8-2023

Clinical and environmental factors affecting the survival outcomes among stage 1A TN0M0 first primary non-small cell lung carcinoma patients in the United States.

Naiya G. Patel
University of Louisville

CLINICAL AND ENVIRONMENTAL FACTORS AFFECTING THE SURVIVAL OUTCOMES AMONG STAGE 1A TN0M0 FIRST PRIMARY NON-SMALL CELL LUNG CARCINOMA PATIENTS IN THE UNITED STATES

By

Naiya G. Patel

D.D.S., Gujarat University India, 2014

M.P.H, Long Island University New York, 2017

A Dissertation
Submitted to the Faculty of the

School of Public Health and Information Sciences at the University of Louisville in Partial Fulfilment of the Requirements for the Degree of

Doctor of Philosophy in Public Health Sciences

Department of Health Management and System Sciences

University of Louisville

Louisville, Kentucky

August 2023

CLINICAL AND ENVIRONMENTAL FACTORS AFFECTING THE SURVIVAL OUTCOMES AMONG STAGE 1A TN0M0 FIRST PRIMARY NON-SMALL CELL LUNG CARCINOMA PATIENTS IN THE UNITED STATES

By

Naiya G. Patel
D.D.S., Gujarat University in India, 2014
M.P.H, Long Island University in New York, 2017

A Dissertation Approved on

June 29, 2023

By the following Dissertation Committee:

_____

Seyed M. Karimi, Ph.D., Dissertation Chair

_____

Bert Little, Ph.D., Dissertation Committee Member

_____

Michael E. Egger, MD, Dissertation Committee Member

_____

Demetra Antimisiaris, PharmD, Dissertation Committee Member

DEDICATION

I dedicate this dissertation to my strongest family support system, especially my parents, who encouraged and envisioned their daughter to become a first-generation Ph.D., regardless of cultural and financial barriers. To my niece and her mother, who constantly remind me of what matters in life. To my husband, who defied cultural norms and encouraged me to pursue my dreams. Finally, I dedicate this dissertation to my late grandfather, who left us because of progressive lung cancer.

ACKNOWLEDGMENTS

I want to acknowledge the members of my dissertation committee, Dr. Seyed Karimi, Dr. Bert Little, Dr. Michael Egger, and Dr. Demetra Antimisiaris, for providing unwavering support. A special thanks to my dissertation chair Dr. Seyed Karimi, who went above and beyond his responsibilities to provide the support needed to continue pursuing my Ph.D. dreams during the most challenging situation. I would also like to thank Dr. Bert Little for providing continued support during the process. To Dr. Christopher Johnson for extending the admission offer into the Ph.D. program and providing the initial support to thrive, without which I would not have reached this far. To the State of Kentucky's state-university partnership (SUP) program for funding research assistantship, which supported my educational goals without worrying about the financial barriers an international student has to face. To all those anonymous kind donors who financially supported my research presentation opportunities through travel awards during the doctoral program.

I would also like to extend my gratitude to John Bartley, the Information Technology manager at the University of Louisville School of Public Health & Information Sciences, who helped us set up the remote work environment and provided data storage support to generate and store 2.7TB of work produced during this thesis work. To my colleagues and friends at the University of Louisville who are compassionate about doctoral journey obstacles and provide social support to thrive. To Dr. Michael Goldsby and Hamid Zarie, who taught me Python, SQL,

KNIME, and Excel, which seemed beyond reach to learn given my non-quantitative background at the time.

Finally, I would like to extend sincere gratitude to my mom, who prepped meals to help me save time when I was caught up working on projects during an academic year. My husband, who single-handedly ran the daily errands and prepped meals in Louisville. He enabled me to dedicate my undivided attention during my qualifying exam preparation during a pandemic and beyond. I would like to thank my sister, niece, and brother-in-law, who provided the support needed during unforeseen circumstances. To my extended family, who would explain to friends and family what my research is about to non-academic social groups.

ABSTRACT
CLINICAL AND ENVIRONMENTAL FACTORS AFFECTING THE SURVIVAL OUTCOMES AMONG STAGE 1A TN0M0 FIRST PRIMARY NON-SMALL CELL LUNG CARCINOMA PATIENTS IN THE UNITED STATES

Naiya G. Patel

June 29, 2023

Background: Lung cancer is the second leading cause of death in the United States (U.S.). The most prevalent histological type of lung cancer is Non-Small Cell Lung Cancer (NSCLC), which has an overall five years survival of 5% if left untreated. Therefore, early treatment of NSCLC is vital to improve overall survival (OS) outcomes. Several factors affect survival outcomes, which can be categorized as modifiable or non-modifiable. The difference in timely receipt of guideline-concordant treatment affects the survival outcomes of patients with stage 1A NSCLC. This dissertation explored factors that affect treatment and survival outcomes among stage 1A NSCLC patients using a nationally representative cancer registry population (*i.e.*, Survival, Epidemiology, and End Results (SEER) 18 plus cancer registry), air pollution and weather data, and local health resource information.

Methods: The first chapter of this study utilized an evidence-based Directed Acyclic Graph (DAG) synthesis method to review the causal pathways explored in

the literature for factors affecting treatment receipt and survival outcomes among stage 1A NSCLC patients in the U.S. Subsequently, guided by the theoretical frameworks proposed previously Andersen and Aday[1] ; Shi and Steven[2], DAGs for the second and third chapter analyses were proposed, respectively. The second chapter utilized logistic regression adjusted for the year of diagnosis and county-specific time-invariant factors with standard errors clustered at the state level. The third chapter employed Kaplan-Meier survival estimates and a Cox proportional regression model adjusted for the year of diagnosis and county-specific time-invariant factors to determine survival outcomes, allowing for right censoring.

Findings: The DAGs identified several causal pathways that were accounted for in Chapters Two and Three analyses. The results of the Chapter Two analysis corroborated with the existing literature that there exists a difference in guideline-concordant treatment receipt. The findings of the Chapter Three analysis confirmed a difference in survival outcomes among stage 1A TN0M0 NSCLC first primary patients exposed to higher versus low levels of air pollution in the U.S.

Policy implications: Black and Medicaid enrolees are less likely to receive guideline-concordant treatment than others are. This warrants future policy decision-making geared toward reducing the difference in treatment receipt, which ultimately improves survival outcomes. Additionally, the shortage of sufficient air pollution monitoring stations in non-metropolitan areas warrants an improvement in determining the health outcomes for non-metropolitan residents. Ambient air pollution control policies are required to improve the survival outcomes of patients with stage 1A TN0M0 NSCLC.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 BACKGROUND

Lung cancer is the leading cause of cancer-associated mortality rates in the U.S. and the second most common cancer among the population.[3] The most prevalent type of lung cancer is Non-Small Cell Lung Cancer (NSCLC), and without treatment, the overall five years survival for stage 1A NSCLC is 75%.[4] Lung cancer survival outcomes depend on the type of treatment received and the comorbidities present at the time of treatment.[5] Only a few studies have explored the survival outcomes among treated lung cancer patients in the presence of key air pollutants.[6] Lung cancer is also a type of cancer that rapidly progresses after diagnosis and leads to more fatalities in the U.S. under a lack of appropriate treatment receipt.[7] While it is evident that the standard guideline for stage 1A Non-Small Cell Lung Cancer (NSCLC) is either lobectomy, limited resection with adjuvant radiotherapy, or radiotherapy, it is important to determine the factors that drive the decision to receive treatment.

Despite convincing evidence that survival among stage 1A NSCLC patients significantly increases if appropriately treated at stage 1A, studies utilizing the Survival, Epidemiology, and End Results (SEER) registry data did not address potentially significant sources of selection bias associated with treatment selection.[8] The sources of bias include patients' preferences, doctors' recommendations, and treatment guideline revision. In particular, accounting for treatment guideline revisions in determining treatment selection is lacking in the existing literature. It is vital to understand whether there is a trend in the type of treatment received due to revisions of

existing treatment guidelines or whether there is a racial, geographic region, physician recommendation, medical insurance coverage, and general socioeconomic difference in receiving cancer treatment regardless of the revisions.

Studies that use retrospective human data, such as data from the SEER cancer registry, have not studied the association between exposure to air pollutants and weather in a multi-pollutant model after the type of treatment received on Overall Survival (OS) among patients with stage 1A NSCLC. Air pollution levels vary by geographic area because of the presence of its source of production and certain weather conditions. This leads to differential exposure over longer periods in the presence of certain weather conditions.[9] In addition, there is a gap in determining the effect size on OS among the standard guided care: lobectomy or limited resection with adjuvant radiotherapy for operable fit and radiotherapy for inoperable fit in the presence of weather and air pollutants.[6] Finally, pre-diagnosis patient exposure has not yet been considered in estimating the risk of death in relevant past literature. While the hypothesis of air pollution playing a vital role in lung cancer survival is made,[6,10,11] it has rarely been included in the analyses of stage 1A lung cancer survival.[6] Only one study attempts to account for air pollution in determining the survival period after a lung cancer diagnosis. However, the study did not specifically focus on patients with stage 1A TN0M0 first primary tumor. Instead, it generalizes staging and focuses only on one U.S. state rather than the entire U.S. population. Additionally, it did not account for pre-diagnosis exposure assignment to patients for air pollutants as well as weather components to account for pre-diagnosis exposure effect on OS.[6] Finally, the study did not account for other primary air pollutants such as SO_2, CO, and weather components. It is evident that air pollutant concentrations vary spatiotemporally for several reasons, such as chemical reactions between pollutants or local physical reactions of the source

of production with other factors.[12] Hence, it is vital to understand the effects of multiple pollutants and weather conditions on chronic health outcomes.

1.2 SIGNIFICANCE

There are differences in incidence, diagnosis, and OS among different U.S. states, with some states suffering significant incidence, diagnosis, and OS difference compared to others. Kentucky is one of the 50 states in the U.S. that is disproportionately affected by these differential rates and outcomes.[13] The difference is indicative of several underlying factors, such as contextual community factors, healthcare access barriers, population demographics, and exposure to environmental components. Contextual factors include the causes of lung cancer that are categorized into risk and preventive factors. Both of these factors could be changed and prevented. Risk factors include tobacco smoke, second-hand smoke, radon exposure, asbestos exposure, radioactive element exposure, and specific agonist dietary supplements. Other risk factors that cannot be changed include previous radiation therapy to the lungs, air pollution, and a history of familial lung cancer. Prevention factors include avoiding tobacco consumption, eating a healthy diet, avoiding cancer-causing radioactive agents, preventing exposure to particulate air pollutants, and limiting exposure to those agents when avoidance is not possible.[14] Among the healthcare access barriers, nonfinancial barriers are common reasons behind unmet needs or delayed care among adults in the U.S. compared to affordability barriers.[15] Additionally, several underlying disparities associated with continuous medical insurance coverage have also been determined to be the primary cause of the inability to access improved cancer screening and treatments for better outcomes.[16] Hence, this study is particularly relevant in the context of chronic diseases in Kentucky, as it could help determine key racial,

medical insurance coverage, geographic region, and general socioeconomic difference factors affecting such states disproportionately.

This study is particularly relevant in the context of chronic diseases in Kentucky for the reasons mentioned above and for determining the specific racial, medical insurance coverage, geographic region, and general socioeconomic difference factors that exist. The age-adjusted new lung cancer incidence rate in Kentucky is 89 per 100000, which is significantly higher than the national rate (58 per 100,000).[13] The lung cancer incidence rate in Kentucky is greater than the national rate for both whites and Blacks with 89 per 100,000 for whites and 82 per 100,000 for Black Kentuckians.[13] Interestingly, the lung cancer incidence rate in Black Kentuckians is lower than that in white Kentuckians, while the opposite is observed at the national level aggregation.[13,17] Nationally, 24% of cases are diagnosed at stage 1, whereas in Kentucky, the stage 1 diagnosis percentage falls into the below-average tier as per American Lung Association which defines tier as state rate compared to national and other state rate.[13] For example, 22% of lung cancer cases among white Kentuckians were diagnosed at stage 1, while the national stage 1 rate for whites is 25%. Among Black Kentuckians, the stage 1 diagnosis rate was 20%, compared to the national rate of 21%. In general, minorities have a 3%-4% lower probability of being diagnosed with stage 1 lung cancer diagnosis than white Americans: whites (25%), Blacks (21%), Latinos (22%), Asian Americans or Pacific Islanders (21%), indigenous people (22%).[13,17] In general, the State of Kentucky has the highest lung cancer incidence rate, almost 2.3 times that of Utah, which has the lowest rate nationally.[17]

It is evident that racial, socioeconomic, geographic, survival, and medical insurance differences exist in present study context as discussed prior in this section. However, several empirical and theoretical gaps exist in the literature (e.g, utilizing

Andersen and Aday[1], Shi and Steven[2] theory framework to guide statistical decision model factors, building DAGs to illustrate comprehensive study relationship to determine gold standard statistical modelling, including identified key confounders like treatment guideline revision years, patient preference , doctor's recommendation, state-level analysis without accounting for local-level effect through standard-error clustering, adjustment of local resources through inclusion of county-specific, year of diagnosis time-invariant unobserved confounders, weather and air pollution components such as wind, SO_2, CO, and finally assigning pre-diagnosis exposure to patients in an attempt to determine carry-over effect) that attempt to estimate those differences for stage 1A NSCLC in the U.S.. Hence, this dissertation is aimed at filling the empirical and theoritcal gaps by advanced data techniques allowing to include key identified confounders and exposure assignment, statistical modelling, as well as causal diagram development that would inform statistical modeling. Identifying an association between treatment guideline revision years, patient and physician preference, and other relevant areas and provider-level variables in the study context for stage 1A NSCLC primary tumors and treatment receipt can guide future studies to build on existing results, data techniques, or study limitations. Identifying the close-to-true hazards of exposure to air pollutants, weather, and the type of treatment received on OS can guide future studies focusing on stage 1A primary tumors to build on existing results, and pre-diagnosis exposure assignment and multiple longitudinal data merging techniques.

Hence, the objective of the proposed research is threefold: (1) to identify factors that affect treatment receipt and OS among the stage 1A primary NSCLC population using real-world data, (2) to determine factors predictive of treatment receipt, and (3) to determine (a) whether inclusion of environmental factors improve close-to-true all-cause hazards and (b) whether there is a difference in OS.

The method utilized in this study generates post-treatment guideline revision years after each revision is introduced. Patient preference and doctor's recommendation variables were generated following standard treatment guideline revision years and the logic model of treatment decision-making for stage 1A NSCLC TN0M0 patients. Additionally, relevant area-level and provider-level information were merged with SEER data to include provider and county socio-demographic factors in determining its role in treatment receipt. In addition, the nearest air pollution monitors and weather stations to the subjects' county of residence were identified, and their recorded values were assigned to the subjects before and after the diagnosis. Utilizing advanced data techniques to assign pre-diagnosis exposure to patients might aid in improving the determination of the association between environmental exposure and OS, which seems to be a potential factor affecting cancer survival, as per the Canadian cancer risk population health management model[18] and Shi and Steven's[2] general framework model for studying vulnerable populations. To our knowledge, this study is the first to utilize Shi and Steven's[2] vulnerable population model to determine the association between air pollutants, weather, and survival outcomes, and to assign pre-diagnosis exposure to SEER. Additionally, this is the first study to utilize the Andersen and Aday[1] conceptual framework to determine the factors affecting treatment receipt by including key factors, such as treatment guideline revision years and patient and physician preferences.

1.3 PAPER OVERVIEW

The aim of the study was to determine factors affecting treatment reciept and survival outcomes among stage 1A NSCLC patients in the U.S. utilizing real world

data. As mentioned prior the existence of identified empirical and theoretical gaps guided each aim to attempt to fill those gaps in literature. Hence, the three aims are as described in following paragraphs of this section.

1. AIM 1: To identify factors associated with non-treatment among stage 1A TN0M0 NSCLC with first primary tumor and determine risk factors associated with lung cancer-specific survival after surgery and radiation therapy through a systematic review.

The published literature in study context was surveyed to answer several questions (*i.e.,* What factors affect the survival of patients with stage 1A primary NSCLC in the U.S.? What determines the treatment choice?). Existing literature that have analzyed these factors were surveyed to determine literature gaps. This chapter provided information for devising the statistical models of decision-making for the second and third chapters by building evidence synthesis-directed acyclic graphs (DAG) to determine confounders, colliders, and mediators. The systematic review was divided into two parts to determine: a) factors affecting treatment receipt, and, b) factors affecting OS. Guided by a systematic review, integrated DAGs (iDAG) were built, based on which we proposed the final iDAGs relevant to the present study that informed the statistical modelling of Chapters Two and Three.

2. AIM 2: To identify factors associated with treatment receipt among stage 1A TN0M0 first primary NSCLC by utilizing U.S. population-based cancer data, local socio-demographic information from U.S. censuses, and Area Health Resource Files.

What determines treatment receipt? Does treatment receipt differ by patient demographics and tumor characteristics? We hypothesized racial, medical insurance, and general socioeconomic difference in treatment receipt among the stage 1A NSCLC

first primary tumor population.[19,20] Chapter Two identified factors that affect treatment receipt by utilizing and comparing cohorts with and without air pollution and weather monitoring exposure assignment data. This allowed us to identify whether cohort compositional effects affected the study results and helped us determine whether there was a difference in racial, type of medical insurance coverage, geographical region factors affecting treatment receipt differences. The rationale behind analyzing two separate study samples for regression analysis was to validate if estimates from study sample corroborate with estimates of sub-sample. To our knowledge, this is the only study that attempts to apply advanced data techniques to merge information with SEER data by obtaining AHRF from historical data and categorizing them into groups defined by the Social Determinants of Health (SDOH) beta files by the Agency for Health Research and Quality (AHRQ). Our literature search strategy in Chapter One across three frequently used databases (*i.e.,* PubMed, EMBASE, and Web of Science) and grey literature search informed our claim made in the prior sentence. The study used a logistic regression model adjusted for the year of diagnosis and county-specific time-invariant factors with standard errors clustered at the state level.

3. AIM 3: To evaluate whether exposure to levels of certain air pollutants is associated with OS among patients with stage 1A TN0M0 first primary NSCLC undergoing treatment of choice by utilizing U.S. population-based cancer data and U.S. environmental air pollution data.

We hypothesized a difference in OS; in other words, all-cause hazards among higher versus lower exposure groups in the presence of treatment type.[6] In Chapter Two, we determined racial, type of medical insurance coverage, geographical and general socioeconomic difference factors associated with treatment receipt. In Chapter Three, using a sub-sample with additional information regarding patients

exposure to weather and air pollutants is included to determine their OS. One published study evaluated the association between air pollutants and all stages of lung cancer using the California Cancer Registry. However, the study did not account for weather components, SO_2 and CO air pollutant exposures, and their interactions with treatment types. The previous study also did not assign exposure before diagnosis to determine cumulative hazards over a longer period of time that might have a carry-over effect on outcomes.[6] Air pollution and type of treatment affect lung function and morbidity,[21,22] so it is important to understand the time to death depending on the type of treatment received and in the presence or absence of environmental pollutant and weather condition pattern with air pollution exposures. Air pollutants (*i.e.*, NO_2, SO_2, CO, PM, Lead, Benzene) is an apparent factor affecting cancer survival, according to the Canadian cancer risk population health management model.[18]. This study used Shi and Steven's[2] vulnerable population model to analyze survival outcomes. Prior published reports using SEER data merged with U.S. air pollution data (*i.e.*, NO_2, SO_2, CO, PM, Ozone, Lead, Benzene) and Area Health Resource Files (AHRF) did not assign pre-diagnosis exposure specifically to stage 1A NSCLC TN0M0 patients in the registry. Similarly, no prior published studies were located that included weather exposure in air pollution model in a given study context. Kaplan-Meier survival and Cox regression analyses were used to analyse OS, with right censoring. The analysis was to determine in the all-pollutant (multi-pollutant) and single-pollutant models for one, three, and five years of survival. Exposure to air pollutants and weather components was analzyed in the three groups noted above (one, three, and five years) before the diagnosis until the survival assessment period. This is a major strength of the present study, and extends the findings from prior studies whose major flaw was the absence of exposure prior

to diagnosis. It is logically naïve to neglect factors before diagnosis because cancers usually develop following long-term time-varying exposures. Even associational studies that do not include pre-diagnosis expsoures are confounded in unknown ways that make them highly difficult to interpret. The results of the present analysis may be used to help interpret prior studies that were confounded by no exposure prior to diagnosis.

The three chapters jointly help to identify clinical and environmental factors that determine factors affecting treatment receipt and survival outcomes for stage 1A NSCLC patients with a first primary tumor in the U.S.. This thesis will serve as a foundation for future studies utilizing SEER to account for unaccounted factors and help researchers build evidence-based exposure assignment models by merging environmental data for all cancer types in the SEER registry's real-world outcomes. This study will also serve as a foundation for future research to determine the hazards of air pollutants accounting for the type of treatment for stage 1A NSCLC survival outcomes. Finally, the systematic review performed in this study will serve as a foundation for meta-analysis for future studies that might be interested in exploring study quality indicators for health services research.

The novelty of this study above prior research is the analysis of pre-diagnosis exposure, and holds the greatest promise for refining the direction of future studies of environmental and weather exposures prior to lung cancer diagnosis. This research will close the loop on environmental and cancer association studies because it is one step closer to evaluating causality.

