi, M., Beausang, J., Bottino, R., Kim, S.K., & Quake, S.R. (2017). Single-cell analysis of human pancreas reveals transcriptional signatures of aging and somatic mutation patterns. Cell *171*, 321–330.e314. https://doi.org/10.1016/j.cell.2017.09.004

Fenizia, F., Pasquale, R., Roma, C., Bergantino, F., Iannaccone, A., & Normanno, N. (2018). Measuring tumor mutation burden in non-small cell lung cancer: tissue versus liquid biopsy. *Traditional Lung Cancer Research, 7*(6), 668–677. https://doi.org/10.21037/tlcr.2018.09.23

Fos, P. J. (2011). *Epidemiology foundations: the science of public health*. Jossey-Bass. San Francisco, CA.

Gingras, M. (2018). Scholar-Practitioner final project. PUBH-8540: Epidemiology Topic Seminar A00563966 Biomarkers Screening for Detection of Lung and Bronchus Cancer: a systematic review Abstract.

Griffiths, A. J., F., Miller, J. H., & Suzuki, D. T. (2000). *An introduction to genetic analysis. (7th ed.).* New York, NY., Freeman.

Gyoba, J., Shan, S., Wilson, R., & Bédard, E.L. R. (2016). Diagnosing lung cancers through examination of Micro-RNA biomarkers in blood, plasma, serum and sputum: A review and summary of current literature. *International Journal of Molecular Sciences, 17*(494), 1–14. https://doi.org/10.3390/ijms17040494

Hammond, S. M. (2015). An overview of microRNAs. *Advanced Drug Delivery Reviews, 7*, 3–14. https://doi.org/10.1016/j.addr.2015.05.001

Hayashi, M. T. (2017). Telomere biology in aging and cancer: early history and perspective. *Genes Genetic System*, *92*(3), 107–118. https://doi.org/10.1266/ggs.17-00010

Huang, J. Y., Larose, T. L., Luu, H. N., Wang, R., Fanidi, A., Alcala, K…& Yuan, J.-M. (2019). Circulating markers of cellular immune activation in prediagnostic blood sample and lung cancer risk in the lung cancer cohort consortium (LC3). *International Journal of Cancer*, *00*(00–00), 1–12. https://doi.org/10.1002/ijc.32555

Healthy People 2020. (2019). *Older adults.* Retrieved from https://www.healthypeople.gov/2020/topics-objectives/topic/older-adults

Honda, O., Johkoh, T., Sekiguchi, J., Tomiyama, N., Mihara, N., Sumikawa, H…, & Nakamura, H. (2009). Doubling time of lung cancer determined using three-dimensional volumetric software: comparison of squamous cell carcinoma and adenocarcinoma. *Lung Cancer*. *66*(2), 211–217. https://doi.org/10.1016/j.lungcan.2009.01.018

Izzotti, A., Balansky, R., Ganchev, G., Iltchevam, M., Longobardi, M., Pulliero, A., …

Flora, S. D. (2016). Blood and lung microRNAs as biomarkers of pulmonary tumorigenesis in cigarette smoke-exposed mice. *Oncotarget, 7*(51), 84758–84774. https://doi:10.18632/oncotarget.12475

Jia, Y., Zang, A., Feng, Y., Li, X.-F., Zhang, K., Li, H., Wang, R., Wei, Y., & Huo, R. (2016). miRNA-486 and miRNA-499 in human plasma evaluate the clinical stages of lung cancer and play a role as a tumor suppressor in lung tumorigenesis not pathogenesis. *Bangladesh Journal Pharmacology*, *11*(1), 264–268. https://doi.org/10.3329-bjp.v11i1.25318

Jin, X., Liu, X., Zhang, Z., Guan, Y., XV, R., & Li., J. (2018). Identification of key pathways and genes in lung carcinogenesis. *Oncology Letters, 16*(2018), 4185–4192. https://doi.org/10.3892/01.2018.9203