CHAPTER 2: FACTORS AFFECTING TREATMENT RECEIPT AND SURVIVAL OUTCOMES FOR STAGE 1A NON-SMALL CELL LUNG CANCER UTILIZING EVIDENCE SYNTHESIS FOR CONSTRUCTING DIRECTED ACYCLIC GRAPHS.

2.1 INTRODUCTION

A long-standing association exists between lung cancer survival and socioeconomic factors in epidemiologic studies on early-stage Non-Small Cell Lung Cancer (NSCLC).[23] Several factors contribute to the geographic differences among early-stage NSCLC regarding comorbidity status and carcinogen exposure.[24] Geographic area is a critical factor in treatment utilization and survival for early-stage NSCLC.[24] Treatment modalities for stage 1A NSCLC include surgery for medically fit candidates or radiation therapy for medically unfit candidates.[25] Differentiating treatment modalities are associated with survival outcome differences.[24,26] However, limited scientific literature exists that develops causal diagrams in Directed Acyclic Graphs (DAG) regarding factors affecting treatment receipt and survival outcomes among stage 1A TN0M0 NSCLC.

DAG is a simple graphical representation of causal relation assumptions in the study context and multiple factors that must be accounted for to obtain the unconfounded relationship between the exposure and the outcome variable.[27] Conventional statistical models contain several parametric assumptions that may or may not be correct.[27] It is a drawback when identifying assumptions in a study context and model violation.[27] Causal diagrams depicted through DAG represent those study assumptions that can complement statistical models.[27] Causal diagrams illustrate causal

relationships without any parametric assumptions, as in the case of conventional statistical models.[27] However, causal diagrams capture the series of causation in the current study context, which a conventional model might not be equipped to do. Some causal relation assumptions might be untested or unknown, but a causal diagram can capture all possible causal pathways.[27]

This chapter aimed to identify factors associated with non-treatment among stage 1A TN0M0 NSCLC with the first primary tumor and determine risk factors associated with lung cancer-specific survival after surgery and radiation therapy through a systematic review. What factors affect OS in patients with stage 1 primary NSCLC in the United States (U.S.)? What determines the treatment choice? These are the key questions that we aimed to seek through a comprehensive DAG-guided systematic literature review of the topics. We reviewed the existing literature that has sought to measure these factors, especially from the perspective of treatment selection and lung cancer-specific survival among national cancer registry data or clinical trials. This chapter will guide the model decision-making of Chapters Two and Three by determining the associational pathways that need to be accounted for.

2.2 METHODS

A literature search strategy was developed with the assistance of a librarian expert, oncologist, and health economist to ensure that exhaustive literature was included. Three key databases were identified: Embase, PubMed, and Web of Science. For gray literature, the Connected Papers database,[28] manual searching by bibliographic browsing of key research articles relevant to the study topic, and the Clinical trials.gov[29] database were used. Only studies focused on the U.S. were moved toward the final

sample from the body screening stage, as the clinical staging and treatment guidelines differ internationally. Therefore, the present study focuses on the U.S.

Additionally, the current study aims to develop causal diagrams to supplement the statistical model variables for Chapters Two and Three, which utilize U.S. data. The homogeneity of the included sample in terms of the country was emphasized better to understand the causal relationship within the location context; for the literature database, the publication year was set to 2002-2023 only to capture studies relevant to recent medical advances in this field as well as clinical staging AJCC 6th and higher. An approach was strategized during screening phase of study in an attempt to be consistent across study periods dealing with three different AJCC staging. The included studies ranged from AJCC 6th to 8th edition hence a strategy of following TN0M0 staging convention first in hierarchy decision was developed to be consistent in definition of non-metastatic tumor as mentioned in the screening questions section of this paper, in an attempt to avoid exclusion of studies that do not refer to specific AJCC staging information yet focus on overall stage 1. The hierarchical strategy was informed by American College of Surgeon (ACS) and American Cancer Society that emphasises on TNM staging serving as foundation to defining overall AJCC staging system. The definition of stage 1A from AJCC 6th to 8th moves from broader categorization of T staging to more granular and for the same reason the study characterics Appendix Table 2.1 informs about particular tumor staging each finalized study included since the definition of stage 1 was relative across AJCC 6th to 8th.Clinical staging informs definitive treatment decision-making and affects survival outcomes,[30] so it is very important to consider studies published after 2001, as the AJCC 6th edition was implemented after that year. No publication year filter was applied for studies found through gray literature searching to capture studies that might be important and

relevant to the current topic context. Limited empirical evidence is available to develop a systematic review bibliographic search strategy for healthcare. However, an experiment determined that significant relevant studies were found on Embase, ranking it second highest of all the pertinent databases of search results.[30] PubMed comprises citations from Medline, another relevant medical literature database, while Web of Science provides only peer-reviewed studies. The decision to use these three databases was made after consultation with a librarian expert and an oncologist. The search strategy across each database is as described in Appendix. DAGitty, an online software program, developed implied DAGs and integrated DAGs graphics.[31]

2.2.1 PROTOCOL

The protocol for developing the final integrated DAGs (iDAGs) from the shortlisted literature was informed by 'Evidence Synthesis for Constructing Directed Acyclic Graphs' (ESC-DAGs).[32] The current review is divided into two components: a) factors affecting treatment receipt and b) factors affecting survival outcomes. The conceptual model utilized in the translation stage of ESC-DAGs for factors affecting treatment receipt is Andersen and Aday[1] health services research behavioral model. While for factors affecting survival outcome, the conceptual model by Shi and Steven[2] for vulnerable populations is utilized. Therefore, the translation-stage decisions were guided solely by these conceptual frameworks for each component regarding temporality and construct validity. The title, abstract, and body screening questions were as follows: (1) is it about stage 1 first primary NSCLC TN0M0; (2) is the document an article?;(3) is it quantitative research?(4) is it about assessing the factors affecting treatment receipt? OR is it about assessing the factors affecting survival outcomes?

2.3 RESULTS

A total of 36 of 9,421 studies qualified for final data extraction (Figure 2.1). The baseline characteristics of each study are described in Appendix Table 2.1. It describes the study setting, study period, data registry utilized for the study, the age

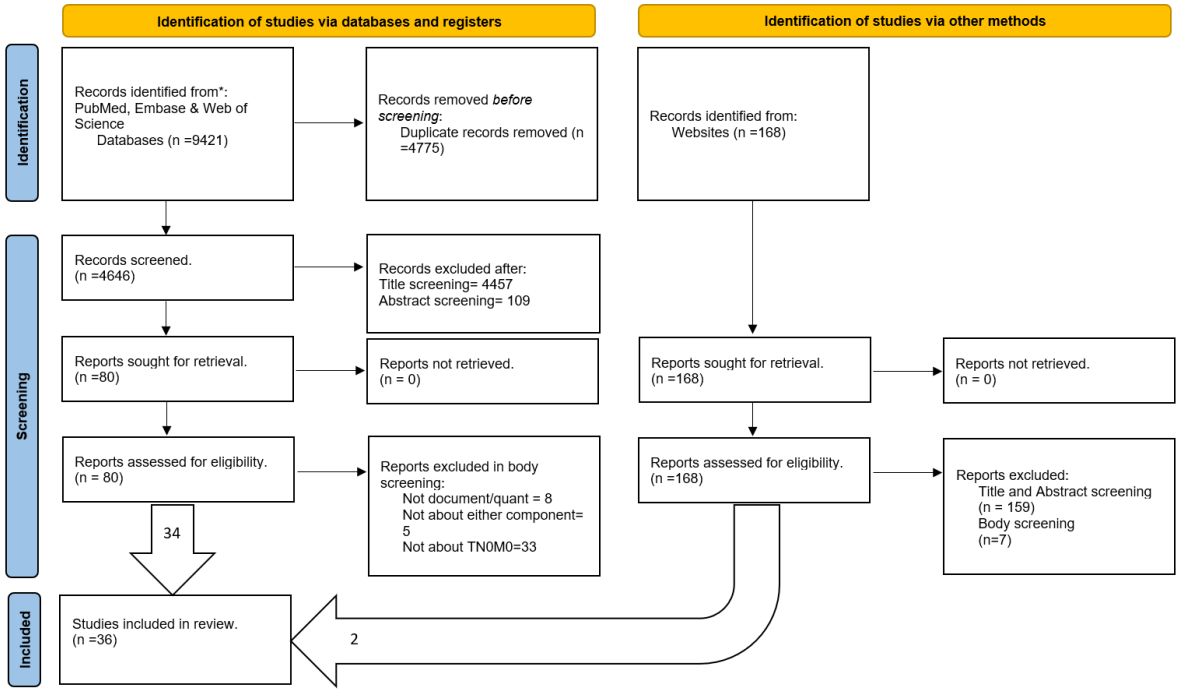


Figure 2.1 PRISMA flowchart

range of the population included, sample size, type of intervention (exposure variable), outcome, AJCC staging version used for study inclusion, and the component under which the study falls. The qualified literature study period ranged from 1988 to 2021, and the study designs were both observational and clinical trials. The observational studies utilized the SEER, National Cancer Database (NCDB), SEER-Medicare linked, California Cancer Registry (CCR part of the SEER registry), and primary data collection in clinical trials.

2.3.1 MAPPING STAGE

In this stage, each qualified study was used to extract data to develop implied DAGs (Appendix Table 2.2 and 2.3) using DAGitty software separately for each component of the review, as explained in the protocol section of this chapter (Tables 2.1 and 2.2). The studies were read in detail to determine the significant confounders, unobserved/unadjusted confounders, mediators, and colliders. In the implied DAG, gray bubbles were the identified confounders in each study that were not adjusted in their statistical analysis model. The green bubbles are the study's exposure variables and the blue bubble is the study's outcome variable. Green arrows are the front door causal pathways, while purple indicates the back door pathways that must be closed to achieve a true causal relationship within the study relationship context. In this stage, implied DAGs were developed as determined by the studies, and arrow edges were directed as identified in the study results or conclusions. No decision about maintaining or removing the arrow edge was made at this stage.

	Exposure variable group	Outcome variable	Odds Ratio(95% CI)	Confounders based on study conclusion	Identified mediators	Identified colliders	Statistical Analysis approach
1[33]	Race	Treatment receipt	Black men: 0.13 (0.04-0.47) Black women: 0.87 (0.13-3.69)	Age at diagnosis Smoking status Coronary artery disease COPD Tumour histology Diagnosis after screening	None	None	Multinomial logistic regression
2[34]	Race	Treatment receipt	Black: 1.43 (1.18-1.72)	Age Race Gender Marital status Insurance type Neighbourhood Socioeconomic Status NCI designation of hospital Tumour size	None	None	Multivariable logistic regression
3[35]	Race	Treatment receipt	Black: 1.58	Race Gender Marital Status Age Income Insurance status	None	None	Multivariate logistic regression
4[36]	Geographic region	Treatment receipt	Rural: 0.04 Urban: 0.24	Gender Race Insurance status Income Age Health status	None	None	Multivariate logistic regression

5[37]	Geographic region	Treatment receipt	Urban: 0.92 (0.85-1.01) Rural: 1 (0.82-1.22)	Age Race Lower education Lower median income	None	None	Multivariate logistic regression
6[38]	Race	Treatment receipt	Black: 0.59	Insurance status Education Race Age Gender Comorbidity score Tumor size Geographical area of residence	None	None	Multivariate logistic regression
7[39]	Race	Treatment receipt	Black: 0.61 (0.58-0.64)	Gender Geographic region Type of treatment facility Rural urban region Insurance Status Comorbidity score	None	None	Multivariate logistic regression
8[40]	Disability status	Treatment receipt	Disabled: 0.27	Age	None	None	Bivariate logistic regression

Table 2.1 Factors affecting treatment receipt mapping stage of ESC-DAG

	Exposure variable group	Outcome variable	Odds Ratio (95% CI)	Hazards Ratio (95% CI)	Confounders based on study conclusion	Identified mediators	Identified colliders	Statistical analysis approach
1[41]	Treatment type	Survival		1.38 (0.70-2.73)	Tumor size Tumor histology Patient performance status Age	None	None	Propensity Score Matching
2[42]	Treatment type	Survival		1.66(1.51-1.83)	Age Gender Tumor size Histology type	None	None	Cox regression
3[43]	Treatment type	Survival		Surgery: 0.18 Radiation: 0.51 Both:0.36	Age Gender Tumor histology T staging	None	None	Cox regression
4[44]	Treatment type	Survival	Surgery: 3.65 Radiation: 7.43		Age Histology type	None	None	Kaplan Meier and Log rank test
5[45]	Histologic type	Survival		Lobectomy: 0.92 (0.83-1.02)	Treatment type Age Gender Tumor grade Number of resected lymph nodes Tumor size	None	None	Propensity Score Matching Cox regression
6[25]	Treatment type	Survival		Surgery: 0.91(0.86-0.96) Radiotherapy: 0.77(0.71-0.83)	Year of diagnosis Gender Race Age	None	None	Cox regression
7[46]	Marital status	Survival		Married: 0.85(0.82-0.89) Divorced: 1.08(1.02-1.15)	Gender Race Tumor grade Age Tumor size	None	None	Cox regression
8[47]	Treatment type	Survival		Segmentectomy: 0.83(0.71-0.96) Radiotherapy: 0.65(0.52-0.81)	Age Gender Tumor size Tumor grade Adjuvant therapy	None	None	Propensity Score Matching Cox regression

9[48]	Treatment type	Survival	Radiofrequency ablation (RFA): 1.25		Hospital region Year of diagnosis Tumor size	None	None	Propensity score matching
10[49]	Treatment type	Survival	RFA: 1.23 SBRT: 0.13		None	None	None	Propensity score matching
11[50]	Treatment type	Survival		RFA: 0.97(0.86-1.11)	Gender Age Tumor size Histologic type Tumor grade Insurance status	None	None	Propensity score matching and Cox regression
12[51]	Treatment type	Survival		Lobectomy: 0.78(0.41-1.48) Adjuvant radiotherapy:0.14 (0.03-0.64)	Gender Number of Lymph nodes examined	None	None	Propensity score matching and Cox regression
13[52]	Treatment type	Survival		Sublobectomy: 1.40(1.25-1.58)	Age Gender Tumor grade Histologic type Tumor size Number of lymph nodes sampled	None	None	Propensity score matching and Cox regression
14[53]	Treatment type	Survival		Radiotherapy: 2.42(2-3)	Age Gender Histologic type Number of Lymph nodes examined Tumor grade Year of diagnosis Tumor size	None	None	Propensity score matching and Cox regression
15[54]	Treatment type	Survival		Lobectomy: 0.76(0.60-1) Segmentectomy: 0.80(0.54-1.18)	Age Gender Lymph nodes examined status Tumor grade	None	None	Cox regression
16[55]	Treatment type	Survival		Segmentectomy: 1.44(1.11-1.86)	Age Tumor grade	None	None	Propensity score matching and Cox regression
17[56]	Treatment type	Survival		Lobectomy: 0.82(0.77-0.87)	Race Tumor size Gender Tumor grade Age	None	None	Cox regression
18[57]	Treatment type	Survival		SBRT: 1.56(1.50-1.62) RFA: 1.91(1.73-2.10) VATS: 0.55(0.52-0.60)	Age Gender Race Treatment facility type Income Treatment facility location Comorbidity score Tumor size Tumor grade Tumor histology	None	None	Propensity score matching and Cox regression
19[58]	Treatment type	Survival		Segmentectomy: 0.89(0.54-1.46) Wedge resection: 1.29(0.97-1.72)	Lymph node dissection	None	None	Cox regression
20[59]	Treatment type	Survival	Segmentectomy: 0.88(0.76-1.02)		Age Tumor grade Tumor histology Number of Lymph nodes dissected Gender Tumor size	None	None	Kaplan Meier Log rank and Multivariate analysis
21[60]	Treatment type	Survival		Segmentectomy: 1.35(1.18-1.54)	Age Year of diagnosis Gender	None	None	Cox regression

				Adjuvant radiotherapy: 1.91(1.58-2.30)	Tumor size Marital status Insurance status Tumor grade Histologic type Number of lymph nodes dissected			
22[61]	Treatment type	Survival		Thermal ablation: 1.40(1.04-1.86) Adjuvant radiotherapy: 1.68(1.40-2.05)	Race Tumor size	None	None	Propensity score matching and Cox regression
23[62]	Treatment type	Survival		Sublobectomy: 1.45(1.35-1.56)	Gender Tumor size Number of lymph nodes sampled Age	None	None	Cox regression
24[63]	Treatment type	Survival		Lobectomy: 1.12(0.93-1.35)	None	None	None	Propensity score matching and Cox regression
25[64]	Radiation therapy	Survival		Radiotherapy: 0.90(0.84-0.97)	Number of Lymph nodes examined Age Gender Tumor size Tumor grade Tumor histology	None	None	Cox regression
26[65]	Treatment type	Survival		Lobectomy: 1.01(0.93-1.11)	Age Gender Tumor size Tumor histology Tumor grade Number of lymph nodes examined	None	None	Cox regression
27[66]	Treatment type	Survival		SABR: 0.86(0.45-1.65)	Age Gender Race Region of enrolment Tumor histology Tumor size Mediastinal lymph node examination	None	None	Kaplan Meier
28[67]	Treatment type	Survival	RFA: 0.21 (0.00-0.44)		Age Gender Region of enrolment Performance status Vital capacity Lung function	None	None	Kaplan Meier

Table 2.1 Factors affecting survival outcomes mapping stage of ESC-DAG.

2.3.2 TRANSLATION STAGE

At this stage, the extracted implied DAGs for each study were utilized to build a DAG edge index (Appendix Tables 2.4 and 2.5) to determine the arrow directionality decision-making between an implied set of variables. To reach objective decisions, the proposed theoretical frameworks for each component were utilized to determine

whether the arrow directionality was accurate. While deciding to remove or retain the edge, the construct validity and temporality of the edge direction were determined using the theoretical framework. Bi-directionality was determined for each study individually by utilizing their implied graph to determine if the edge direction of the arrow was bi-directional, given the set of variables in the existing study context. For factors affecting treatment receipt (Appendix Table 2.4), Andersen and Aday's[1] theoretical framework was used to guide the construct validity and temporality of a particular arrow direction in a given set of variables. Likewise, Shi and Steven's[2] theoretical framework for vulnerable populations was utilized to identify the factors affecting survival outcomes (Appendix Table 2.5). The arrow edges of studies beginning and ending with factors affecting treatment receipt were removed, as no specific exposure variable was assigned to determine the construct validity of the study context. The rationale for removing the edge was to eliminate broader and vague construct validity factors. The arrow edges ending at factors affecting treatment receipt tend to specify each variable within that study as an individual exposure variable without identifying confounders, mediators, or colliders.

Moreover, decision-making regarding temporality is vague if several factors serve as exposure variables without determining significant covariates. In addition, the present study aimed to identify confounders, mediators, and colliders to inform future statistical modeling variables within a study context. Eliminating the vaguely defined arrow edges helps develop reliable integrated DAGs to determine the true causal pathways.

2.3.3 INTEGRATION STAGE

2.3.3.1 FACTORS AFFECTING TREATMENT RECEIPT

The Andersen and Aday[1] model was used to determine the construct validity and temporality of the exposure variables on outcome treatment receipt. As structure-level components occur first temporally and studies have extracted determined exposure variables, type of treatment facility, geographic area of patient, race, and disability status, the integrated DAGs (iDAGs) include those variables as exposure variables (Figure 2.2a). The outcome variable of interest was treatment receipt. Studies with no specific exposure variable were inconclusive regarding the back-door causal pathways in iDAGs. Hence, the edge index from those studies was removed, as the foundation of DAG is a specific exposure and outcome variable to determine all possible causal pathways. The iDAGs included unobserved confounders of the extracted studies as adjusted confounders to determine the potential back door paths relevant to the current study context. The rationale was to identify all possible confounders, mediators, or colliders in the final DAGs. An individual study might have been misspecified as a confounder but might have appeared as a mediator or collider when included in the integrated DAG.

There were 10 causal paths and 13 covariates in the iDAG for the exposure variables of interest on the outcome variable. The total effect adjustment for the given effect of interest suggests controlling for only the following necessary variables to close all the back door paths (purple lines): age, Chronic Obstructive Pulmonary Disease (COPD), comorbidity score, coronary artery disease, education, sex, health status, income, insurance status, marital status, patient preference, physician preference, and tumor size. The front-door paths (green lines) represent the effects of interest in the extracted studies.

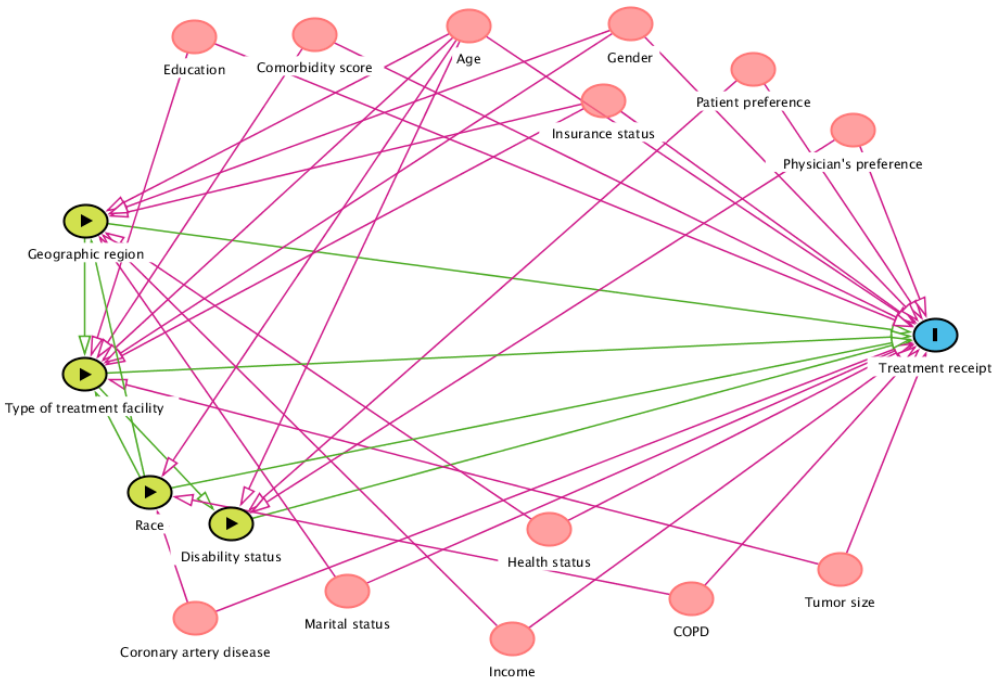


Figure 2.1a Integrated DAG for factors affecting treatment receipt.

The following conditional independence testable implications are identified by the iDAG results for total effect adjustment. In other words, the following must be true in a statistical analysis model, given the effect of exposure on the outcome and multicollinearity. After adjusting for age and type of treatment facility, the comorbidity score was not related to disability status. In addition, the comorbidity score was unrelated to geographic region, insurance status, marital status, patient preference, physician preference, tumor size, education, sex, income, and race.

Adjusting for age and race, coronary artery disease was unrelated to disability status, geographic region, and type of treatment facility. After adjusting for age and type of treatment facility, coronary artery disease was not related to disability status. It is also unrelated to insurance status, marital status, patient preference, physician preference, tumor size, type of treatment, age, COPD, education, gender, or income.

Adjusting for age, sex, geographic region, insurance, and race, disability status was not related to health status, marital status, or income, while adjusting for age, race, and

disability status was unrelated to COPD. Adjusting for age and type of treatment facility, disability status was unrelated to income, race, COPD, education, sex, geographic region, health status, insurance status, tumor size, and marital status. In addition, the geographic region is unrelated to patient and physician preferences, education, and tumor size. Finally, while adjusting for age and race, the geographic region was not related to COPD.

Health status is unrelated to insurance status, marital status, patient preference, physician preference, tumor size, age, education, sex, income, and race. Adjusting for age, sex, geographic region, insurance status, and race, health status was not related to the type of treatment facility.

Insurance status is unrelated to marital status, patient preference, physician preference, tumor size, age, COPD, education, sex, income, and race.

Marital status is unrelated to patient preference, physician preference, tumor size, age, COPD, education, sex, income, and race. After adjusting for age, gender, geographic region, insurance status, and race, marital status was not related to the type of treatment facility.

Patient preference was not related to tumor size, type of treatment facility, age, COPD, education, sex, income, or race. Likewise, physicians' preferences are unrelated to tumor size, type of treatment facility, age, COPD, education, sex, income, and race.

Tumor size was unrelated to age, COPD, education, sex, income, and race. COPD is not related to education, sex, or income. Gender was not related to income or race in the current iDAG.

Adjusting for age, sex, geographic region, insurance status, and race, the type of treatment facility is not related to income. In addition, after adjusting for age and race, the type of treatment facility was not related to COPD.

2.3.3.2 FACTORS AFFECTING SURVIVAL OUTCOMES

The iDAG (Figure 2.3a) has two causal pathways and 32 covariates for the two exposures of interest, marital status, treatment type, and the outcome variable (survival). The two exposure variables are informed by the extracted studies, and Shi and Steven's[2] theoretical framework verifies its temporality. The total effect adjustment for the given effect of interest suggests controlling for only the following necessary variables to close all the back door paths (purple lines): access to care, adjuvant therapy, age, cardiopulmonary function, comorbidities, enrolment bias, sex, hospital region, imaging information, insurance status, lung function, mediastinal lymph node examination, number of lymph nodes examined, number of lymph nodes resected, patient functional status, patient preference, provider bias, quality of life, race, recurrence rate, region of enrolment, smoking status, surgeon expertise, T staging, treatment facility location, treatment facility type, treatment selection criteria, tumor grade, tumor histology, tumor markers, tumor size, and year of diagnosis. In addition, the iDAG includes unobserved confounders of the extracted studies as adjusted to determine the back door paths, if any, relevant to those variables in the current study context. The rationale is similar to that of the prior component for identifying mediators and colliders if misspecified as confounders in individual studies.

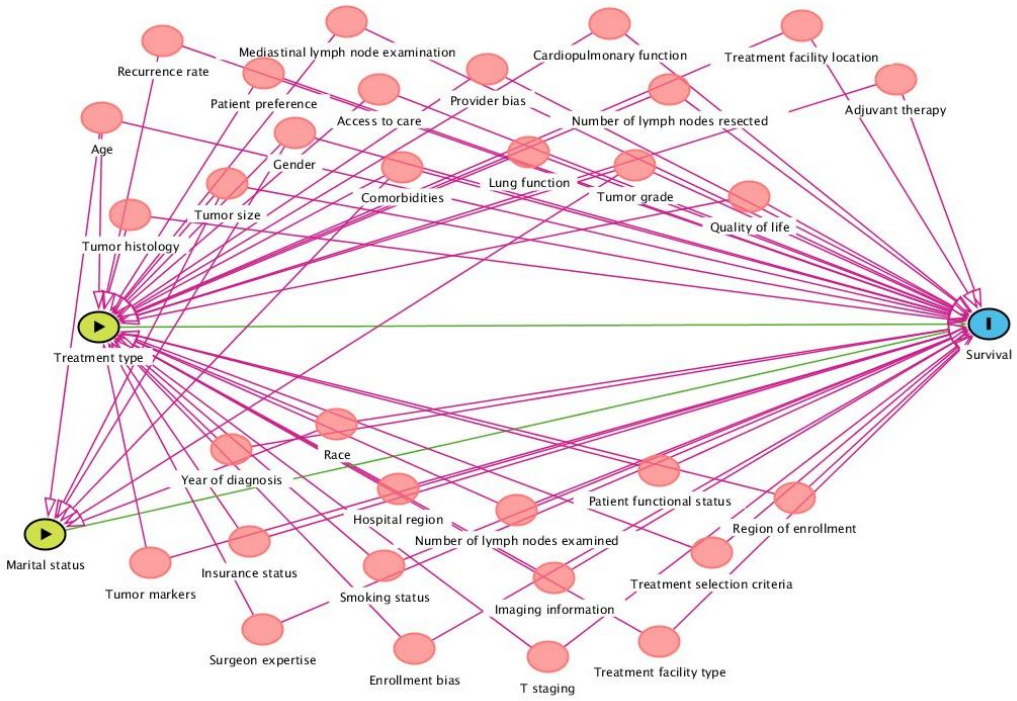


Figure 2.2a Integrated DAG for factors affecting survival outcomes.

The following conditional independence testable implications were identified by DAGitty diagnostics for the total effect adjustment of the developed iDAG. In other words, the following must be true in a statistical analysis model, given the effect of exposure on the outcome and multicollinearity.

Access to care was not related to enrolment bias, imaging information, lung function, marital status, mediastinal lymph node examination, number of lymph nodes examined, number of lymph nodes resected, patient functional status, patient preference, provider bias, quality of life, recurrence rate, region of enrolment, smoking status, surgeon expertise, T staging, treatment facility type, adjuvant therapy, treatment selection criteria, tumor grade, tumor histology, tumor markers, tumor size, year of diagnosis, age, comorbidities, sex, race, and cardiopulmonary function.

Adjuvant therapy was not related to cardiopulmonary function, enrolment bias, hospital region, imaging information, insurance status, lung function, marital status, mediastinal lymph node examination, number of lymph nodes resected, functional status, patient preference, provider bias, quality of life, recurrence rate, region of enrolment, smoking status, surgeon expertise, T staging, treatment facility location, treatment facility type, treatment selection criteria, tumor grade, tumor histology, tumor markers, tumor size, year of diagnosis, age, comorbidities, sex, and race.

The cardiopulmonary function was not related to enrolment bias, hospital region, imaging information, insurance status, lung function, marital status, mediastinal lymph node examination, number of lymph nodes examined, number of lymph nodes resected, patient preference, provider bias, quality of life, recurrence rate, region of enrolment, smoking status, surgeon expertise, T staging, treatment facility location, treatment facility type, tumor grade, tumor histology, tumor markers, tumor size, year of diagnosis, age, comorbidities, sex, and race.

Enrolment bias was not related to hospital region, imaging information, insurance status, lung function, marital status, mediastinal lymph node examination, number of lymph nodes examined, number of lymph nodes resected, patient functional status, patient preference, provider bias, quality of life, recurrence rate, region of enrolment, smoking status, surgeon expertise, T staging, treatment facility location, treatment facility type, treatment selection criteria, tumor grade, tumor histology, tumor markers, tumor size, year of diagnosis, age, comorbidities, sex, and race.

The hospital region was not related to imaging information, insurance status, lung function, marital status, mediastinal lymph node examination, number of lymph nodes examined, number of lymph nodes resected, patient functional status, patient preference, provider bias, quality of life, recurrence rate, region of enrolment, smoking

status, surgeon expertise, T staging, treatment facility location, treatment facility type, treatment selection criteria, tumor grade, tumor histology, tumor markers, tumor size, year of diagnosis, age, comorbidities, sex, or race.

Imaging information was not related to insurance status, lung function, marital status, mediastinal lymph node examination, number of lymph nodes examined, number of lymph nodes resected, patient functional status, patient preference, provider bias, quality of life, recurrence rate, region of enrolment, smoking status, surgeon expertise, T staging, treatment facility location, treatment facility type, tumor grade, tumor histology, tumor markers, year of diagnosis, age, comorbidities, sex, or race.

Insurance status was not related to lung function, marital status, mediastinal lymph node examination, number of lymph nodes examined, number of lymph nodes resected, patient functional status, patient preference, provider bias, quality of life, recurrence rate, region of enrolment, smoking status, surgeon expertise, T staging, treatment facility location, treatment facility type, treatment selection criteria, tumor grade, tumor histology, tumor markers, tumor size, year of diagnosis, comorbidities, sex, or race.

Lung function was not related to marital status, mediastinal lymph node examination, number of lymph nodes examined, number of lymph nodes resected, patient functional status, patient preference, provider bias, quality of life, recurrence rate, region of enrolment, smoking status, surgeon expertise, T staging, treatment facility location, treatment facility type, treatment selection criteria, tumor grade, tumor histology, tumor markers, tumor size, year of diagnosis, age, comorbidities, sex, or race.

Marital status was not related to mediastinal lymph node examination, number of lymph nodes examined, number of lymph nodes resected, patient functional status, patient preference, provider bias, quality of life, recurrence rate, region of enrolment, smoking status, surgeon expertise, T staging, treatment facility location, treatment

selection criteria, tumor histology, tumor markers, and year of diagnosis. Adjustments for age, comorbidities, sex, race, tumor grade, tumor size, and marital status were not related to treatment type.

Mediastinal lymph node examination was not related to patient functional status, patient preference, provider bias, quality of life, recurrence rate, region of enrolment, smoking status, surgeon expertise, T staging, treatment facility location, treatment facility type, treatment selection criteria, tumor grade, tumor histology, tumor markers, tumor size, year of diagnosis, age, comorbidities, sex, or race.

The number of lymph nodes examined was not related to the number of lymph nodes resected, patient functional status, patient preference, provider bias, quality of life, recurrence rate, region of enrolment, smoking status, surgeon expertise, T staging, treatment facility location, treatment facility type, treatment selection criteria, tumor grade, tumor histology, tumor markers, tumor size, year of diagnosis, age, comorbidities, sex, and race.

The number of lymph nodes resected was not related to patient functional status, patient preference, provider bias, quality of life, recurrence rate, region of enrolment, smoking status, surgeon expertise, T staging, treatment facility location, treatment facility type, treatment selection criteria, tumor grade, tumor histology, tumor markers, tumor size, year of diagnosis, age, comorbidities, sex, and race.

Patient functional status was not related to patient preference, provider bias, recurrence rate, region of enrolment, smoking status, surgeon expertise, T staging, treatment facility location, treatment facility type, tumor grade, tumor histology, tumor markers, tumor size, year of diagnosis, age, sex, or race.

Patient preference was not related to provider bias, quality of life, recurrence rate, region of enrolment, smoking status, surgeon expertise, T staging, treatment facility

location, treatment facility type, treatment selection criteria, tumor grade, tumor histology, tumor markers, tumor size, year of diagnosis, age, comorbidities, sex, or race.

Provider bias was not related to the quality of life, recurrence rate, region of enrolment, smoking status, surgeon expertise, T staging, treatment facility location, treatment facility type, treatment selection criteria, tumor grade, tumor histology, tumor marker, tumor size, year of diagnosis, age, comorbidities, sex, and race.

Quality of life was not associated with recurrence rate, region of enrolment, smoking status, surgeon expertise, T staging, treatment facility location, treatment facility type, treatment selection criteria, tumor grade, tumor histology, tumor markers, tumor size, year of diagnosis, age, sex, or race.

The recurrence rate was not related to the region of enrolment, smoking status, surgeon expertise, T staging, treatment facility location, treatment facility type, treatment selection criteria, tumor grade, tumor histology, tumor markers, tumor size, year of diagnosis, age, comorbidities, sex, or race.

Region of enrolment

The region of enrolment was not related to smoking status, surgeon expertise, T staging, treatment facility location, treatment facility type, treatment selection criteria, tumor grade, tumor histology, tumor markers, tumor size, year of diagnosis, age, comorbidities, sex, or race.

Smoking status was unrelated to surgeon expertise, T staging, treatment facility location, treatment facility type, treatment selection criteria, tumor grade, tumor histology, tumor markers, tumor size, year of diagnosis, age, comorbidities, sex, or race.

Surgeon expertise was not related to T staging, treatment facility location, treatment facility type, treatment selection criteria, tumor grade, tumor histology, tumor markers, tumor size, year of diagnosis, age, comorbidities, sex, and race.

The treatment facility type was unrelated to the treatment selection criteria, tumor grade, tumor histology, tumor markers, tumor size, year of diagnosis, age, comorbidities, sex, and race.

Tumor grade was unrelated to tumor histology, tumor markers, tumor size, year of diagnosis, age, comorbidities, sex, and race. Tumor histology is unrelated to tumor markers, tumor size, year of diagnosis, age, comorbidities, sex, or race. Tumor markers were unrelated to tumor size, year of diagnosis, age, comorbidities, sex, and race. The tumor size was not related to the year of diagnosis, age, comorbidities, sex, or race. The year of diagnosis was not associated with age, comorbidities, sex, or race. Age was not related to comorbidities, sex, or race. Sex is not related to race, and comorbidities are not related to sex or race.

2.3.3.3 PROPOSED DAG DEVELOPED FOR CHAPTERS TWO AND THREE INFORMED BY CORRESPONDING IDAGS

The iDAG (Figure 2.2a) for factors affecting treatment received informed the development of the proposed DAG (Figure 2.2b), which supplemented Chapter Two statistical modeling. Similarly, iDAG for factors affecting OS (Figure 2.3a) informed the development of the proposed DAG (Figure 2.3b), which supplemented the Chapter Three statistical modeling. Several factors affecting lobectomy receipt were identified, such as treatment guideline revision years, race, doctor recommendation, patient preference, rural-urban continuum, marital status, tumor size, tumor grade, age, sex, and insurance status.

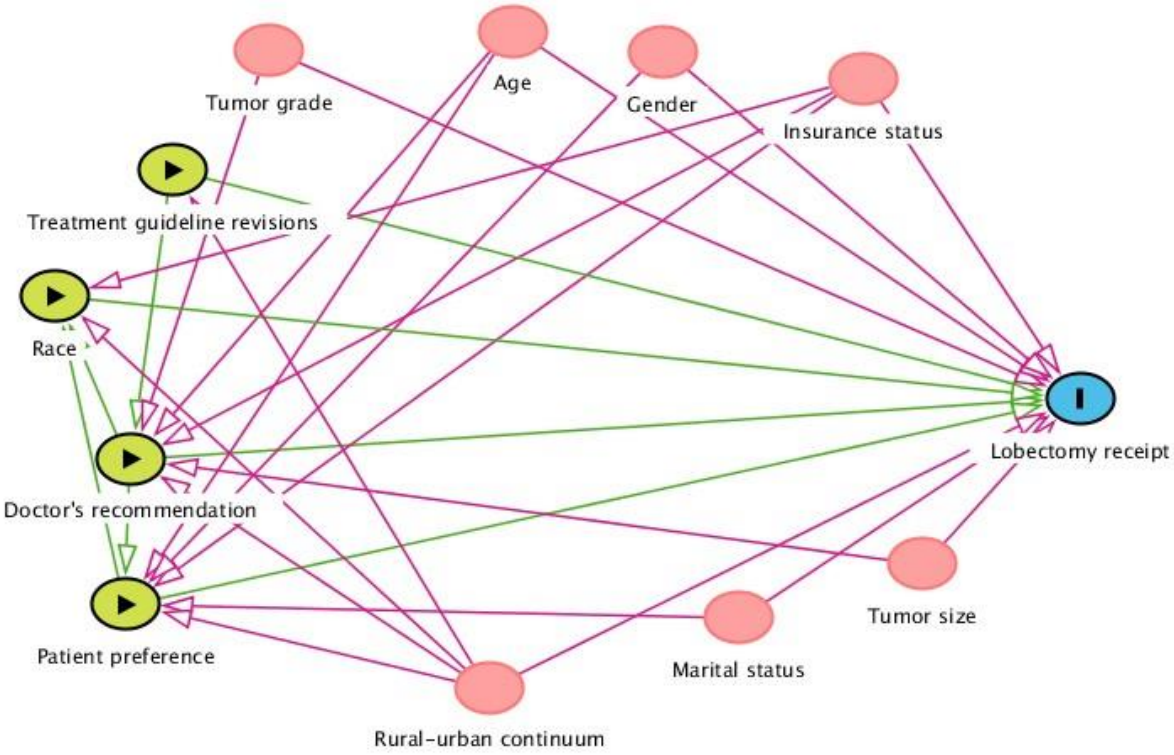


Figure 2.3b Proposed DAG for factors affecting treatment receipt.

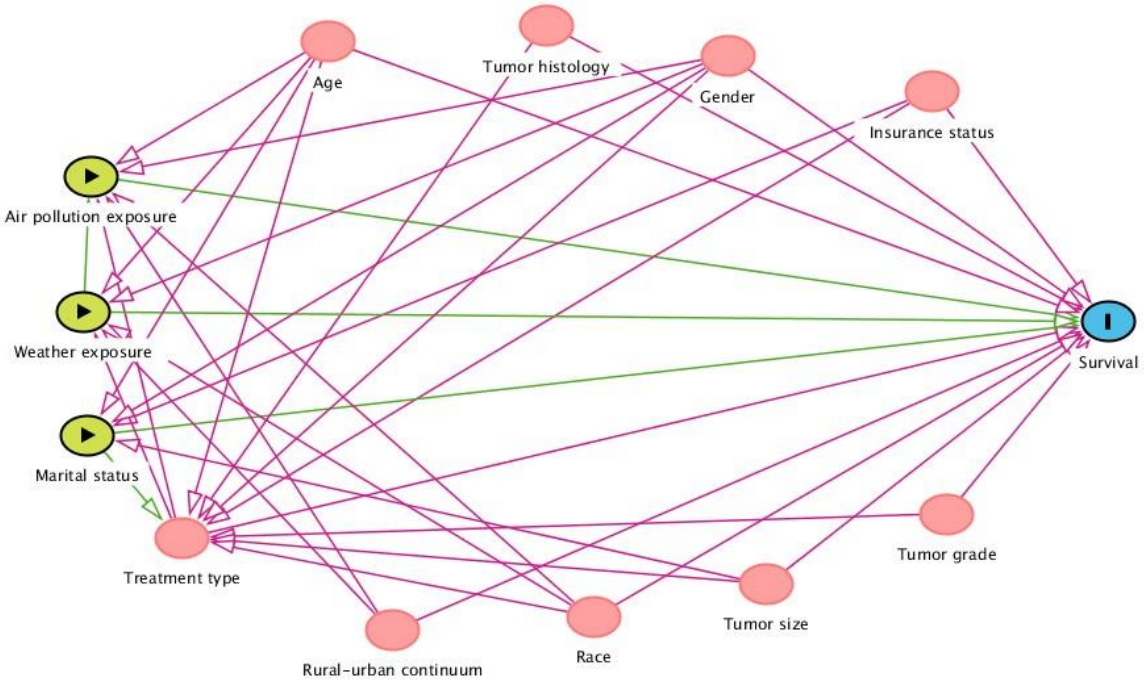


Figure 2.4b Proposed DAG for factors affecting survival outcomes.