John, U., Hanke, M., Meyer, C., & Schumann, A. (Kanashiki, M., Tomizawa, T., Yamaguichi, I., Kurishima, K., Hizawa, N., Ishikawa, H.,… & Satoh, H. (2012). Volume doubling time of lung cancers detected in a chest radiograph mass screening program: comparison with CT screening. *Oncology letters*, *4*(1), 513–516. https://doi.org/10.3892/ol.2012.780

Krol, J., Loedige, I., & Fillipowicz, W. (2010). The widespread regulation of microRNA biogenesis, function and decay. *Genetics*, *11*, 597–610. https://doi.org/10.1038/nrg2843

Lee, B., Lee, T., Lee, S-H., Choi, Y. L., & Han, J. (2016). Clinicopathologic characteristics of EGFR, KRAS, and ALK alterations in 6595 lung cancers. *Oncotarget*, *7*(17), 23874–23884. https://doi.org/10.18632/oncotarget.8074

Li, C., Yin, Y., Liu, X., Xi, X., Xue, W., & Qu, Y. (2017). Non-small cell lung cancer associated microRNA expression signature: Integrated bioinformatics analysis, validation and clinical significance. *Oncotarget, 8*(15), 24564–24578. https://doi.org/10.18632/oncotarget.15596.

Li, Y., Gu, F., Zhu, Q., Ge, D., & Lu, C. (2018). Transcriptomic and functional network features of lung squamous cell carcinoma through integrative analysis of GEO and TCGA data. *Scientific Reports, 8*:15834. https://doi.org/10.1038/s41598-018-3460-w

Liang, J., Lv, J., & Liu, Z. (2015). Identification of stage-specific biomarkers in lung adenocarcinoma based on RNA-seq data. *Tumor Biology, 36*(8), 6391–6399. https://doi.org/10.1007/s13277-015-3327-0

Liu, M., Zhou, K., & Cao, Y. (2016). MicroRNA-944 affects cell growth by targeting EPHA7 in non-small cell lung cancer. *International Journal of Molecular Sciences, 17*(1493), 1–12. https://doi.org/10.3390.ijms17101493

Liu, X., Chen, J., Guan, T., Yao, H., Zhang, W., Guan, Z., & Wang, Y. (2019). miRNAs and target genes in the blood as biomarkers for the early diagnosis of Parkinson's disease. *BMC System Biology*, *13*(10), 1–8. https://doi.org/10.1186/s12918-019-0680-4

Lozano, M., Echeveste, J. I., Abengozar, M., Meijias, L. D., Idoate, M. A., Calvo, A…& Andrea, C. E. (2018). Cytology smears in the era of molecular biomarkers in non-small cell lung cancer – Doing more with less. *Molecular testing on cytology smears, 142*(2018), 291–298). https://doi.org/ 10.5858/arpa.2017-0208-RA.

Mayo Clinic (2019). Lung cancer. https://www.mayoclinic.org/diseases-conditions/lung-cancer/symptoms-causes/syc-20374620

McClelland III, S., Page, B. R., Jaboin, J. J., Chapman, C. H., Deville Jr, C., & Thomas, C. R. (2017). The pervasive crisis of diminishing radiation therapy access for vulnerable populations in the United States, part 1: African-American patients. *Adv Rdiat Oncology, 2*(4), 523–531. https://doi.org/10.1016/j.adro.2017.07.002

Miller, Y. E. (2005). Pathogenesis of lung cancer 100 year report. *American Journal of Respiratory Cell and Molecular Biology, 33*(3), 216–223. https://doi.org/10.1165/rcmb.2005-0158OE

Mitsuhashi, A., Goto, H., Kuramoto, T., Tabata, S., Yukishige, S., Abe, S., Hanibuchi, M…, & Nishioka, Y. (2013). Surfactant protein A suppresses lung cancer progression by regulating the polarization of tumor-associated macrophages. *American Journal of Pathology, 182*(5), 1843–1853. https://doi.org/10.1016/j.ajpath.2013.01.030