Similarly, several factors were identified that affect survival outcomes, such as air pollution, weather exposure, marital status, treatment type, rural-urban continuum, race, tumor size, tumor grade, age, sex, tumor histology, and insurance status.

There are a total of seven covariates and 12 causal paths between doctor's recommendation, patient preference, treatment guideline revision years, race exposure variables, and lobectomy receipt outcome variable. The following conditional independence testable implications were identified by the DAGitty diagnostics for the total effect adjustment of the proposed DAG (Figure 2.2b). In other words, the following must be true in a statistical analysis model, given the effect of exposure on the outcome and multicollinearity. Treatment guideline revisions were not related to tumor grade, race, or sex. Doctors' recommendations were not related to marital status

or sex. Patient preference was not related to treatment guideline revisions, tumor grade, or tumor size. Insurance status is not related to treatment guideline revisions, tumor grade, tumor size, or the rural-urban continuum. Marital status was not related to the rural-urban continuum, treatment guideline revisions, tumor grade, tumor size, sex, or race. The rural-urban continuum was not related to tumor grade, size, age, or sex. The tumor grade was not related to tumor size, age, sex, or race. Tumor size was not associated with age, sex, or ethnicity. Age was not related to sex or race, whereas sex was not related to race.

Nine covariates and eight causal paths were identified between air pollution, weather exposure, marital status exposure variables, and survival outcome variables. The following conditional independence testable implications were identified by the DAGitty diagnostics for the total effect adjustment of the proposed DAG (Figure 2.3b). In other words, the following must be true in a statistical analysis model, given the effect of exposure on the outcome. Adjusting for age, sex, race, and treatment type, air pollution exposure was not related to insurance status, marital status, tumor grade, or tumor size. Insurance status was not related to the rural-urban continuum, tumor grade, tumor histology, tumor size, or race. After adjusting for age, sex, race, and treatment type, insurance status was not related to weather exposure. Marital status was not related to the rural-urban continuum, tumor grade, tumor histology, or race. After adjusting for age, sex, race, and treatment type, marital status was not related to weather exposure. The rural-urban continuum was not related to tumor grade, tumor histology, tumor size, age, or sex. The tumor grade was not related to tumor histology, age, sex, race, or tumor size. Adjustment for age, race, sex, treatment type, and tumor grade was not related to weather exposure. Tumor histology was not related to tumor size, age, sex, or race. After adjusting for age, sex, race, and treatment type, tumor histology was

not related to weather exposure. Tumor size was not associated with age, sex, or ethnicity. Age was not associated with sex or race. Finally, sex was not found to be related to race.

2.4 CONCLUSION

Eight studies provided information on the factors affecting treatment receipt. In comparison, 28 studies provided information on the factors affecting survival outcomes. The factors that affect treatment receipt are age, comorbidity score, education, sex, income, insurance status, marital status, patient preference, physician's preference, tumor size, geographic location, and treatment facility type, which aligns with the existing literature. Therefore, adjusting for these factors in a regression model can help improve the prediction abilities of the model in determining the direct effect on the treatment receipt outcome variable.

The factors that affected survival outcomes were access to care, adjuvant therapy, age, cardiopulmonary function, comorbidities, enrolment bias, sex, hospital region, imaging information, insurance status, lung function, mediastinal lymph node examination, number of lymph nodes examined, number of lymph nodes resected, patient functional status, patient preference, provider bias, quality of life, race, recurrence rate, region of enrolment, smoking status, surgeon expertise, T staging, treatment facility location, treatment facility type, treatment selection criteria, tumor grade, tumor histology, tumor markers, tumor size, and year of diagnosis, which can help determine the total as well as the direct effect of exposure variables on the outcome variable.

The integrated DAGs developed in this study might serve as a supplement to inform statistical modeling decision-making for including covariates in future studies to

determine the factors affecting treatment receipt and survival outcomes among patients with stage 1A NSCLC TN0M0. The results of this study are not a substitute for other relevant regression diagnostics, such as correlations or multicollinearity. DAG is a graphical representation that helps identify all possible backdoor pathways to evaluate the total effect of exposures on outcome variables. Several factors, such as sample size, time trend, statistical modeling, composition of the study sample, sample selection bias, age group, type of data, study design, and type of intervention, contribute to the significance of confounders in the study context. Further implication testing can be carried out by future studies through statistical analysis to determine the effect of significance in a given study setting using meta-analysis and regression.

CHAPTER 3: FACTORS THAT AFFECT TREATMENT RECEIPT FOR STAGE 1A NON-SMALL CELL LUNG CANCER PATIENTS IN THE UNITED STATES

3.1 INTRODUCTION

The treatment modality for stage 1A TN0M0 Non-Small Cell Lung Cancer (NSCLC) in the United States (U.S.) is surgery for medically fit patients, while radiotherapy is for medically unfit patients.[68] Stage 1A TN0M0 NSCLC has more efficacious treatment modalities than more advanced stages.[40] For medically fit patients, lobectomy or limited resection (segmentectomy or wedge resection) with adjuvant radiotherapy. Those who are frail, or otherwise medically unfit for surgical procedure(s), usually receive radiotherapy treatment.[34] Differences in receiving curative surgical treatment according to guidelines exist by race, facility type, and geography.[69] Socioeconomic variation in receipt of curative intent treatment for this stage are not yet published. Notably, there are studies attempting to determine lack of guideline-concordant treatment stratified by race but do not determine it in presence of key identified confounders (*i.e.,* treatment guideline revision years, patient preference, doctor's recommendation, year of diagnosis and county specific time invariant unobservables).[35,39]

Early-stage lung cancer morbidity and mortality are largely preventable, but barriers to cancer care exist.[36] However, no published empirical studies to identify barriers to lung cancer treatment through the application of relevant theoretical models Andersen and Aday[1], and statistical model decision-making. The present study is the first analysis of barriers to cancer treatment in the Anderson model. State-level analysis is done to determine rate of receipt of curative intent surgery survival in masking the

local-level access to care problem identification and variation in treatment receipt.[69] Studies in the past have attempted to identify the associational relationship in the study context without accounting for the clustering of standard errors at the state level, year of diagnosis, and time-invariant unobserved effects for analysis performed at the county level. Those studies do not account for key confounders treatment guideline revision years, patient preference, and physician recommendations leading to biased estimates. In this study, Area Health Resource Files (AHRF) were merged with National Cancer Registry lung cancer patient data for analysis of the variation and difference in treatment receipt in the U.S. according to racial, type of medical insurance coverage, and geographical region difference factors.

In Chapter One, factors were identified that affect receipt of treatment for stage 1A NSCLC TN0M0 through an integrated Directed Acyclic Graph (iDAG). Factors affecting treatment receipt (Figure 2.2a) were disability status, type of treatment facility, geographic location, and race.[40,70–73] Significant observed confounders in the causal model were age at diagnosis, smoking status, coronary artery disease, Chronic Obstructive Pulmonary Disease (COPD), tumor histology (i.e., adenocarcinoma, squamous cell carcinoma, adenosquamous carcinoma etc.), diagnosis after screening,[70] race, sex, marital status, age, income, insurance status, health status, treatment type,[70,71] and geographic area of residence.[40,73] Unobserved confounders identified by these studies included physician preference, patient preference, treatment facility information, specialist availability in the area, frailty, comorbidities, lung, and overall health functional status.[34,37,40]

We proposed key factors missing from previous literature for the iDAG (Figure 2.2b) and attempted to account for these factors in our statistical modeling in this study to identify factors that affectreceipt of treatment. The iDAG (Figure 2.2b) identified

receipt of treatment receipt factors (*e.g.*, treatment guideline revision years, patient preference, doctor recommendation, race, tumor grade, age, sex, insurance status, tumor size, marital status, and rural-urban continuum) that affect whether or not a lobectomy is appropriate for stage 1A NSCLC TN0M0 first primary patients in the U.S..

The current study aimed to identify factors associated with treatment receipt among stage 1A TN0M0 first primary NSCLC by utilizing US population-based cancer data and AHRF. We attempted to answer the following questions: Is there a difference in treatment receipt among patients with varying demographics and tumor characteristics? What determines the type of treatment received? Which type of treatment has higher/lower odds of death, adjusting for relevant enabling, predisposing, and environmental factors? We hypothesized a difference in the type of treatment received among the US population with first primary stage 1A TN0M0 NSCLC by SDOH and local resource status.[19,20] An empirical gap in the literature exists regarding unobserved factors. The Surveillance, Epidemiology, and End Results (SEER) website states that critical biases exist in "who receives the treatment," including patient preference, physician recommendations, comorbidities, and proximity to treatment providers. The current study will fill some of these gaps by quatifying key confounders (e.g., patient preference, doctor recommendations) to determine factors that affect treatment receipt. The doctor recommendation variable followed the treatment decision logic model. Since the treatment guidelines are clear for this stage and histology of lung cancer, we assume that a doctor's recommendation for a certain type of treatment is solely based on the health status, clinical characteristics, and comorbidity score.

Specialist availability, such as the pulmonary specialist Doctor of Medicine (MD), thoracic surgeon, and oncology radiologist variables, were obtained by merging

historical AHRF data to account for additional unobserved confounders identified in the literature.[37] Hence, combining SEER and AHRF data accounts for patient- and provider-level factors. Finally, there is a theoretical gap, as prior literature specific to stage 1A NSCLC TN0M0 lacks utilization of the widely accepted Andersen and Aday[1] health services research behavioral conceptual framework. The DAG developed guided by the theoretical framework was not used to inform statistical model decision-making in prior studies either. It is vital to account for unobserved confounders to identify factors affecting treatment receipt for stage 1A NSCLC. This can help inform treatment decision-making and policies to improve overall survival (OS) by addressing targeted difference factors that are one of the deliverables in this dissertation.

3.2 METHODS

3.2.1 STUDY DESIGN AND RESEARCH APPROACH

3.2.1.1 STUDY DESIGN

This was an observational, retrospective cohort study.

3.2.1.2 DATA SOURCES AND COLLECTION

The SEER 18 Research Plus data access request was approved on 04/18/2022, with reference number SAR0028589, to access the data through the SEER*Stat account. AHRF files are publicly available data, and the website from which they were retrieved is explained in a later section of this paper. The SEER 18 Research Plus and AHRF were used from 1988 to 2016. The construction of the data file for the final

analysis is illustrated in Figure 3.1. The Surveillance Research Program (SRP) of the NCI's Division of Cancer Control and Population Sciences (DCCPS) supports SEER.

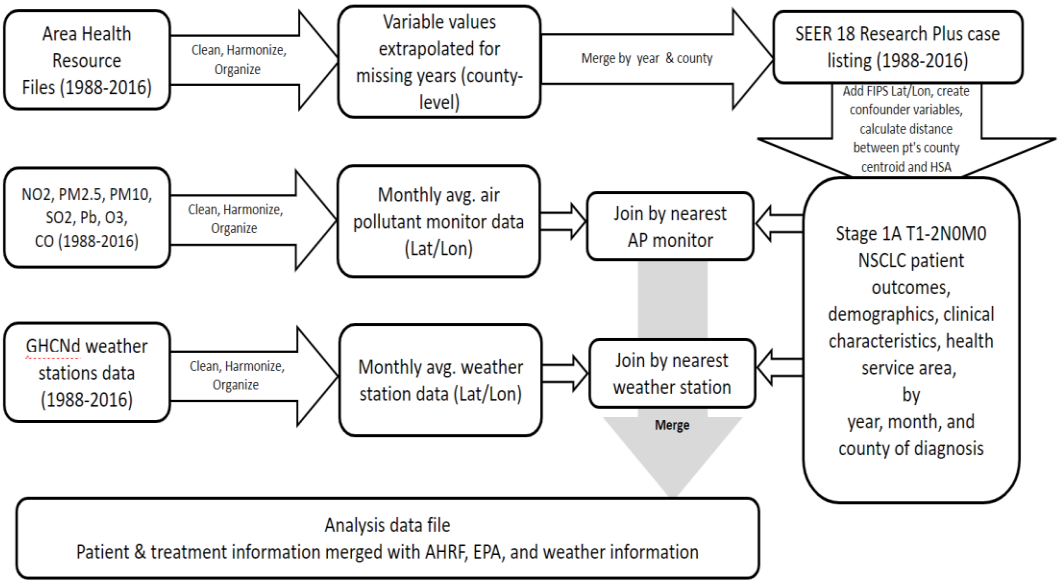


Figure 3.1 Data analysis file construction

SEER collects and publishes cancer incidence and survival data for every cancer case reported from 22 U.S. geographic areas from population-based cancer registries, covering approximately 48% of the U.S. population. Registries routinely collect data on patient demographics, primary tumor site, tumor morphology and stage at diagnosis, the first course of treatment, and follow-up for vital status (survival).[74] The AHRF collects its data from over 50 national sources. The data are all at the county level or are aggregated up to the county level. In addition, data were collected by the Health Resources Service Administration's (HRSA) Bureau of Health Professions in each of the nation's counties.[75] AHRF files were converted from software-independent archival files to software-dependent files and later cleaned before merging with SEER data. In STATA, several steps were taken to output the final cleaned files.

1. The Excel files generated in Python were renamed for their column names, and the variables were labelled according to the field names mentioned in the AHRF technical documentation.
2. The files generated in the previous step were used for each year of data-file creation. Variables that were not required in the study were dropped, and variables were renamed to reflect the type of data captured. Social Determinants Of Health (SDOH) beta files by the Agency for Health Research and Quality (AHRQ) have categorized variables into the following main domain context: social, economic, education, physical infrastructure, healthcare, and geography.[76] In this step, the data variables required for the study analysis were categorized into respective domains.
3. Keeping county Federal Information Processing Standard (FIPS) as a merging criterion, each domain and each year file was utilized to extract same-year variables under each domain. Finally, similar-year variables were merged to generate a master yearly file corresponding to each domain. The AHRF files for each year have multiple-year data that do not necessarily correspond to the year the data file belongs to for each variable. Therefore, in this step, we separated these multiple-year variables into their corresponding yearly files. Finally, using county FIPS as the key variable, we merged byyear variables and produced a master file that was appropriately time stamped with the year to which each variable belongs.

Under each domain, the yearly variables extracted in the previous step were merged. For the variables absent in that particular year, a missing value was generated such that at the end, when all-year files were merged, there were no unmatched observations. Finally, each domain for each year was merged using county FIPS codes as merge key.

3.2.1.3 EMPIRICAL MODEL AND DATA ANALYSIS SOFTWARE

Data in the final analysis file with complete information on variables supplemented by the proposed iDAG in Chapter One were used in the final and preliminary analyses. STATA 16,[77] Microsoft Excel and Pycharm for Python were used for data analysis and file construction. Surgery is the first recommendation for medically fit patients among guideline-concordant care followed by radiotherapy for medically unfit patients. Surgery could be either lobectomy, limited resection, or limited resection with adjuvant radiotherapy for unclear margins. Of the surgeries, lobectomy is the first recommendation for medically fit patients. However, patients who are white and have insurance coverage other than Medicaid (i.e., Medicare, or any private) are more likely to undergo lobectomy than Black and Medicaid beneficiaries. It is clear that difference in lobectomy receipt exists, even though it is the first recommended treatment for medically fit candidates. Lobectomy is the most comprehensive type of surgery that attempts to remove the majority of the tumor-affected region, depending on other clinical and patient characteristics, compared to other less aggressive and non-aggressive surgeries. Therefore, it is important to determine a statistical model for lobectomy. The following empirical model represents the rationale behind the final logistic regression model:

$$\textit{Receipt of Lobectomy}_i\,(0/1) = \beta_0 + \beta_1.\textit{Patient Demographics}_i +$$

$$\beta_2.\textit{Clinical Characteristics}_i + \beta_3.\textit{County}_i + \beta_4.\textit{Year of Diagnosis}_i + e_i$$

where *i* denotes an individual patient. The variable "Receipt of Lobectomy" was a binary variable that took a value of 1 if the patient underwent lobectomy and 0 if the patient underwent limited resection with adjuvant radiotherapy. Other treatment types

were excluded because there were fewer observations within the radiotherapy and limited resection categories. The variable "Receipt of Lobectomy" represents the outcome construct of Andersen and Aday's[1] theoretical framework. The variable "Patient Demographics" represents Andersen and Aday's[1] structure and process level factors such as socioeconomic status, age, race, sex, marital status, access to healthcare, insurance coverage, and geographic region of residence. The variables in the "Clinical Characteristics" category represents Andersen and Aday's[1] health policy, structure, and process level factors, such as treatment guideline revision years, tumor size, tumor grade, comorbidity scores, health status, coronary artery disease, disability status, and other cardiopulmonary diseases. The variables "Year of Diagnosis" and "County" represents all time-invariant unobservable variables in that particular year of diagnosis and county. It represents health policy, structure, and process level constructs of the theoretical model that includes differential treatment-relevant policy implementation at the federal, state, and local levels, also county socio-demographic and healthcare resources.

3.2.1.3.1 ETHICAL CONSIDERATIONS

The University of Louisville (UofL) Institutional Review Board (IRB) approved this study (IRB number 22.0281). The study is exempt according to 45 CFR 46.101(b) under Category 4: Secondary research, for which consent is not required.

3.2.1.4 SAMPLING STRATEGY

3.2.1.4.1 POPULATION AND SAMPLE

The study population comprised stage 1A NSCLC patients as per AJCC 3rd, or 6th, or 7th edition information provided by SEER database with the first primary and

only TN0M0 tumor to form a more homogenous sample, given specific treatment guidelines. The inclusion criteria for the present study included the diagnosis year 1988-2016. Tumor staging information for patients is present in the SEER 18 Research Plus cancer registry but is absent before 1988. The age range for inclusion was 18-80 years as treatment guidelines for this range are homogenous[78,79], microscopic histologic confirmation as confirmed pathologically, histology subtypes (adenocarcinoma, squamous cell carcinoma, or large cell carcinoma) the three main subtypes.[80] Other tumor subtypes had ≤ 1% of observations and were excluded from the analysis. We divided our analysis sample into study and sub-samples. The sub-sample included only patients from the study sample with air pollution and weather exposure non-missing values for the analysis in Chapter Three.

The exclusion criteria for the SEER 18 research plus cancer registry are described in Figure 3.2, along with the number of patients excluded beyond the cancer registry (Figure 3.3). In addition, the pre-regression diagnostics determined several missing values for the variables in the AHRF database, which were >15%. These variables belonged to several domain context variables (*i.e.*, healthcare, physical infrastructure, and education), and the majority of variables from the social context and economics (Table 3.1) were excluded from the final data analysis.

Two analysis samples were created, one for the present study and one for Chapter Three, to determine whether the study estimates were robust or whether the compositional effect persisted. Therefore, there might have been a selection bias for analysis sample one as we specifically focused on certain inclusion criteria.

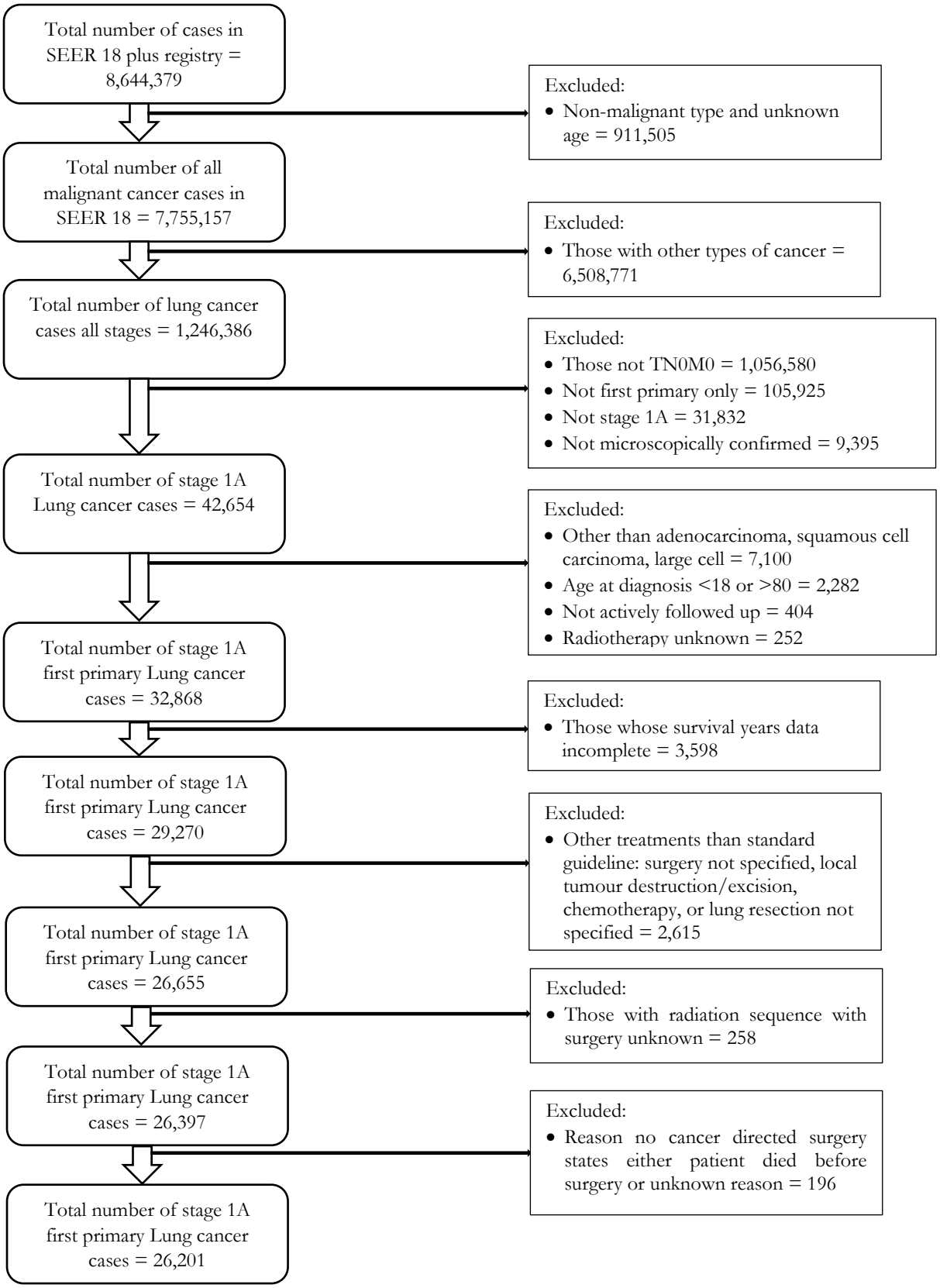


Figure 3.2 SEER 18 Research Plus cancer registry sample selection

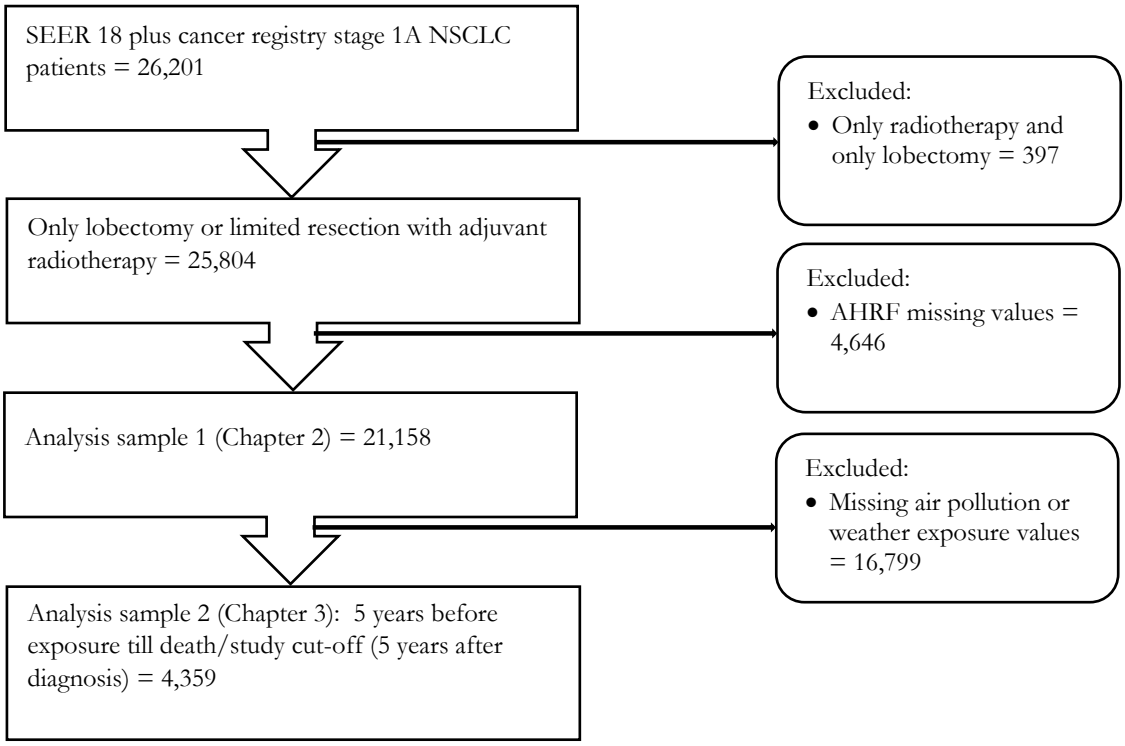


Figure 3.3 Sub-sample selection

Domain Context	Variable names
Geography	Census division code, Census division name, Census regional code, Census regional name, County name, State FIPS code, State and County FIPS code, State name, County FIPS code
Social Context	Median age, Population male/female by 5 years age groups, Number of households, Number of persons in household, Percent Black population, Percent white population, Percent other race population, Population estimates in 100's
Economic	Per Capita Income, Median family income, Percent person below poverty, Percent family below poverty, Unemployment rate
Education	Percent person with high school education, Percent person with 4 years of college degree

Physical infrastructure	Number of occupied housing units, Median gross rent
Healthcare	Number of pulmonary MD's, Number of thoracic surgeon MD's, Number of radiation oncologist MD's, Total number of hospitals, Total number of long-term care hospitals, Total number of chronic disease hospitals, Total number of inpatient days, Total number of hospital beds, Distribution of hospitals as per their utilization rate

Table 3.1 Final Area Health Resource Files (AHRF) variable domains with names

However, the selection of analytical sample two (sub-sample) was beyond the investigator's control, which was solely guided by presence or absence of air pollution and weather monitoring stations within 40 miles of air pollution and 20 miles of distance for weather stations from the centroid of the patient's county of residence. Finally, descriptive statistics for excluded samples were used to determine the generalizability of the results to the specific included this study sample.

There is a difference in characteristics between the excluded sample (Appendix Tables 3.1a and 3.1b) and the included samples (Tables 3.2a and 3.2b). In the excluded sample, the treatment guideline revision year post-2010 is present as AHRF data files with missing 2010 year data, creating a missing value for the AHRF of that year in the data analysis construction file, leading to its exclusion in the analysis sample. This missingness of one treatment revision year in the analysis sample leads to the loss of its effect analysis in the given study context. Moreover, the excluded sample comprised a higher proportion of non-metropolitan areas and Medicaid patients, indicating that one or several of the exclusion criteria might specifically apply to it. This includes the absence of functioning air pollution or weather monitoring stations in non-metropolitan areas and a higher proportion of Medicaid-eligible individuals residing there.

	Study sample		Sub-sample	
N	21,158		4,359	
	Frequency	Percentage	Frequency	Percentage
Tumor size				
Upto 1cm	1,415	6.69	240	5.51
> 1cm & <=2cm	6,691	31.62	1,028	23.58
>2cm	4,700	22.21	832	19.09
Unknown size	8,352	39.47	2,259	51.82
Tumor Grade				
Grade I	4,420	20.89	746	17.11
Grade II	9,041	42.73	1,806	41.43
Grade III	5,846	27.63	1,399	32.09
Grade IV	153	0.72	46	1.06
Unknown	1,698	8.03	362	8.30
Treatment guideline revision years				
Pre 1996 revision	2,274	10.75	747	17.14
Post 1996 revision	8,517	40.25	1,885	43.24
Post 2007 revision	10,367	49.00	1,727	39.62
Rural Urban Continum				
Large central metro	8,022	37.91	2,471	56.69
Large fringe metro	5,840	27.60	1,337	30.67
Medium metro	4,030	19.05	480	11.01
Non-metropolitan	3,266	15.44	71	1.63
Race				
Black	1,859	8.79	516	11.84
White	17,893	84.57	3,532	81.03
Unknown	1,406	6.65	311	7.13
Sex				
Female	11,744	55.51	2,195	50.36
Male	9,414	44.49	2,164	49.64
Insurance type				
Only Medicaid	990	4.68	160	3.67
Only Medicare	5,942	28.08	989	22.69
Only Private	3,142	14.85	537	12.32
Uninsured	150	0.71	22	0.50
Unknown	10,934	51.68	2,651	60.82
Marital Status				
Married	12,482	58.99	2,519	57.79
Widowed	3,081	14.56	657	15.07
Divorced	2,671	12.62	531	12.18
Single	2,222	10.50	502	11.52
Unknown	702	3.32	150	3.44
Doctor's recommendation				
Didn't recommend surgery	-	0.00	-	0.00
Recommend surgery	21,158	100.00	4,359	100.00
Patient preference				
Patient refused doctor's recommendation	-	0.00	-	0.00
Patient prefered doctor's recommendation	21,158	100.00	4,359	100.00

Table 3.2*a* Frequency statistics of study sample and sub-sample

	Study sample			Sub-sample		
N	21,158			4,359		
	Median	Mean	SD	Median	Mean	SD
Age at diagnosis	67	65.81	9.24	68	67.07	8.86
County-level charactristics						
Population estimates	787,698	1,767,143	2,784,384	899,593	2,217,825	2,893,918
Unemployment rate	56	57.78	33.00	55	56.27	30.85
Percapita income	36127	38677.91	13866.60	37362	40363.90	14849.96
Total # hospitals	10	23.18	37.70	14	29.88	41.95
Total # hospital beds	2197	5155.60	8312.50	3413	6676.36	8935.01

Table 3.2*b* Descriptive statistics of study sample and sub-sample

Lack of guideline-concordant care and increased receipt of other treatment types (*i.e.*, local tumor excision, chemotherapy, and lung resection) or a higher proportion of incomplete data in registry reporting to SEER from non-metropolitan areas. Moreover, similar to "only radiotherapy" and "only limited resection groups", the "lung resection not specified" only included 14 observations which is too low of a category power to be included in the statistical analysis as a separate category or merge with another treatment category. The difference in mean survival years might be due to the restriction of the study period to 10 years in analysis sample two.

3.2.1.5 STUDY VARIABLES

3.2.1.5.1 OUTCOME VARIABLE

The lobectomy treatment variable was dummy coded as 0 (if a patient did not receive lobectomy), 1 (if the patient received lobectomy). We excluded observations with only radiotherapy or limited resection, as there were only a few (<1%) observations compared with lobectomy and limited resection with adjuvant radiotherapy.

3.2.1.5.2 PREDICTOR VARIABLES

A paper by Kann et al. 2020 indicates NSCLC treatment guideline revision publication release years in its Figure 1 of the page volumes by cancer type from 1996 to 2019.[81] The National Comprehensive Cancer Network (NCCN) NSCLC treatment guidelines were revised and issued in 1996, 2005, 2006, 2007, 2010, 2012, 2013, 2015, 2018, and 2019. A treatment guideline revision year variable was created based on the information available in the paper to account for years of policy changes in the treatment guidelines. The categorical variable, revision years, comprised three mutually exclusive categories. Those diagnosed before 1996 were categorized as the pre-1996 treatment revision years. Those diagnosed between 1996 and 2007 were categorized as post-1996 treatment revision years, and post-2007 for period beyond 2007 revision year. The treatment recommendations that occur at the time of diagnosis depend on the treatment guidelines, as mentioned previously. Hence, the year of diagnosis facilitated the categorization of observations into treatment guideline revision years. The patient's insurance information was only available from 2007; therefore, the insurance status information before 2007 was categorized as unknown. Similarly, tumor size information was only available from 2004; therefore, tumor size information before 2004 was categorized as unknown. The tumor size categories were created from several numeric values per American Cancer Society categories.[82] A conservative approach

was undertaken because insurance status and tumor size variables are important in the present study context, as informed by the DAG in Chapter One. Instead of excluding observations without insurance and tumor size information, we categorized them into unknown categories.

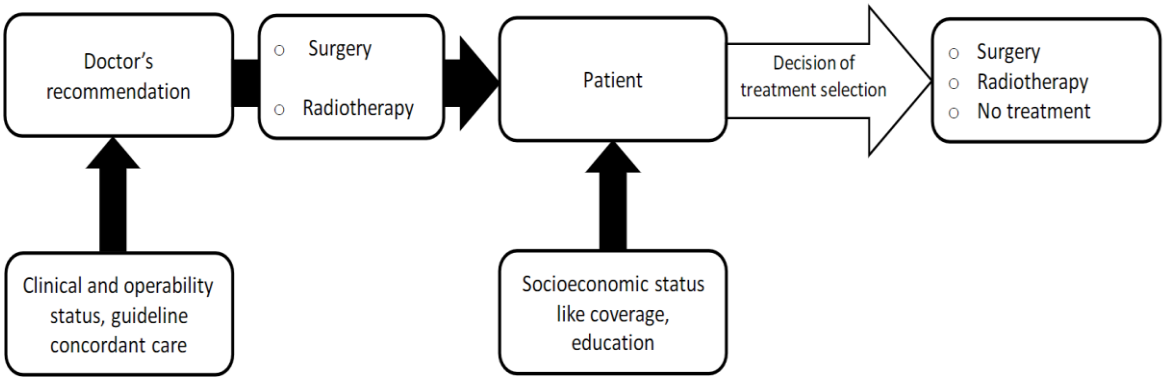


Figure 3.4 Treatment decision-making logic model.

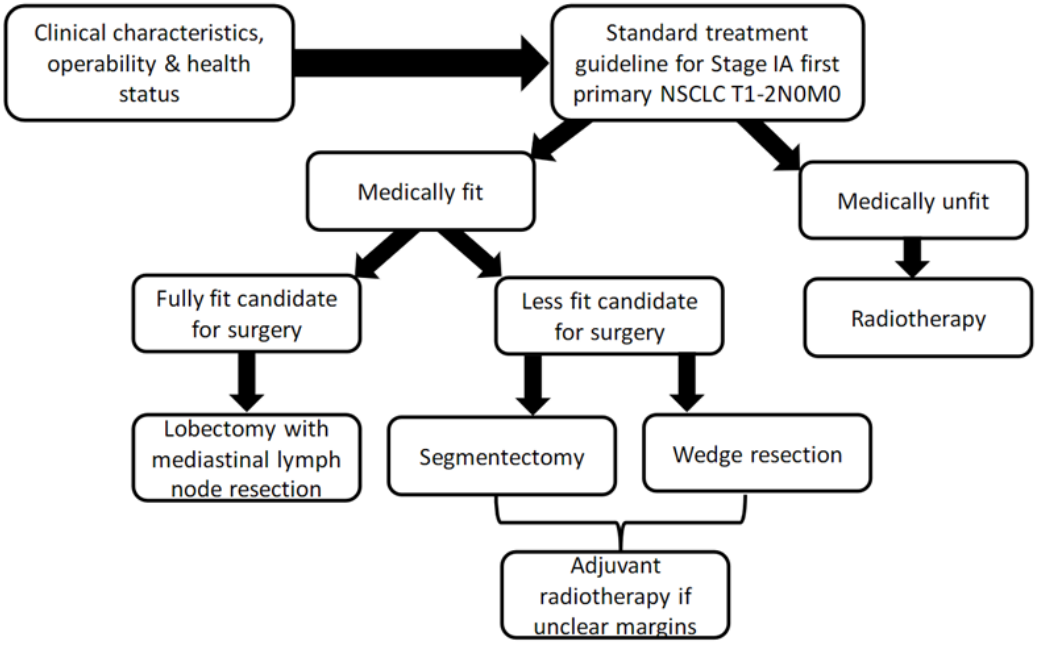


Figure 3.5 Standard treatment guideline for stage 1A NSCLC TN0M0

The doctor's recommendation dummy variable was created using the reason that there was no cancer-directed surgery variable. The patient preference dummy variable was created using the doctor's recommendation variable and the variable treatment information, as illustrated in logic model Figures 3.4 and 3.5. The patient refuses or accepts what is offered through a doctor's recommendation (*i.e.*, for the given lung cancer stage, the doctor would recommend surgery or radiotherapy), depending on the operability status of the patient. The rural-urban continuum categorical variable was created by the U.S. census bureau.[83] Since there were few observations in the small metro, micropolitan, and non-core categories, we combined them into one category titled "non-metropolitan" to adapt to a more conservative approach. Hence, the rural-urban continuum categorical variable comprises four categories: large central metro, large fringe metro, medium metro, and non-metropolitan. Likewise, the unknown and other race categories were combined because the other race categories had few observations. The insurance-type categorical variable was constructed from the original recode variable and patient age. The original variable categories were "Any Medicaid," "Insured/No specific," "Insured," and "Uninsured." The new categorical variable categories are Medicaid, Medicare, private insurance, and the uninsured. Those aged ≤ 64 and belonging to the insured or insured/no specific categories of the original variable were assigned the "only private" category. Those with any Medicaid were renamed Medicaid only, and those aged ≥ 65 years falling under the insured category or insured with no specifics were assigned only the Medicare category. The year of diagnosis dummy variables was constructed to account for the unobservable time-invariant. The other variables informed by the proposed iDAG and present in the final data analysis file with complete information were included.

3.2.2 DATA ANALYSIS METHODS

Ideally, a longitudinal analysis of treatment type receipt would require sufficient variation in outcome variables and their determinants within the county for a panel logistic regression with county-specific and year of diagnosis time-invariant factors with standard errors clustered at the state level to account for unobserved time-invariant confounders. However, overtime variation within the model variables was insufficient for a fixed-effects regression model. Therefore, logistic regression analysis was adjusted for clustering of standard errors at the state level as well as for time-invariant unobserved factors for the year of diagnosis and county. Because the observations were clustered at the county level and counties were clustered within a state, clustering at the state level was utilized. Simultaneously, the time-invariant unobservable for each year of diagnosis and county specificity were accounted for by including their dummy variables in the regression model with standard errors clustered at the state level. Interactions between independent variables were also examined in the preliminary analysis and diagnostics. The diagnostics included the exploration of an alternative relevant regression model. The final regression analysis used the study sample and sub-sample to determine the compositional effect on the model estimates. Finally, post-estimation tests for goodness of fit and predicted probabilities were used to determine the prediction abilities of the employed model.

3.3 RESULTS

The descriptive statistics results determined 21,158 patients in the study sample and 4,359 patients in the sub-sample with non-missing air pollution and weather exposure assignments. The mean age at diagnosis for patients in the study sample was 65 years, whereas that in the sub-sample was 67 years. The median ages of the study sample and

sub-sample were 67 years and 68 years, respectively. Other county-level and patient characteristics are listed in Tables 3.2a and 3.2b.

The compositional differences between the two samples are also presented in Tables 3.2a and 3.2b. In the sub-sample, most of the observations are present in metropolitan areas compared with the study sample, which has approximately 3,266 observations residing in non-metropolitan areas. This indicated higher air pollution and weather monitoring in metropolitan areas than in rural regions. One of the possible reasons might be higher levels of air pollution in urban areas than in rural areas, deeming the installation of more monitors in urban areas compared to rural areas. The study sub-sample comprised approximately 11% Black and majority white racial groups. This might also indicate the higher residential occupation of white racial groups in metropolitan areas compared to their Black counterparts, for whom the descriptive statistics indicate that they tend to live in rural areas. The number of patients with Medicare (23%) and private (12%) insurance tended to be higher in the sub-sample than in Medicaid (4%). This might indicate that a higher proportion of patients residing in metropolitan areas had Medicare or private insurance than those with Medicaid. The difference in composition between the two samples affects their regression estimates, which is further explained in the results section. Hence, the generalizability of the study results highly depends on the regression estimates obtained for that study sample/sub-sample. Furthermore, Chapter Three results depend only on the included sub-sample demographics of the patients.

The odds of Black patients receiving lobectomy was less than their white counterparts by 0.15 points (Table 3.3). The odds of patients receiving lobectomy with tumor grade IV were lower than those with tumor grade II by 0.37 points. The odds of widowed individuals receiving lobectomy were less than those of married by 0.09 points. The

odds of patients with tumor sizes up to 1 cm who underwent lobectomy were less than those with tumor sizes of > 1 cm and ≤ 2 cm by 0.44 points. In contrast, the odds of patients with a tumor size > 2 cm who underwent lobectomy were higher than those with tumor sizes > 1 cm and ≤ 2 cm by 0.96 points. The odds of receiving lobectomy for an increase in patients' age every year than younger patients was less by 0.02 points. The odds of receiving lobectomy for patients with Medicaid insurance than those with Medicare alone was less by 0.17 points. The odds of patients receiving lobectomy for being treated during post-2007, treatment guideline revisions was less than pre-1996 by 0.64 points. Finally, the odds of patients receiving lobectomy who reside in large fringe metro areas were less than their counterparts living in large central metro areas by 0.25 points. Additionally, odds of patients receiving lobectomy and living in the medium metro and non-metropolitan was more than their large central metro counterparts by 0.31 and 0.92 points, respectively.

The likelihood of receiving lobectomy was smaller among Black patients than among white patients by about 2% points (Table 3.3.). The likelihood of receiving lobectomy was smaller for Medicaid beneficiaries than for their Medicare counterparts by about 2% points. The likelihood of receiving lobectomy was smaller for widowed patients than for married patients by about 1% points. The likelihood of receiving lobectomy was smaller for every year increase in age than for younger patients by about 2% points. Similarly, the likelihood of receiving lobectomy during post-treatment guideline revision years 2007 was smaller than pre 1996 by about 10% points respectively.