Moyer, V. A. (2014). Screening for lung cancer: U.S. Preventive Services Task Force Recommendation Statement. *Annals of Internal Medicine,160*(5), 330–338. https://doi.org/10.7326/M13-2771

National Cancer Institute [NCI]. (n.d. a). *Process of Cancer Data Collection.* https://training.seer.cancer.gov/registration/data/collection.html

National Toxicology Program [NTP], (2016). Tobacco-Related Exposures. In: *Report on Carcinogens. Fourteenth Edition.* U.S. Department of Health and Human

Services, Public Health Service, National Toxicology Program, 2016. https://ntp.niehs.nih.gov/pubhealth/roc/index-1.html

Ni, M., Liu, X., Wu, J., Zhang, D., Tian, J., Wang, T., & Zhang, X. (2018). Identification of candidate biomarkers correlated with the pathogenesis and prognosis of non-small cell lung cancer via integrated bioinformatics analysis. *Frontiers in Genetics, 9*(469), 1–14. https://doi.org/10.3389/fgene.2018.00469

Patnaik, S. K., Kannisto, E. D., Mallick, R., & Vachani, A. (2017). Whole blood microRNA expression may not be useful for screening non-small cell lung cancer. *PLoS ONE, 12*(7), e0181926. https://doi.org/10.1371/journal.pone.0181926

Pickett, K. E., & Wilkinson, R. G. (2015). Income inequity and health: A causal review. *Social Science & Medicine*. *128*(2015), 316–326. https://doi.org/10.1016/j.socscimed.2014.12.031

Pepe, M. S., Li, C. I., & Feng, Z. (2015). Improving the quality of biomarker discovery research: the right samples and enough of them. *Cancer Epidemiology Biomarkers and Prevention, 24*(6), 944–950. https://doi.org/10.1158/1055-9965

Pu, H.-Y., Xu, R., Zhang, M.-Y., Yuan, L.-J., Hu, J.-Y., Huang, G.-L., & Wang, H.-Y. (2017). Identification of microRNA-615-3p as a novel tumor suppressor in non-small cell lung cancer. *Oncology, 13*, 2403–2410. https://doi.org/10.3892/ol.2017.5684

Pyenson, B., & Dieguez, G. (2016). 2016 reflections on the favorable cost-benefit of lung cancer screening. *Annuals of Translational Medicine*,*4*(8), 1–8. https://doi.org/10.21037/atm.2016.04.02

Quail, D. F., & Joyce, J., A. (2013). Micro environmental regulation of tumor progression and metastasis. *National Medical*, *19*(11), 1423–1437. https://doi.org/10.1038/nm.3394

Roberti, A., Valdes, A. F., Torrecillas, R. Fraga, M. F., & Fernandez, A. F. (2019). Epigenetics in cancer therapy and nanomedicine. *Clinical Epigenetics*, *11*(81), 1–18. https://doi.org/10.1186/s13148-019-0675-4

Robbins, H.A., Engels, E. A., Pfeiffer, R. M., & Shiels, M. (2015). Age at cancer diagnosis for blacks compared with whites in the United States. *Journal of National Cancer Institute*, *107*(3), 1–8. https://doi.org/10.1093/jnci/dju489

Ryan, B. M. (2018). Lung cancer health disparities. *Carcinogenesis*, *39*(6), 741–751. https://doi.org/10.1093/carcin/bgy047

Salamanca, J. C., Meehan-Atrash, J., Vreeke, S., Escobedo, J. O., Peyton, D. H., & Strongin, R. M. (2018). E-cigarettes can emit formaldehyde at high levels under conditions that have been reported to be non-averse to users. *Scientific Reports*, *8*(7559), p.1–6. https://doi.org/10.1038/s41598-018-25907-6

Schueller, G., & Herold, C. J. (2003). Lung metastases. *Cancer imaging*, *3*(2), 126–128. https://doi.org/10.1102/1470-7330.2003.0010

Siegel, R. L., & Miller, K. D. (2019). Cancer statistics, 2019. *CA: A Cancer Journal for Clinicians*, *69*(1), 7–34. https://doi.org/10.3322/caac.21551

Shao, Y., Liang, B., Long, F., & Jiang, S.-J. (2017). Diagnostic microRNA biomarker discovery for non-small lung adenocarcinoma by integrative bioinformatics analysis. *BioMed Research International*, *2017*(2563085), 1–9.

https://doi.org/10.1155/2017/2563085

Shen, J., Todd, N. W., Zhang, H., Yu, L., Lingxiao, X., Mei, Y., Guarnera, M., Liao, J., Chous, A., & Lu, C.L. (2011). Plasma microRNAs as potential biomarkers for non-small-cell lung cancer. *Lab Investigation*, *91*, 579–587. https://doi.org/10.1038/labinvest.2010.194

Siroglavić, K.-J., Vižintin, M. P., Tripković, I., Šekerija, M., & Kukulj, S. (2017). Trends in incidence of lung cancer in Croatia from 2001 to 2013: gender and regional differences. *Croatian Medical Journal, 58*(5), 358–636. https://doi.org/10.3325/cmj.2017.58.358

Soo, R. A., Kubo, A., Ando, M., Kawaguchi, T., Ahn, M.-J., & Ou, S.-J. I. (2017). *Clinical Lung Cancer, 18*(5), 535–542. https://doi.org/10.2016/j.cllc.2017.01.005

Sozzi, Boeri, Rossi, Verri, Suatoni, Bravi, …& Pastorino (2014). Clinical utility of a plasma microRNA biomarker within lung cancer screening. *Journal of Clinical Oncology*, *32*(14), 768–773. https://doi.org/10.1200/JCO.2014.56.7610

Stram, D. O., Park, S. L., Haiman, C. A., Murphy, S. E., Patel, Y., Hecht, S. S., & Marchand, L. L. (2019). Racial/ethnic differences in lung cancer incidence in the multiethnic cohort study: an update. *Journal of National Cancer Institute, 111*(8), 1–9. https://doi.org/10.1093/jnci/djy206/5303811

Srivastava, S. (2012). Biomarkers in cancer screening: a public health perspective. *Cancer Biomarkers, 10*(2011/2012), 1–2. https://doi.org/10.3233/CBM-2012-0241

Strimbu, K., & Tavel, J. A. (2010). What are biomarkers? *Curr Opin HIV/AIDS, 5*(6),

463–466. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3078627/

Swanton, C., McGranahan, N., Starrett, G. J., & Harris, R. S. (2015). APOBEC enzymes: mutagenic fuel for cancer evolution and heterogeneity. *Cancer Discovery, 5*(7), 704–712. https://doi.org/10.1158/2159-8290.CD-15-0344

Toh, T. B., Lim, J. J., & Chow, E. K.-H. (2017). Epigenetics in cancer stem cells. *Molecular Cancer, 16*(29). https://doi.org/10.1186/s12943-017-0596-9

Tyson, J. J., & Novak, B. (2014). Control of cell growth, division, and death; information processing in living cells. *Interface Focus, 6*(4). https://doi.org/10.1098/rsfs.2013.0070

U. S. Cancer Statistic. (n.d.). Rate of cancer deaths in the United States. Retrieved from https://gis.cdc.gov/Cancer/USCS/DataViz.html

U. S. Census Bureau (n. d.). Decennial Census of Population and Housing. Retrieved from https://census.gov/programs-surveys/decennial-census/data/datasets.2010.html

U.S. Preventive Services Task Force [USPSTF]. (2018*). Lung cancer screening.* Retrieved from https://www.uspreventiveservicestaskforce.org/Page/Document/UpdateSummaryDraft/lung-cancer-screening1