Study sample				
	Odds (95% CI)		Marginal effects (95% CI)	
Tumor size (reference: > 1cm & <=2 cm)				
Upto 1cm	0.56***	(0.49 , 0.66)	-0.06***	(-0.07 , -0.04)
>2cm	1.96***	(1.70 , 2.27)	0.07***	(0.05 , 0.08)
Unknown	1.62	(0.48 , 5.53)	0.05	(-0.08 , 0.18)
Tumor Grade (reference: II)				
Grade III	0.92	(0.82 , 1.03)	-0.01	(-0.02 , 0.01)
Grade IV	0.63***	(0.46 , 0.86)	-0.05***	(-0.08 , -0.02)
Grade I	0.91	(0.76 , 1.07)	-0.01	(-0.03 , 0.01)
Unknown	0.93	(0.83 , 1.03)	-0.01	(-0.02 , 0.01)
Treatment guideline revision years (Reference pre 1996)				
Post 1996 revision	0.43	(0.14 , 1.40)	-0.09	(-0.20 , 0.03)
Post 2007 revision	0.359***	(0.13 , 0.97)	-0.104***	(-0.21 , 0.003)
Rural-Urban continum (reference: Large central metro)				
Large fringe metro	0.75***	(0.70 , 0.79)	-0.03***	(-0.04 , -0.02)
Medium metro	2.31***	(2.22 , 2.40)	0.09***	(0.08 , 0.09)
Non-metropolitan	2.92***	(2.74 , 3.12)	0.11***	(0.10 , 0.12)
Age at diagnosis	0.98***	(0.97 , 0.99)	-0.02***	(-0.03 , 0.02)
Race (reference: White)				
Black	0.85*	(0.70 , 1.03)	-0.02*	(-0.04 , 0.01)
Unknown	1.50***	(1.29 , 1.75)	0.04***	(0.03 , 0.06)
Sex (reference: Female)				
Male	1.05	(0.92 , 1.20)	0.01	(-0.01 , 0.02)
Insurance type (reference: Only Medicare)				
Only Medicaid	0.83**	(0.70 , 0.97)	-0.02**	(-0.04 , 0.02)
Only private	0.98	(0.82 , 1.16)	-0.03	(-0.02 , 0.02)
Uninsured	0.84	(0.50 , 1.39)	-0.02	(-0.07 , 0.03)
Unknown	1.21	(0.91 , 1.61)	0.02	(-0.01 , 0.05)
Marital status (reference: Married)				
Divorced	1.03	(0.88 , 1.19)	0.03	(-0.01 , 0.04)
Single	0.90	(0.78 , 1.03)	-0.01	(-0.02 , 0.01)
Widowed	0.91*	(0.82 , 1.01)	-0.01*	(-0.02 , 0.01)
Unknown	0.92	(0.74 , 1.13)	-0.01	(-0.03 , 0.01)

Model fit		
	Clusters	12
	Wald chi square	1945 (12), p 0.00
	Psuedo R2	0.09
	Correctly classified	88%
	N	21,158

p value: * < 10%, ** < 5%, *** < 1%

Table 3.3 Study sample regression results

The likelihood of receiving lobectomy for patients with tumor grade IV was smaller than that for those with tumor grade II by about 5% points.

The likelihood of receiving lobectomy for patients with tumor sizes up to 1 cm was smaller than that for patients with tumor sizes > 1 cm and ≤ 2 cm by about 6% points. In contrast, the likelihood of receiving lobectomy for patients with tumors > 2 cm in size was more than those with tumor sizes > 1 cm and ≤ 2 cm by about 7% points.

The results of the sub-sample analysis (Table 3.4) are similar in the direction of effect size; however, some of the estimates lose their statistical significance. Among these, marital status, insurance status, race, and tumor grade. The odds of patients receiving lobectomy for being treated during post-1996, and post-2007 treatment guideline revisions was less than pre-1996 by 0.94, and 0.83 points respectively. This might be the case for the pure compositional effect, as described in the descriptive statistics. The odds of patients receiving lobectomy with tumor grade I were lower than those with tumor grade II by 0.17 points were. The odds of patients undergoing lobectomy with tumor sizes up to 1 cm were less than those with tumor sizes > 1 cm and ≤ 2 cm by 0.46 points. In contrast, the odds of patients with a tumor size > 2 cm who underwent lobectomy were higher than those with tumor sizes > 1 cm and ≤ 2 cm by 0.20 points. The odds of receiving lobectomy for an increase in patients' age every year than younger patients was less by 0.02 points. Finally, the odds of patients receiving lobectomy who reside in large fringe metro, and non-metropolitan areas was less than their counterparts living in large central metro areas by 0.47 and 0.45 points, respectively. Additionally, the odds of patients receiving lobectomy and living in the medium metro were greater than those of their large central metro counterparts by 0.63 points.

The likelihood of receiving lobectomy was smaller for every year increase in age than for younger patients by about 1% points. Similarly, the likelihood of receiving lobectomy during post-treatment guideline revision years 1996, 2007 was smaller than

pre-1996 by about 31%, and 19% points, respectively. The likelihood of receiving lobectomy for patients with tumor grade I was smaller than that for those with tumor grade II by about 2% points. The likelihood of receiving lobectomy for patients with tumor sizes up to 1 cm was smaller than that for patients with tumor sizes > 1 cm and ≤ 2 cm by about 7% points. In contrast, the likelihood of receiving lobectomy for patients with > 2 cm tumor size was greater than that for patients with tumor sizes > 1 cm and ≤ 2 cm by about 9% points.

Sub-sample				
	Odds (95% CI)		Marginal effects (95% CI)	
Tumor size (reference: > 1cm & <=2 cm)				
Upto 1cm	0.54***	(0.38 , 0.75)	-0.07***	(-0.10 , -0.03)
>2cm	2.20***	(1.55 , 3.13)	0.09***	(0.05 , 0.12)
Unknown	0.37	(0.03 , 5.30)	-0.11	(-0.39 , 0.18)
Tumor Grade (reference: II)				
Grade III	1.05	(0.90 , 1.23)	0.01	(-0.01 , 0.02)
Grade IV	1.19	(0.49 , 2.88)	0.02	(-0.08 , 0.11)
Grade I	0.83*	(0.67 , 1.03)	-0.02*	(-0.04 , 0.01)
Unknown	1.11	(0.71 , 1.72)	0.01	(-0.04 , 0.06)
Treatment guideline revision years (Reference pre 1996)				
Post 1996 revision	0.06**	(0.01 , 0.67)	-0.31**	(-0.57 , -0.05)
Post 2007 revision	0.17**	(0.03 , 0.96)	-0.19**	(0.04 , -0.38)
Rural-Urban continum (reference: Large central metro)				
Large fringe metro	0.53***	(0.46 , 0.62)	-0.07***	(-0.08 , -0.05)
Medium metro	1.63***	(1.30 , 2.04)	0.05***	(0.03 , 0.08)
Non-metropolitan	0.55***	(0.41 , 0.75)	-0.06***	(-0.10 , -0.03)
Age at diagnosis	0.98***	(0.97 , 0.99)	-0.01***	(-0.02 , 0.02)
Race (reference: White)				
Black	0.81	(0.59 , 1.10)	-0.02	(-0.06 , 0.01)
Unknown	1.47***	(1.14 , 1.88)	0.04***	(0.02 , 0.07)
Sex (reference: Female)				
Male	1.17	(0.93 , 1.45)	0.02	(-0.01 , 0.04)
Insurance type (reference: Only Medicare)				
Only Medicaid	1.13	(0.67 , 1.89)	0.01	(-0.04 , 0.07)
Only private	1.07	(0.85 , 1.35)	0.01	(-0.02 , 0.03)
Uninsured	1.13	(0.43 , 2.99)	0.01	(-0.09 , 0.12)
Unknown	2.42*	(0.91 , 6.40)	0.10*	(-0.01 , 0.20)
Marital status (reference: Married)				
Divorced	1.22	(0.75 , 1.97)	0.02	(-0.03 , 0.07)
Single	0.89	(0.65 , 1.22)	-0.01	(-0.05 , 0.02)
Widowed	0.96	(0.85 , 1.10)	-0.01	(-0.02 , 0.01)
Unknown	0.78	(0.53 , 1.14)	-0.03	(-0.07 , 0.01)

Model fit		
	Clusters	11
	Wald chi square	2252.81 (11), p 0.0
	Psuedo R2	0.10
	Correctly classified	86%
	N	4,359

p value: * < 10%, ** < 5%, *** < 1%

Table 3.4 Regression results for sub-sample

3.4 CONCLUSION

The present study's results corroborate those of previous studies and factors identified in iDAG that aim to determine differences in treatment receipt.[35,43] The results of the study help conclude the study hypothesis that there exists a difference in treatment type received among stage 1A NSCLC patients in the U.S. by socioeconomic determinants of health and local resource status. Although past studies lack adjustment of treatment guideline revision years in their regression analysis, the present study found that patients diagnosed post-2007 were less likely to undergo lobectomy. Advances in medical technologies, diagnostics, and the introduction of Stereotactic Body Radiotherapy (SBRT) as an alternative treatment occurred around the same time (post-2007).[44,84] These advances in treatment guidelines might have resulted in the increased receipt of less invasive surgery, such as SBRT,[39] and the decreased receipt of more invasive surgery, such as lobectomy. Additionally, the present study results also corroborate that Black are less likely to receive guideline-concordant care compared to their counterparts' racial groups.[43]

The present study results help identify factors determining who is more likely to receive guideline-concordant standard treatment. It also determines the presence of difference factors in receiving a specific type of treatment for stage 1A when the set treatment guidelines are used for lobectomy in medically fit patients and limited resection with adjuvant radiotherapy in medically less fit patients.

This study has several strengths, including statistical modelling adjusted for standard errors at the state level, year of diagnosis, and county-level unobserved time-invariant confounders. The study also accounts for key policy information that is believed to affect the decision to receive guideline-concordant care (*i.e.*, treatment guideline revision years). The study model’s prediction ability is approximately 88%,

which might be improved if future studies account for unobserved time-varying patient-level confounders.

This study had inherent SEER 18 Research Plus and AHRF database limitations. The data limitations include a lack of information on treatment timing, dose, and treatment types such as SBRT, Radio Frequency Ablation (RFA), comorbidity status, patient functional status, tumor recurrence status, depth of tumor, nearest health service center where the patient sought care, physician preference, patient preference, and provider-level information. Although we attempted to generate physician and patient preferences, the variations were insufficient. There were zero observations for options listed, as 'doctor did not recommend surgery 'and' patient refused doctor's recommendation. This might have occurred due to the exclusion of radiotherapy and limited resection categories, which did not have sufficient observations. The unaccounted factors that affect the relationship for factors affecting treatment receipt serve as unobserved confounders leading to biased estimates.

Some omitted variables in the study context are coronary artery disease, income, health status, COPD, and comorbidity score. It is essential to determine the direction of the estimated bias if these omitted variables, due to data limitations, are not adjusted in the statistical analysis. The comorbidity score seems to negatively affect the type of treatment received, as per standard treatment guidelines. The higher the comorbidity score is, the lower the patient's fitness for more invasive surgery, current first-line therapy per guideline-concordant care. This indicates that comorbidity status has a negative correlation with the more invasive curative intent surgery (*i.e.,* lobectomy). Similarly, comorbidity score seems to have a negative correlation with doctor's recommendation, as a higher comorbidity score and lower chance of being recommended lobectomy surgery. Hence, the resulting estimates could be positive and

overestimate the effect of doctors' recommendations on lobectomy treatment. One such omitted variable is COPD, which appears to be negatively correlated with the type of treatment received. The compromised lung function due to chronic illness tends to be negatively correlated to the receipt of more invasive surgery like lobectomy as the invasive surgery might present perioperative surgery-associated complications, further contributing to a reduction in lung function. Similarly, COPD tends to be negatively correlated with doctors' recommendations and patient preferences for the same reasons mentioned previously. This is indicative of an overestimation of the relationship and estimates to be positively biased in the presence of omitted variables. Income is another unaccounted-for variable that serves as an omitted variable in the present study context. It tends to be negatively correlated with patient preference and race and positively correlated with guideline-concordant care receipt. This leads to an underestimation of the derived estimates in the study context owing to negative bias. Specialist availability is another key factor affecting contextual study relationships. Specialist availability in an area tends to be positively correlated with receiving guideline-concordant care. Similarly, it tended to be positively correlated with doctors' recommendations for guideline-concordant care. Overall, the derived estimates without adjusting for omitted variables could lead to an overestimation of the effect in the given study context. Conversely, specialist availability tends to be negatively correlated with the rural-urban continuum, as professional opportunities tend to be higher in urban areas than in resource-deprived rural areas. This leads to an underestimation of the derived estimates if the variable is not accounted for in the regression analysis. Finally, smoking status tends to be negatively correlated with doctors' and patients' preferences, as smoking tends to aggravate perioperative complications postoperatively. As per literature evidence smoking is associated with several pulmonary complications (i.e., atelectasis,

pneumonia, hypercapnia, broncho plueral fistula, empyema, chylothorax, hemothorax, pulmonary embolism, lobar gangrene) post-surgery for NSCLC causing higher chances of morbidities.[85,86] Similarly, it is negatively correlated with the receipt of guideline-concordant care such as invasive surgery. Overall, omitting the variable from the analysis could lead to an overestimation of the effect in the study context. Moreover, it is important to consider factors such as treatment timing. Treatment timing after the first diagnosis tended to be negatively associated with the type of treatment received as those with delayed reciept in time to treatment are highly likely to present with higher comorbidity score and disease progression compromising patient's medical fitness to undergo guideline-concordant care (*i.e.*, lobectomy).[87] The more the time between the first diagnosis and treatment received, the less likely they were to receive lobectomy or limited resection due to several factors that affect tumor metastasis. Treatment timing also tends to be negatively correlated with factors such as race and geographic region.[88–90] This might have led to overestimating the effect in the given study context, leading to biased estimates. For the same reason, our study only measured associational relationships because we did not account for these identified unobserved confounders in the analysis, nor was the study design as a randomized control trial or natural experiment.

Finally, the study results are only generalizable to patients with stage 1A NSCLC undergoing either lobectomy or limited resection with adjuvant radiotherapy. According to the standard NCCN guidelines, the treatment types for stage 1A NSCLC TN0M0 are lobectomy and limited resection, or limited resection with adjuvant radiotherapy, for medically fit patients. However, those who are medically unfit for surgery are recommended to undergo radiotherapy. Hence, it is vital to understand what drives the treatment receipt for radiotherapy or limited resection without adjuvant

radiotherapy. Due to the very few observations in the data that underwent either radiotherapy or limited resection without adjuvant radiotherapy, we could not account for these treatment category types in our analysis.

Future studies should adjust for the unobservable confounders identified in the iDAG to determine the robustness of the present results. As mentioned in the previous paragraph, several omitted variables in the study context could lead to biased estimation, and future studies could include those variables in their analysis to reduce estimation bias. Some of the databases that they could use are medical claims data that provide comprehensive information on several identified unaccounted confounders. The claims data provide information on treatment timing, dosage, specific types of treatment, perioperative treatment information and associated complications, and several key patient contextual factors such as smoking status, family history, comorbidity index, disability status, occupation, patient's overall health status, and lung function. Information on patient insurance eligibility could also further aid in employing statistical modelling techniques such as regression discontinuity and instrumental variables that help reduce omitted variable bias effects. Finally, studies that want to build upon existing results might want to include other sub-stages to determine the robustness of the effect from an external validity perspective.

CHAPTER 4: AIR POLLUTION AND WEATHER AFFECT SURVIVAL OUTCOMES AMONG STAGE 1A NON-SMALL CELL LUNG CANCER PATIENTS IN THE UNITED STATES

4.1 INTRODUCTION

Several modifiable Social Determinants of Health (SDOH) of improved lung cancer survival exist beyond smoking cessation. However, specific modifiable determinants still exist that are yet to be investigated.[6] Ambient air pollution is a modifiable determinant of lung cancer survival, affecting incidence and mortality.[91–93] To date, only ten studies have explored the dose-relationship association of ambient air pollution on lung cancer incidence and mortality in the United States (U.S.).[93] Limited research has been performed to determine the effect of ambient air pollution on stage 1A Non-Small Cell Lung Cancer (NSCLC) survival post-diagnosis using U.S. population cancer registry data.[6] It is evident in the literature that air pollutants affect a specific type of lung cancer histology.[93,94] However, only limited studies have focused specifically on histology type and specific clinical stages of lung cancer using U.S. national cancer registry data.[6,92]

Assessing air pollutant exposure at broader vicinity levels, like community, underestimates the dose-response relationship of its exposure on health burden.[95] These errors of inaccurate exposure assignment bias the outcome results. Moreover, the dose-response relationship also varies depending on which neighborhood people live in, as certain areas might have a source of pollutant release compared to farther vicinity. Hence averaging the community-level exposure values and assigning the exposure values erroneously affect study results.[95] Lung cancer survival effect assessment

requires accurate exposure assignment, such as distance from county centroid, at a granular level. Although only one study to date has established a dose-response relationship between localized lung cancer survival and ambient air pollution exposure,[6] that study did not account for weather components that might affect exposure levels within the vicinity. Some studies in the literature identifying the dose-response relationship between ambient air pollution and lung cancer survival utilize interpolation or other data techniques to replace missing pollutant levels.[6,92] However, ambient air pollutant levels differ depending on the vicinity, as explained previously, as well as natural events such as volcanic eruptions and forest fires. The drawback of interpolating or extrapolating missing pollutant values might inherently misclassify exposure assignments due to the absence of relevant information such as natural events and weather components such as snow, precipitation, temperature, and interaction with other pollutants.

Moreover, in all pollutant (multi-pollutant) models, primary pollutants are precursors of secondary pollutants such as ozone, which, if accounted for in the analysis, might provide false estimations, such as ozone being inherently associated with primary pollutants. Changes in weather conditions also facilitate chemical reactions between primary pollutants (NO_2, SO_2, CO, and PM) and other atmospheric chemicals, resulting in secondary pollutant production.[96] The weather components such as temperature maximum are also correlated with air pollutants as the rise in air pollutants aids in the urban heat index phenomenon.[96] Hence, accounting for secondary pollutants such as ozone and weather components such as temperature maximum might provide biased estimation results in a given study context. Therefore, it is vital to understand the effect of primary air pollutants in the presence of weather components,

such as precipitation, snow, and temperature, on the survival outcomes of stage 1A TN0M0 NSCLC.

Finally, utilizing the Survival, Epidemiology, and End Results (SEER) 18 Research Plus cancer registry data to determine lung cancer survival in the presence of air pollution and weather components exposure will provide more heterogeneity in the sample population, which is representative of the national population. However, one limitation of such a registry is the absence of information on several key confounders (*i.e.*, treatment time received, patient migration, and comorbidity score). Hence, studies utilizing cancer registry data inherently have to assign pollutants/weather exposure values to patients from the time of diagnosis to the time of death,[6] irrespective of the time of treatment from diagnosis. Moreover, several factors affect treatment decision-making and survival outcomes for stage 1A NSCLC, including clinical and patient characteristics. Hence, it is crucial to identify whether ambient air pollution has a dose-response effect on lung cancer survival outcomes depending on the type of treatment received, and to our knowledge, only one study has aimed to identify it.[6] However, the study did not account for the dose-response relationship in the presence of weather components in a homogenous sample of stage 1A NSCLC TN0M0. It also did not account for other primary air pollutants such as SO_2 and CO. Finally, the study assigned exposure values from the month of diagnosis to death, not before-diagnosis exposures. This could lead to an absence of accounting for the carry-over effect on health outcomes from before diagnosis exposure. Therefore, we aimed to evaluate whether exposure to certain levels of air pollutants is associated with survival outcomes among patients with stage 1A TN0M0 NSCLC undergoing treatment of choice by utilizing U.S. population-based cancer data and U.S. environmental air pollution data. Does accounting for any key confounders missing in the studies reduce selection bias and provide close-to-true

hazards? For example, do the variables informed by Shi and Steven's[2] theoretical framework for vulnerable populations help to identify real risks? How does treatment choice affect survival outcomes in the presence of exposure to the identified air pollutants? We hypothesize that there exists a difference in all-cause hazards of death among treated people exposed to high versus low air pollution levels.[6]

Environmental exposure seems to be a potential factor affecting cancer survival according to the Canadian cancer risk population health management model[18] and Shi and Steven's[2] general framework model for studying vulnerable populations[2] utilized in Chapter One. According to a search on Google Scholar and PUBMED.GOV, no publications have used Shi and Steven's[2] vulnerable population model for the analysis lung cancer survival outcomes. Using the same search strategy we found that no previously published study joined data from SEER 18 Research Plus data,, U.S. environmental data, U.S. weather data, and historical Area Health Resource Files (AHRF).

4.2 METHODS

4.2.1 STUDY DESIGN AND RESEARCH APPROACH

4.2.1.1 STUDY DESIGN

This retrospective cohort study compared the survival outcomes between patients exposed to higher versus lower air pollution and those receiving different treatment types (*i.e.*, limited resection with adjuvant radiotherapy and lobectomy) in single- and multi-pollutant models. The pollutant model included NO_2, SO_2, and CO, adjusted for precipitation, snow, and daily minimum temperature values in both the single-pollutant and multi-pollutant models. The multi-pollutant model included NO_2, SO_2, and CO with weather components. The pollutant models were analyzed separately

for three time intervals (one, three, five years) before the diagnosis of exposure. Each year before the diagnosis exposure model, one year, three years, and five years survival outcomes were analyzed. This is a major strength of the present investigation, a contribution to a major gap in analysis of lung cancer treatment, and a novelty not present in prior analyses.