Usuda, K., Saito, Y., Sagawa, M., Sato, M., Kanma, K., Takahashi, S., & Fujimura, S. (1994). Tumor doubling time prognostic assessment of patients with primary lung cancer. *Cancer, 74*(8), 2239–2244. https://doi.org/10.1002/1097-0142

Wagner K.-K., Cameron-Smith, D., Wessner, B., & Franzke, B. (2016). Biomarkers of aging: From function to molecular biology. *Nutrients, 8:*338, 1–3. https://doi.org/10.3390/nu8060338

Wang, B.-H., Zhou, L.-Y., Zhang, H.-L., Li, Y.-Y., Han, J.-Z., Lv, Y.-Q., Zhang, H.-L., & Zhao, L. (2017). Gene methylation as a powerful biomarker for detection and screening of non-small cell lung cancer in blood. *Oncotarget, 8*(19), 31692–31704. https://doi.org/10.18632/oncotarget.15919

Wang, C., Ding, M., Xia, Mingde, Chen,S., Van Le, A., Soto-Gil, R.,…& Zhang, C. (2015). A five-miRNA panel identified from a multicentric case-control study serves as a novel diagnostic tool for ethnically diverse non-small-cell lung cancer patients. *The Lancet*, *2*(10), 1377–1385. https://doi.org/10.1016/j.ebiom.2015.07.034

Wang, C., Liang, H., Lin, C., Li, F., Xie, G., Qiao, S.,... & Zhang, X. (2019). Molecular subtyping and prognostic assessment based on tumor mutation burden in patients with lung adenocarcinomas. *International Journal of Molecular Sciences, 20*(4251), 1–13. https://doi.org/10.3390/ijms20174251

Wang, W., Feng, X., Duan, X., Tan, S., Wang, S., Wang, T.,… & Wu, Y. (2017). Establishment of two data mining models of lung cancer screening based on three gene promoter methylations combined with telomere damage. *International Journal of Biologic Markers, 32*(1), e141–e146. https://doi.org/10.5301/jbm.5000232

Weiss, R. A. (2004). Multistage carcinogenesis. *British Journal of Cancer, 91*(12), 1981–

1982. https://doi.org/10.1038/sj.bjc.6602318

White, M. C., Holman, D. M., Boehm, J. E., Peipins, L. A., Grossman, M., & Henley, S. H. (2014). Age and Cancer Risk: A potentially modifiable relationship. *American Journal of Prevention Medicine.* 46(301):S7-15. doi:10.1016/j.amepre.2013.10.0296.

World Health Organization [WHO]. (2019). *Cancer: key facts.* Retrieved from http://www.who.int/en/news-room/fact-sheets/detail/cancer

World Health Organization. (2011). *Biomarker and Human Biomonitoring.* https://www.who.int/health-topics/children-environmental-health

World Health Organization. (2019). *Cancer.* Retrieved from https://www.who.int/cancer/en/

Zamay, T. N., Zamay, G. S., Kolovskaya, O. S., Zukov, R. A., Petrova, M. M., Gargaun, A.,… & Kichkailo, A. S. (2017). Current and prospective protein biomarkers of lung cancer. *Cancers, 9:*155. https://doi.org/10.3390.cancers9110155

Xia, X., Chen, W., McDermott, J., & Han, J.-D. J. (2017). Molecular and phenotypic biomarkers of aging [version 1; referees: 3 approved]. F1000Research, 6(F100 Faculty Rev):860, 1–10. https://doi.org/10.12688/f1000research.10692.1

Xiao, D., Pan, H., Li, F., Zhang, X., & He, J. (2016). Analysis of ultra-deep targeted sequencing reveals mutation burden is associated with gender and clinical outcome in lung adenocarcinoma. *Oncotarget, 7*(16), 22857–22864. https://doi.org/10.18632/oncotarget.8213

Xie, S-H., Rabbani, S., Petrick, J. L., Cook, M. B., & Lagergren, J. (2017). Racial and

ethnic disparities in the incidence of esophageal cancer in the United States, 1992–2013. https://doi.org/10.1093/aje/kwx221

Xing, A., Pan, L., & Gao, J. (2018). P100 functions as a metastasis activator and is targeted by tumor suppression miRNA-320a in lung cancer. *Thoracic Cancer, 9*, 152–158. https://doi.org/10.11111/1759-7714.12564

Yabroff, K. R., Gansler, T., Wender, R. C., Cullen, K. J., Brawley, O. W. (2019). Minimizing the burden of cancer in the United States: Goals for a high-performing health care system. *CA Cancer Journal for Clinicians, 69*(3), 166-183. http://doi.org/10.3322/caac.21556

Yang, J.-S., Li, B.-J., Lu, H.-W., Chen, Y., Lu, C., Zhu, R.-X., Liu, S.H., Yi, Q.-T., Li, J., & Song, C.-H. (2015). Serum miR-152, miR-148a, miR-148b, and miR-21 as novel biomarkers in non-small cell lung cancer screening. *Tumor Biology, 36*(4), 3035–3042. https://doi.org/10.1007/s13277-014-2938-1

Yang D., Yang K., & Yang M. (2018). Circular RNA in Aging and Age-Related Diseases. In: Wang Z. (eds) Aging and Aging-Related Diseases. *Advances in Experimental Medicine and Biology*, vol 1086. Springer, Singapore.

Yokoyama, A., Kakiuchi, N., Yoshizato, T., Nannya, Y., Suzuki, H., Takeuchi, Y., … & Ogawa, S. (2019). Aged-related remodeling of oesophageal epithelia by mutated cancer drivers. *Nature, 565*(17), 312–317. https://doi.org/10.1038/s41586-018-0811-x

Zamay, T. N., Zamay, G. S., Kolovskaya, O. S., Zukov, R. A., Petrova, M. M., Gargaun,

A.,…& Kichkailo, A. S. (2017). Current and prospective protein biomarkers of lung cancer. *Cancers*, *9*(155), 1–22. https://doi.org/10.3390/cancers/9110155

Zhang, C., & Zeng, X. (2013). Cell proliferation, processes, regulation and disorders. Nova Science Publishers, Incorporated, New York, N. Y.

Zhang, H., Mao, F., Shen, T., Luo, Q., Ding, Z., Qian, L., & Huang, J. (2017). Plasma miR – 145, miR-20a, miR-21 and miR-223 as novel biomarkers for screening early-stage non-small cell lung cancer. *Oncology Letters, 13,* 669–676. https://doi.org/10.3892/ol.2016.5462

Zheng, Y., Joyce, B., Collicino, E., Liu, L., Zhang, W., Dai, Q.,…& Hou, L. (2016). Blood epigenetic age may predict cancer incidence and mortality. *EBioMedicine*, (2016), 68–73. https://doi.org/10.1016/j.ebiom.2016.02.008

Appendix A: Table Showing the Demographic Factors of Lung Cancer

Table 1.

Demographic Factors of Lung Cancer

Characteristics		**Frequency**	**%**
Sex			
	Women	28,035	44.4
	Men	35,075	55.6
	Total	63,107	100
Race/Ethnicity			
	White	55,049	87.2
	Black	8,058	12.8
	Total	63,107	100.0
Age group			
	(45–49) years	1,536	2.4
	(50–54) years	3,831	6.1
	(55–59) years	6,550	10.4
	(60–64) years	8,967	14.2
	(65–69) years	11,548	18.3
	(70–74) years	11,942	18.9
	(75–79) years	10,734	17.0
	(80–84) years	7,999	12.7
	Total	63,107	100.0
Stage			
	I	8,319	13.2
	II	5,943	9.4
	III	15,441	24.5
	IV	33,404	52.9
	Total	63,107	100.0
Geographical regions			
	Metropolitan Counties	52,835	83.7
	Non-Metropolitan Counties	10,272	16.3
	Total	63107	100.0

Note: SEER─18 Registries Data.