4.2.1.2 DATA SOURCES AND COLLECTION

The SEER 18 Research Plus data access request was approved on 04/18/2022, with reference number SAR0028589, to access the data through the SEER*Stat account. AHRF is publicly available data, and the website from which it was retrieved is explained in another section of this paper. The SEER 18 Research Plus and AHRF were used from 1988 to 2016. The construction of the data file for the final analysis is shown in Figure 3.1, and the sample selection process is presented in Figures 3.2 and 4.1. The Surveillance Research Program (SRP) of the National Cancer Institute (NCI) Division of Cancer Control and Population Sciences (DCCPS) supports SEER. SEER collects and publishes cancer incidence and survival data for every cancer case reported from 22 U.S. geographic areas from population-based cancer registries, covering approximately 48 percent of the U.S. population. Registries routinely collect data on patient demographics, primary tumor site, tumor morphology and stage at diagnosis, the first course of treatment, and follow-up for vital status (survival).[74] The Area Health Resource Files (AHRF) collected data from over 50 national sources aggregated at the county level. Data collected by the Health Resources Service Administration's (HRSA) Bureau of Health Professions on each of the nation's counties.[75]

Agency pre-generated daily summary data files from 1988-2016 were downloaded from the following website: aqs.epa.gov/aqsweb/airdata/download_files.html. The air pollutiongases raw data

downloaded included ground level Ozone (O_3), Sulfur Dioxide (SO_2), Carbon Monoxide (CO), and Nitrogen Dioxide (NO_2). For particulate pollutants, the raw data downloaded were Particulate Matter ($PM_{2.5}$ and PM_{10}). We initially investigated the toxic precursor benzene among Hazardous Air Pollutants (HAPs) and Volatile Organic Compounds (VOCs); however, the high rate of missing values made it unfeasible include them in the final data analysis file engineering. The raw data files for weather data were retrieved by using the following link in the computer to access the open ftp files: ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/by_year/. Zip files from to 1988-2016 were downloaded by year and unzipped to retrieve the raw files.

4.2.1.3 STATISTICAL ANALYSIS AND MODELLING

Descriptive statistics, Kaplan–Meier survival graphs, and the Cox regression model were used to determine the sample demographics and time to all-cause mortality, with right censoring due to death or study end. The model tested the association between treatment type, air pollutants, weather, and survival, the interaction between treatment types and air pollutants, and the interaction between weather and treatment types, adjusting for patient demographics, clinical characteristics, and time-invariant unobserved variables of the year of diagnosis and county Federal Information Processing Standard (FIPS). The duration dependence of hazards because of unobserved heterogeneity was accounted for in the model by including year of diagnosis and county-specific time-invariant unobservable factors. Single-pollutant models and multi-pollutant models were computed, adjusting for the same covariates and dummy variables to determine whetehr estimates were biasedbecause of the independent variables omitted in the unadjusted model. The final model was examined for diagnostic criteria and model fit, including testing for multicollinearity between the

exposure variables. Preliminary analysis evaluated the relevant alternative regression models and employed regression diagnostics. After the preliminary analysis and diagnostics, the final regression models included NO_2, SO_2, CO, precipitation, daily minimum temperature, and snow accumulation variables. The variables supplemented by the proposed integrated DAG (iDAG) in Chapter One. Also, those variables present in the final data analysis file with complete information were used for the final post-preliminary data analysis. The factors that affected survival outcomes were air pollution, weather, treatment type, marital status, age, tumor histology, sex, insurance status, tumor grade, tumor size, race, and rural-urban status.

Kaplan-Meier survival curves and dose-response relationships between adjusted NO_2, SO_2, and CO hazards were plotted by pollutant quartile groups to determine survival probability and dose-response relationships. Survivor functions by pollutant groups were plotted for the nearest air pollution monitors up to 30 miles, the weather station at 20 miles with 25% monthly missing values, and up to 40 miles air pollution, 20 miles weather stations, and 50% missing monthly values. The rationale behind plotting two separate graphs by distance was to capture more accurate trends and differences, if any. As explained in the exposure assignment section, we might capture more accurate values by including the nearest monitoring stations, although we compromised the study power. In addition, measurement errors, such as Berkson, will persist in the nearest stations because of population and individual-level exposure differences. STATA 16 and Microsoft Excel were used for the data analysis.

As mentioned previously, survival outcome differences exist among patients with stage 1A TN0M0 NSCLC. Several factors determine the underlying reasons for it hence the model outcome is all cause hazards of death. The following empirical model analyzes survival outcomes for patients treated with fixed-effect dummy variables:

$$H(t)=h_0(t).\exp\{\beta_1.Treatment\ Type_i + \beta_2.Patient\ Demographics_i + \beta_3.Clinical\ Characteristics_i + \beta_4.County_i + \beta_6.Air\ Pollutants_i + \beta_7.Weather\ Components_i + \beta_8.Air\ Pollutants_i \times Treatment\ Type_i + \beta_9.Weather\ Components_i \times Treatment\ Type_i + \beta_{10}.Year\ of\ Diagnosis_i\}$$

where $h_0(t)$ is the baseline hazards, and $exp(\beta s)$ is the hazard ratio or rate ratio. The variables $County_i$ and $Year\ of\ Diagnosis_i$ are county and year of diagnosis time-invariant unobservable factors. In the model, *i* indicates an individual patient. "Treatment Type" is a binary variable that takes values "lobectomy" if the patient underwent lobectomy, "limited resection with adjuvant radiotherapy" if the patient underwent limited resection with adjuvant radiotherapy. Other treatment types were excluded because there were fewer observations within the radiotherapy and limited resection categories. This variable represents the vulnerability construct of Shi and Steven's[2] theory framework. Another variable that represents the vulnerability construct is the exposure of patients to air pollutants and weather components, "Air Pollutants" and "Weather Components", respectively, which differ spatiotemporally. The air pollutant measuring unit is parts per billion (ppb), the weather measuring unit for precipitation is tens of millimeters, snowfall is millimeters (mm), and the daily minimum temperature is degrees Fahrenheit (°F). Interaction terms between the vulnerability factors were also included in the model. The dependent variable, "Survival Years" represents Shi and Steven's[2] theoretical model of health outcome construct. The risk factor construct of Shi and Steven's[2] theory framework is represented by "Patient Demographics", "Clinical Characteristics", "County", and "Year of Diagnosis". "Patient Demographics" represents socioeconomic status, age, race, sex, marital status, insurance coverage, access to care, smoking status, and geographic region of residence.

The variable “Clinical Characteristics” represents tumor size, tumor grade, tumor histology, comorbidity scores, health status, coronary artery disease, disability status, surgeon expertise, mediastinal lymph node examination, quality of life, and other cardiopulmonary diseases (each of these factors were included separately in the analysis). The variables “Year of Diagnosis” and “County” represents all time-invariant unobservable variables in that particular year of diagnosis and county. It represents risk factor constructs of the theoretical model that include differential treatment relevant to yearly policy implementation at the federal, state, and local levels. County socio-demographic factors such as hospital region, treatment facility location, enrolment bias, treatment facility type, and relevant healthcare resources such as provider bias, access to care, and surgeon expertise serve as contextual variables.

4.2.1.4 SENSITIVITY ANALYSES

The robustness of the effect was tested by estimating hazards using the average monthly median and maximum exposure values for one, three, and five years before and after diagnosis obtained from the corresponding daily exposure values. The confounding effect due to omitted exposure variables was tested by running a single and multi-pollutant model.

4.2.1.5 ETHICAL CONSIDERATIONS

The University of Louisville (UofL) Institutional Review Board (IRB) approved this study (IRB number 22.0281). The study is exempt according to 45 CFR 46.101(b) under Category 4: Secondary research, for which consent is not required.

4.2.1.6 SAMPLING STRATEGY

4.2.1.6.1 POPULATION AND SAMPLE

The study sub-sample population in Chapter Two was used. The inclusion and exclusion criteria for the SEER 18 Research Plus cancer registry data and AHRF files are described in Chapter Two (Figures 3.2 and 4.1). In addition to the two databases, as explained in the data construction file, the final sample included monthly averages from daily air pollution values and weather data by miles and the percentage missing for non-missing values of variables in the regression analysis. After preliminary analysis, patients with up to five years before diagnosis exposure were included in the final analysis and followed until death or study cut off from the date of diagnosis to five years after diagnosis. The reason for including these patients was to restrict the compositional effect and misspecification error due to migration during the longer study periods. Including patients post-five years after diagnosis and prior to five years before diagnosis exposure is too long a period more prone to migration chances. As per the U.S. census bureau mobility data from 2017 to 2021 about 4 % and 2 % people of age groups 25-64 and 65 years respectively migrate to a different county. In this paper, the sampling procedure for weather and air pollutant data is described in addition to Chapter Two.

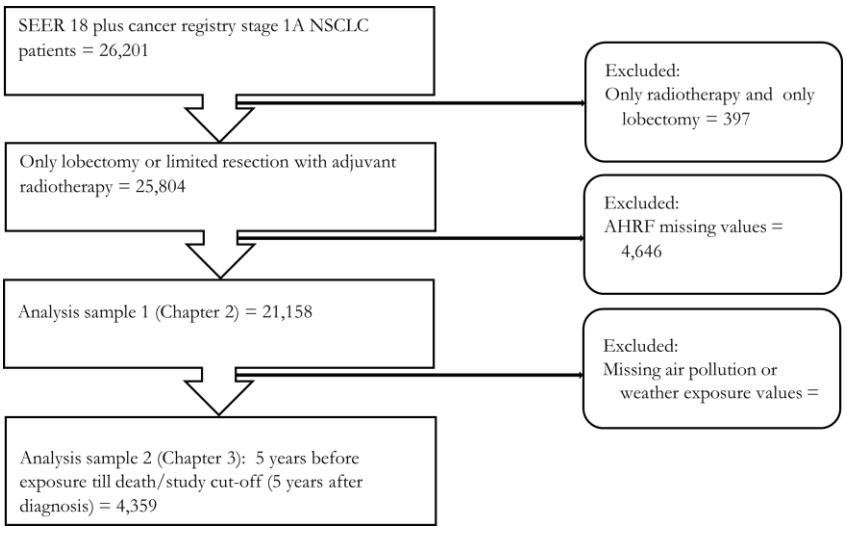


Figure 4.1 Sub-sample selection

4.2.1.6.2 WEATHER DATA

The raw data files are retrieved by using the link below to download files via ftp: ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/by_year/. All raw zip files from 1988 to 2016 were downloaded by year and unzipped for retrieval. Then, the raw files were renamed, reshaped, and cleaned in STATA wide to long format for six components (*i.e.*, daily average temperature minimum, maximum, average, snowfall, wind, and precipitation). Refer to Appendix Table 4.1 in the Appendix for the detailed steps.

4.2.1.6.3 AIR POLLUTANTS DATA

Pre-generated daily summary data files from 1988-2016 are downloaded from the following website: aqs.epa.gov/aqsweb/airdata/download_files.html. The criteria gases raw data downloaded were ozone, SO_2, CO, and NO_2. The criteria for particulate pollutants .csv data downloaded are $PM_{2.5\ and}$ PM_{10}. We initially investigated the toxic precursor benzene. However, we decided not to include them in the analysis because of significant missing values. Refer to Appendix Table 4.2 in the Appendix for the detailed steps.

4.2.1.7 VARIABLES AND EXPOSURE ASSIGNMENT

4.2.1.7.1 EXPOSURE ASSIGNMENT

Air pollution and weather exposure assignments for each patient are shown in Figure 4.2. Each patient in the final sample was assigned one, three, five, and ten years before diagnosis exposure from death or study cut-off (ten years after diagnosis). Preliminary exposure assignment data techniques informed the final exposure assignment technique from the death/study cut-off until one, three, five, and ten years before diagnosis exposure.

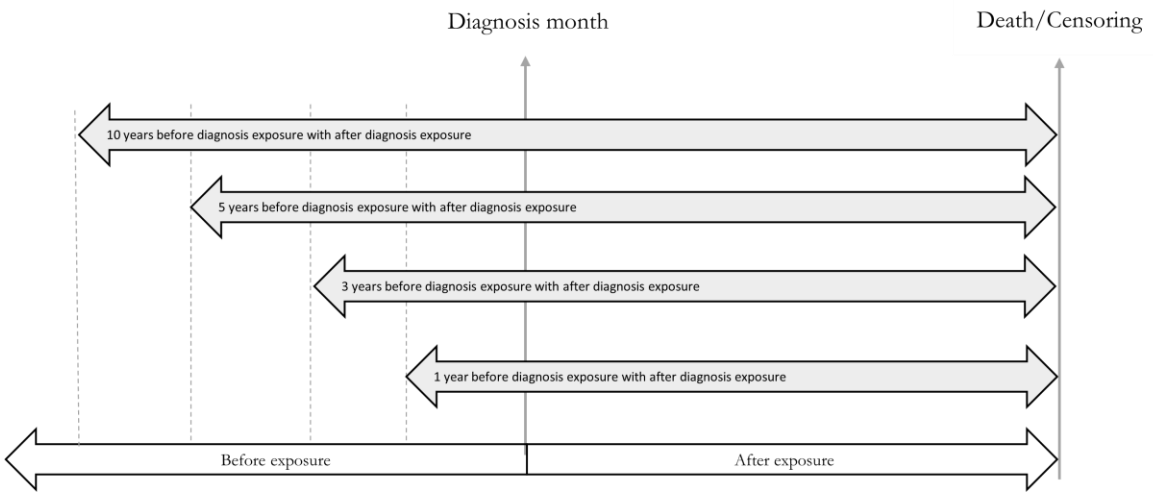


Figure 4.2 Exposure assignment to patients in data

Exposure assignments were excluded when the nearest air pollution monitoring station was >40 miles, the weather station was >20 miles, and the percentage of missing monthly values was > 50%. Preliminary sample analysis of exposure assignments for air pollution ≤ 30 miles, weather ≤ 10 miles, and $< 33.33\%$ missing values determined a very small sample size; therefore, the final analysis sample was least restrictive in terms of the distance of air pollution exposure assignments with the nearest monitoring station < 40 miles, weather station ≤ 20 miles, and missing monthly values $\leq 50\%$. The investigators were aware of the measurement error problems it may cause; however, we aimed to retain study power by being less restrictive, as measurement errors are inevitable in air pollution epidemiologic studies.

Exposure assignment errors can be categorized as measurement and misspecification errors. A recent study relevant to the current study determined that long-term exposure assignment measurement errors are inevitable in epidemiological studies and are random. Although randomly present, the classical and Berkson measurement errors obtain biased results towards the null. If the studies find a

statistically significant association, the estimates are smaller than the true effect size and less likely to be undermined.[97]

Finally, including exposure assignments for nearest stations < 30 miles or ≤ 20 miles would still have the problem of measurement error, in addition to compromising study power. The nearest station monitors might capture more accurate values or events; however, the average population exposure level is still different from the individual-level exposure, validating the presence of exposure assignment measurement errors. The latter error (*i.e.*, misspecification error) is also inevitable in similar studies, as individual patient migration data are absent in national cancer registries such as the present data. Some measures we have taken to control for larger misspecification errors restrict the study period to ten years (five years before and five years after diagnosis). An assumption was made about the absence of migration during the ten years among the included patients.

4.2.1.7.2 INDEPENDENT VARIABLES

The yearly average values of monthly means for all weather and air pollution components were utilized, and the yearly exposure variables were created, as shown in Figure 4.3. Each weather and air pollution components continuous variable included a yearly average of monthly averages before diagnosis exposure of each patient from death or study cut-off (whichever occurred first). The categorical treatment type variable included two categories: lobectomy and limited resection with adjuvant radiotherapy. Owing to the very few radiotherapy observations and limited resection, including those two category observations in the analysis was not feasible. Surgery codes for wedge resection and segmentectomy were not differentiated before 1998 in the data.[54]

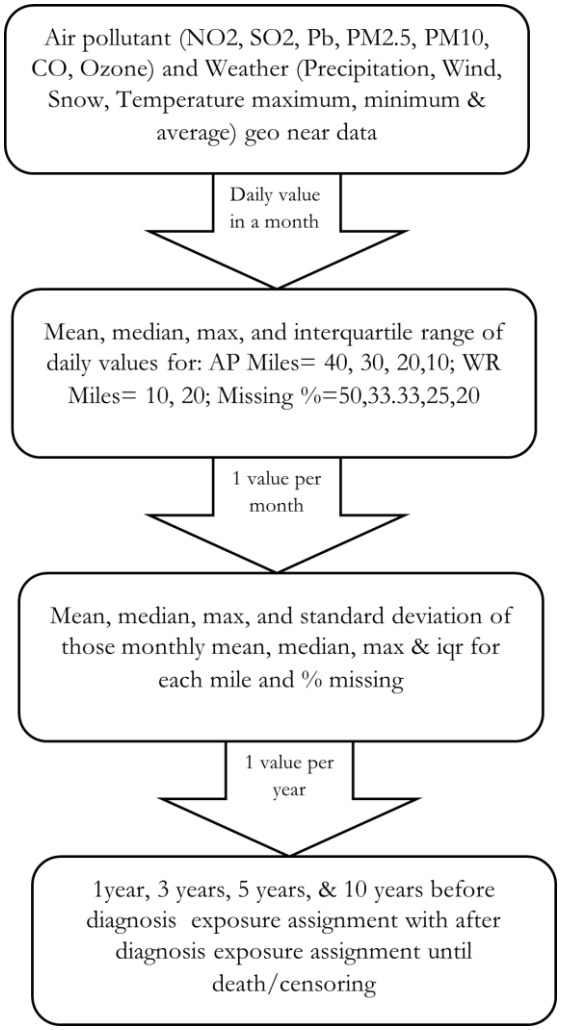


Figure 4.3 Air pollution and weather exposure yearly value creation from daily values

Hence, we adopted a conservative approach and combined the two types of surgery codes into one category, "Limited Resection," as informed by the NCCN treatment guidelines and similar studies.[98,99] The radiation sequence with surgery variable from data was utilized to aid in creating the treatment category of limited resection with adjuvant radiotherapy.

4.2.1.7.3 OUTCOME VARIABLE

Survival time was calculated as the number of months of survival from the first diagnosis to death from any cause (all-cause mortality) cumulating into years.

4.2.1.7.4 COVARIATES

Tumor size categories were constructed as described by the American Lung Cancer Society (ALCS). Due to very few observations in the category "up to 3 cm" and no specific values, the category was merged with the unknown tumor size category.[82] SEER 18 Research Plus cancer registry data lacks information on tumor size before 2004, so the patients before the 2004 diagnosis had missing tumor size values. A more conservative approach was adopted in the current study, and observations with missing information were categorized into the unknown tumor size category. Likewise, for the insurance status information, no data was available before 2007 in data, so an unknown category was constructed for insurance status information before 2007. Dummy variables for the county FIPS and year of diagnosis were constructed to account for time-invariant unobservable variables. The non-metropolitan rural-urban continuum category comprised small metropolitan, micropolitan, and non-core, as these three categories had very few observations, and there was not much demographic difference. Hence, the rural-urban continuum categorical variable comprises four categories: large central metro, large fringe metro, medium metro, and non-metropolitan. The tumor histology and insurance status variables are described in Chapter Two.

4.3 RESULTS

The sample size comprised 4,359 patients for the non-missing values of all the variables included in the regression model (Tables 4.1a and 4.1b).

	Above median		Below median	
N	2,179		2,180	
	Frequency	Percentage	Frequency	Percentage
Tumor Grade				
Grade I	262	12.02	484	22.20
Grade II	877	40.25	929	42.61
Grade III	835	38.32	564	25.87
Grade IV	30	1.38	16	0.73
Unknown	175	8.03	187	8.58
Tumor size				
Upto 1cm	42	1.93	198	9.08
> 1cm & <=2cm	208	9.55	820	37.61
>2cm	189	8.67	643	29.50
Unknown size	1,740	79.85	519	23.81
Treatment type				
Only lobectomy	1,951	89.54	1,815	83.26
Limited resection w/ adjuvant	228	10.46	365	16.74
Rural Urban Continum				
Large central metro	1,333	61.17	1,138	52.20
Large fringe metro	536	24.60	801	36.74
Medium metro	285	13.08	195	8.94
Non-metropolitan	25	1.15	46	2.11
Insurance type				
Only Medicaid	35	1.61	125	5.73
Only Medicare	166	7.62	823	37.75
Only Private	69	3.17	468	21.47
Uninsured	6	0.28	16	0.73
Unknown	1,903	87.33	748	34.31
Race				
Black	288	13.22	228	10.46
White	1,773	81.37	1,759	80.69
Unknown	118	5.42	193	8.85
Sex				
Female	969	44.47	1,226	56.24
Male	1,210	55.53	954	43.76
Marital Status				
Married	1,280	58.74	1,239	56.83
Widowed	380	17.44	277	12.71
Divorced	247	11.34	284	13.03
Single	224	10.28	278	12.75
Unknown	48	2.20	102	4.68

Table 4.1*a* Frequency statistics of study sample by above and below pollutant exposure median

	Above median			Below median		
N	2,179			2,180		
	Median	Mean	SD	Median	Mean	SD
Survival months	27	28.11	17.61	30	31.09	15.93
Panel A: Exposure to air pollutants before and after diagnosis						
NO2 exposure (ppb)	22.25	25.66	3.61	12.71	12.97	3.61
SO2 exposure (ppb)	4.10	3.98	1.20	1.56	1.81	1.20
CO exposure (ppb)	816.75	1010.84	214.13	371.03	447.91	214.13
Panel B: Weather condition before and after diagnosis						
Precipitation	24.06	26.07	8.76	22.41	23.34	10.93
Snow	0.98	1.14	1.15	0.10	1.28	1.54
Daily minimum temperature	76.04	75.90	17.66	82.80	81.92	18.01
Panel C: Individual-level charactristics						
Age at diagnosis	69	67.76	8.52	68	66.38	9.13
Panel D: County-level charactristics						
Population estimates	881,490	3,154,905	3,762,147	933,141	1,281,174	920,018
Unemployment rate	59	63.70	24.39	45	48.85	34.63
Percapita income	30496	32920.76	10118.93	47146	47803.63	15097.07
Total # hospitals	16	45.68	54.17	13	14.09	9.35
Total # hospital beds	3797	10169.78	11463.38	3130	3184.55	1979.29

Table 4.1*b* Descriptive statistics of study sample above and below pollutant exposure median

For descriptive statistics, the cohort was divided into two groups: those above and below the median exposure. The median survival times for patients above and below the median groups were about 2.3 years and 2.6 years, respectively. Likewise, the mean survival times for the above- and below-exposure median groups were approximately 2.5 and 2.7 years, respectively. The median exposure levels for above and below median groups by air pollutants are 22.25 ppb and 12.71 ppb (NO_2), 4.10 ppb and 1.56

ppb (SO_2), and 816.75 ppb and 371.03 ppb (CO), respectively. The median weather conditions for above and below median groups are 24.06 tenths of mm, and 22.41 tenths of mm (precipitation), 0.98 and 0.10 mm (snowfall), 76.04 and 82.80 $^{\circ}F$ (minimum temperatures), respectively. The mean nearest monitoring station distance from the county centroid for CO, NO_2, and SO_2 was 10.53 miles, 11.77 miles, and 15.69 miles, respectively. Likewise, the mean mile monitor distance for precipitation, snow, and daily minimum temperature is 4.13 miles, 4.48 miles, and 5.26 miles, respectively (Table 4.1c).

Element monitor	Distance in miles				
	25th Percentile	Median	75th Percentile	Mean	SD
Panel A: Sub-sample					
CO	5.57	8.85	12.72	10.53	6.49
NO2	6.56	11.66	13.61	11.77	6.99
SO2	8.80	15.92	22.22	15.69	7.14
Precipitation	3.34	3.47	5.40	4.13	1.98
Snow	3.46	3.75	5.64	4.48	2.05
Daily minimum temperature	3.37	5.07	7.05	5.26	2.59
Panel B: Above median					
CO	5.57	10.43	11.48	9.90	5.76
NO2	6.56	11.66	11.77	11.36	6.45
SO2	8.80	15.31	15.92	13.91	6.27
Precipitation	3.34	3.45	4.22	4.02	1.53
Snow	3.46	3.55	5.23	4.38	1.57
Daily minimum temperature	3.50	6.20	7.05	5.68	2.36
Panel C: Below median					
CO	6.85	8.85	17.20	11.15	7.10
NO2	6.56	11.15	15.13	12.19	7.47
SO2	11.35	16.60	22.92	17.46	7.51
Precipitation	2.53	3.58	6.61	4.23	2.35
Snow	2.60	3.95	6.54	4.58	2.42
Daily minimum temperature	2.54	4.19	6.93	4.84	2.75

Table 4.1*c* Air pollution and weather monitor stations distance in miles from county centroid

The characteristics of the sample are listed in Tables 4.1a and 4.1b. The dose-response relationship of pollutants was directly compared to the first quartiles. The hazards were consistently higher with each quartile unit increase in the air pollutant level dose in the multi-pollutant model (Figure 4.4). The dose-response relationship was plotted for mean monthly averages for NO_2, SO_2, and CO pollutants in a multi-pollutant model for one, three, and five years survival for five years before diagnosis exposure. Compared to the first quartile, the hazards of death were generally higher for the second, third, and fourth quartiles for all pollutants. For NO_2 and CO pollutants, the relationship seems to be linear. The NO_2 pollutant second quartile dose-response relationship 95% confidence interval for three and five years survival did not seem to overlap with the third and fourth quartiles. Similarly, the 95% confidence interval for SO_2 pollutant second quartile dose-response relationship for three and five years survival does not seem to overlap the third and fourth quartiles.

The survivor functions for CO, NO_2, and SO_2 (Figures 4.5-4.10) were plotted for one, three, and five years before diagnosis exposure. Each year before diagnosis exposure, the one year, three years, and five years OS were plotted for each pollutant above and below the median groups. Pollutant groups were generated by dividing the exposure variables below and above the median values. Overall, those exposed to above-median pollutant levels had lower OS than those exposed to lower levels. The Kaplan-Meier survival plots for multi-pollutant and single-pollutant models for ≤ 30 miles of air pollution, ≤ 20 miles of weather station, and 25% missing value graphs (Appendix Figures 4.1-4.6) and 40 miles, 20 miles, and 50% missing values were also plotted (Figures 4.5-4.10).

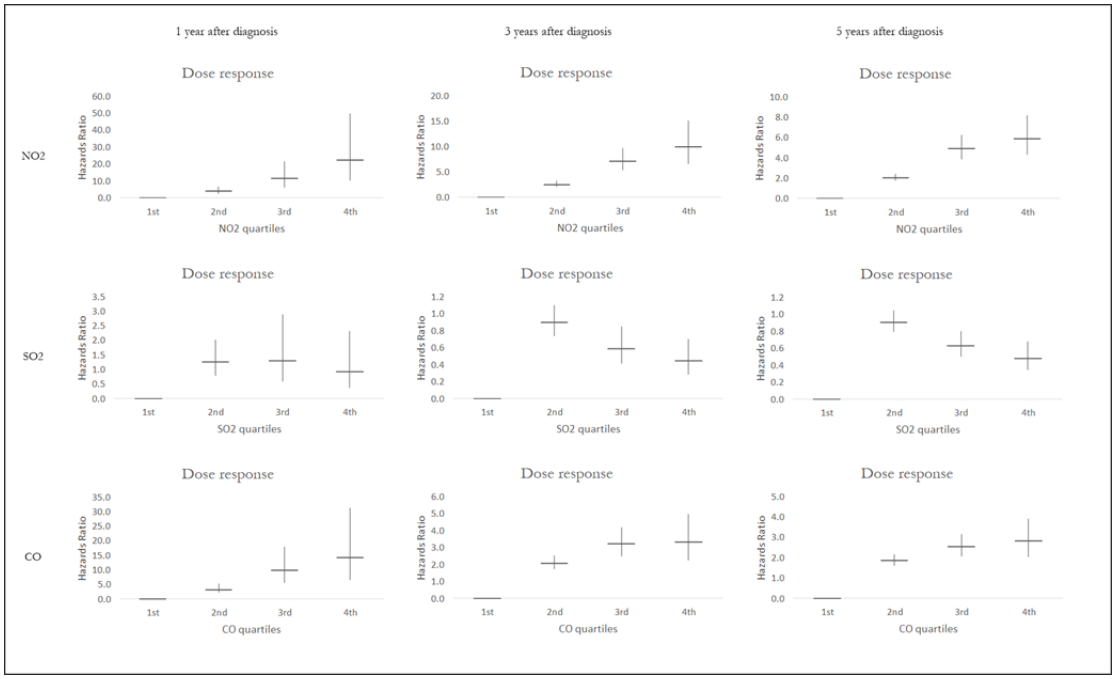


Figure 4.4 Dose response relationship for adjusted hazards ratio by pollutant quartile groups

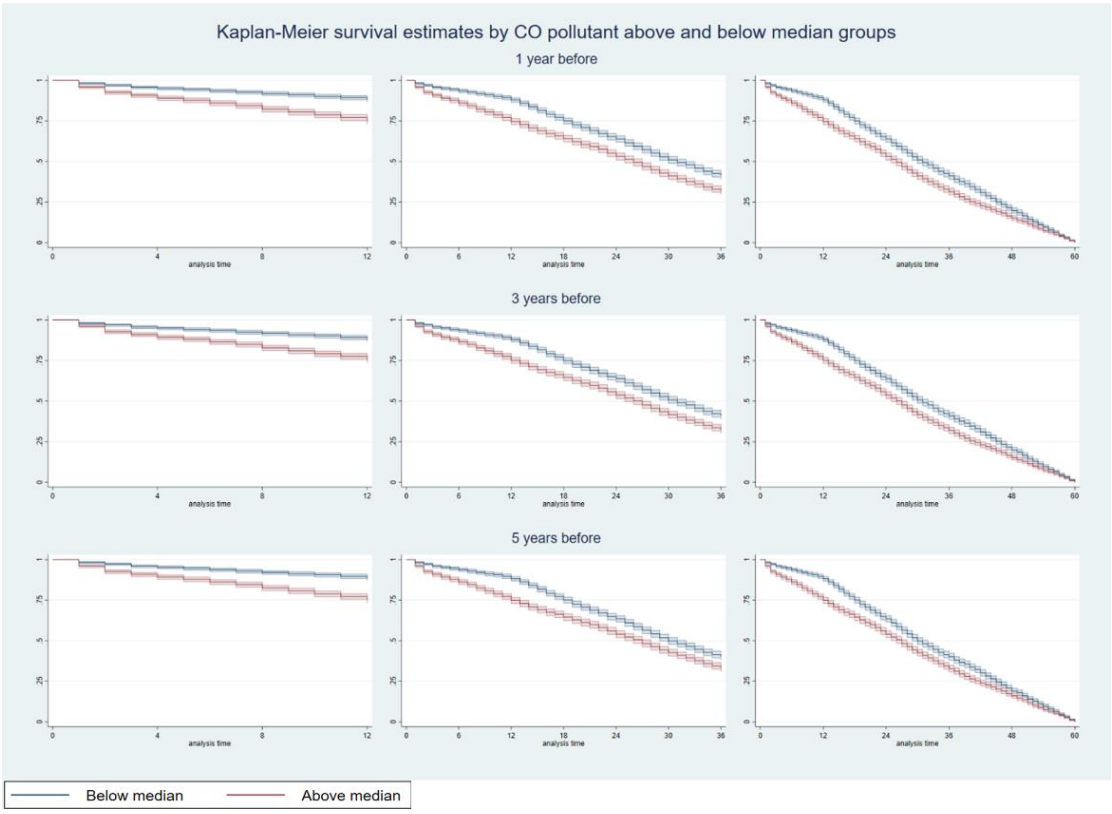


Figure 4.5 Multi-pollutant Model: Kaplan-Meier survival estimates with 95% confidence interval by CO above and below median groups, maximum 40 miles distance 50% missing.

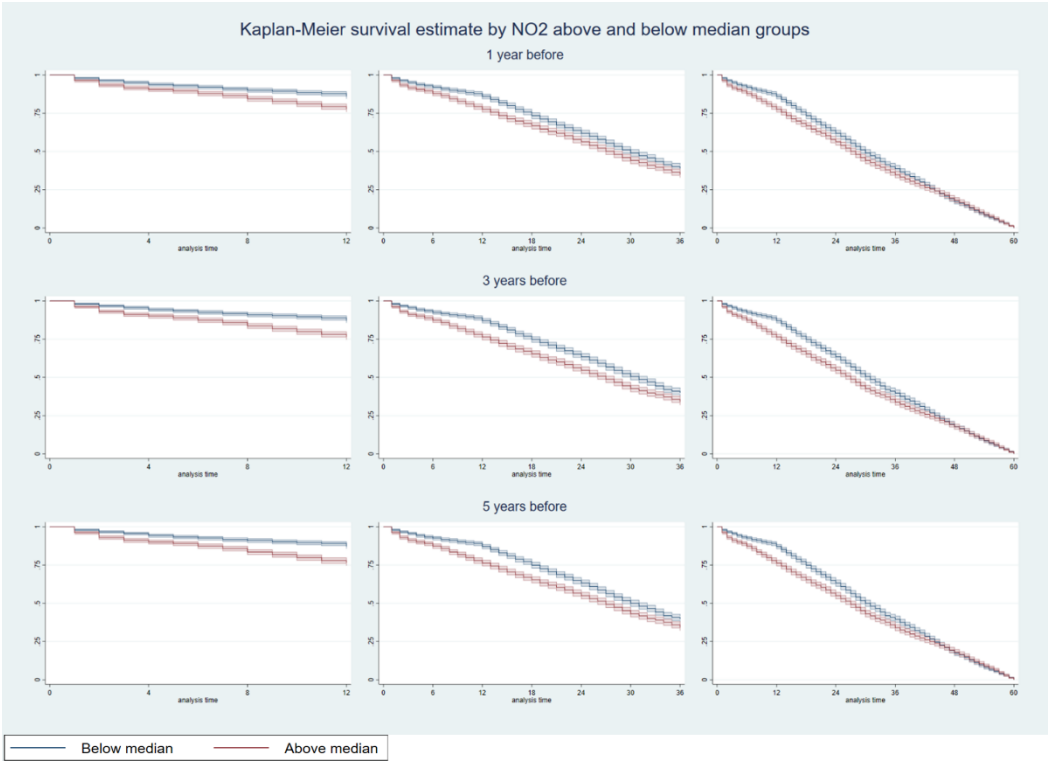


Figure 4.6 Multi-pollutant Model: Kaplan-Meier survival estimates with 95% confidence interval by NO_2 above and below median groups, maximum 40 miles distance 50% missing.

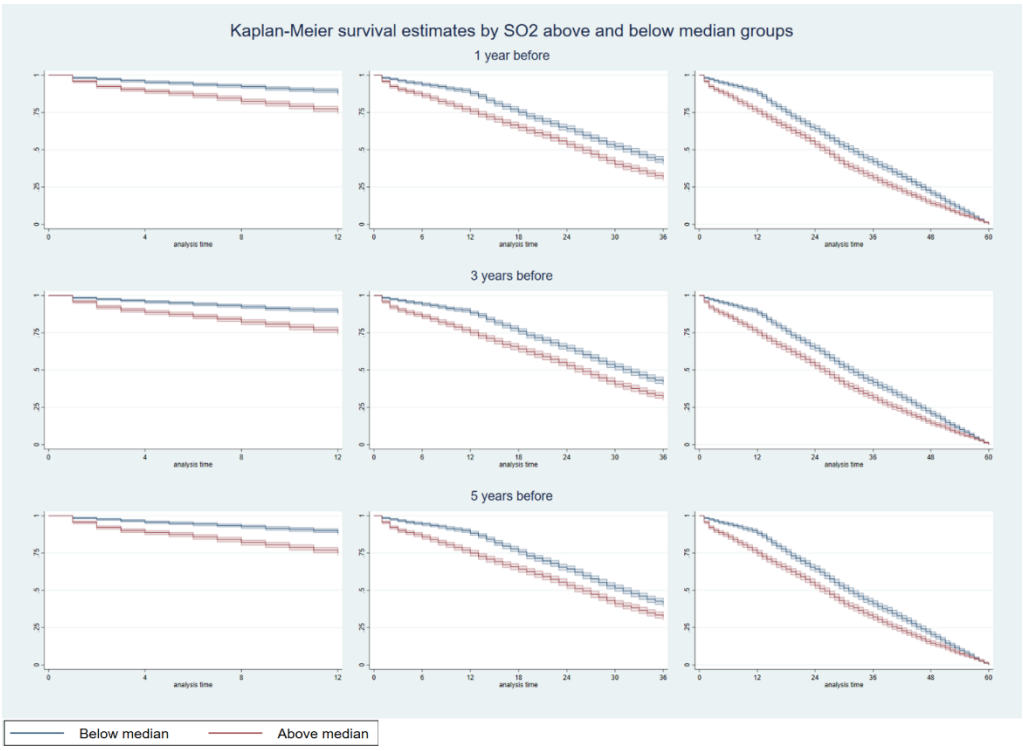


Figure 4.7 Multi-pollutant Model: Kaplan-Meier survival estimates with 95% confidence interval by SO_2 above and below median groups, maximum 40 miles distance 50% missing

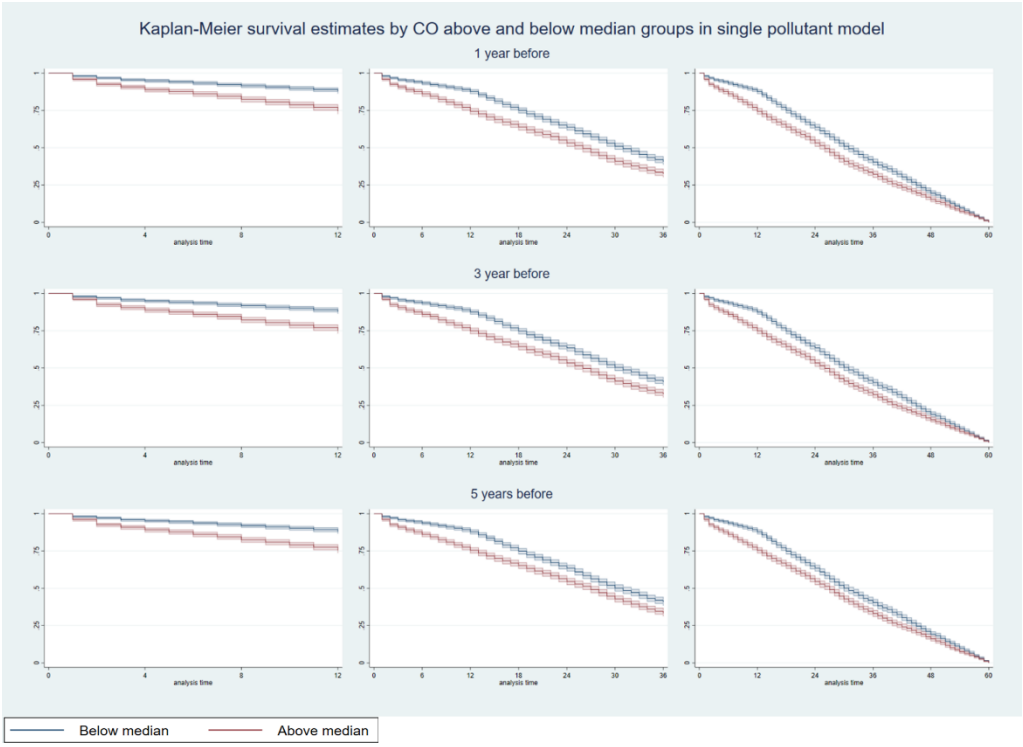


Figure 4.8 Single-pollutant Model: Kaplan-Meier survival estimates with 95% confidence intervals by CO above and below median groups, maximum 40 miles distance 50% missing.

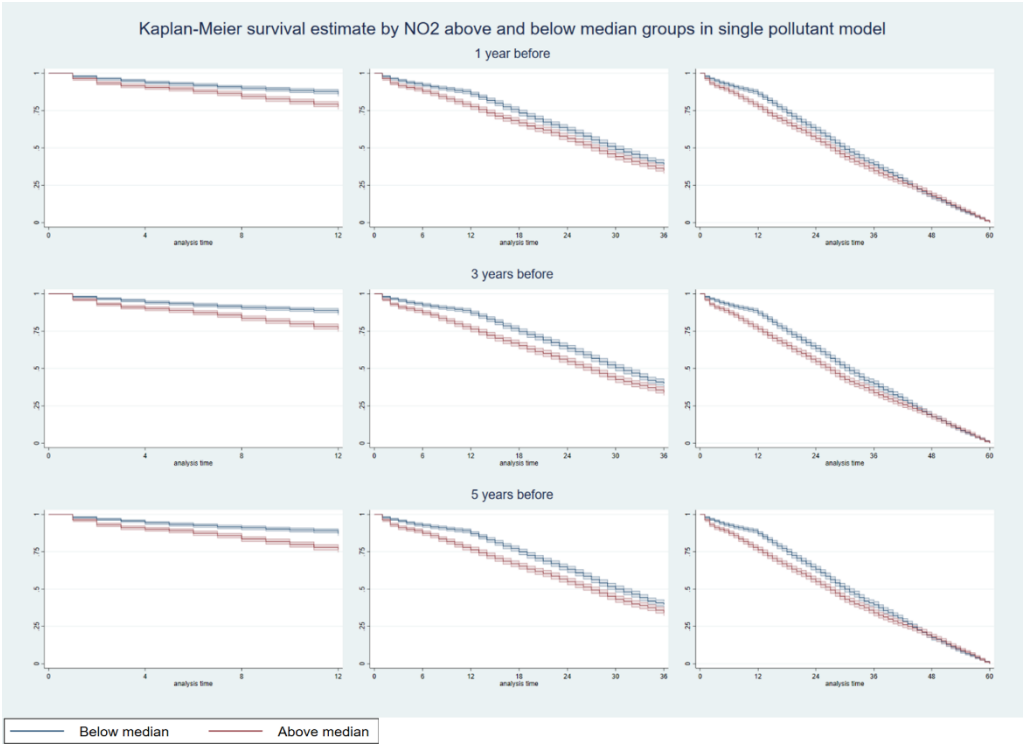


Figure 4.9 Single-pollutant Model: Kaplan-Meier survival estimates with 95% confidence interval by NO_2 above and below median groups, maximum 40 miles distance 50% missing

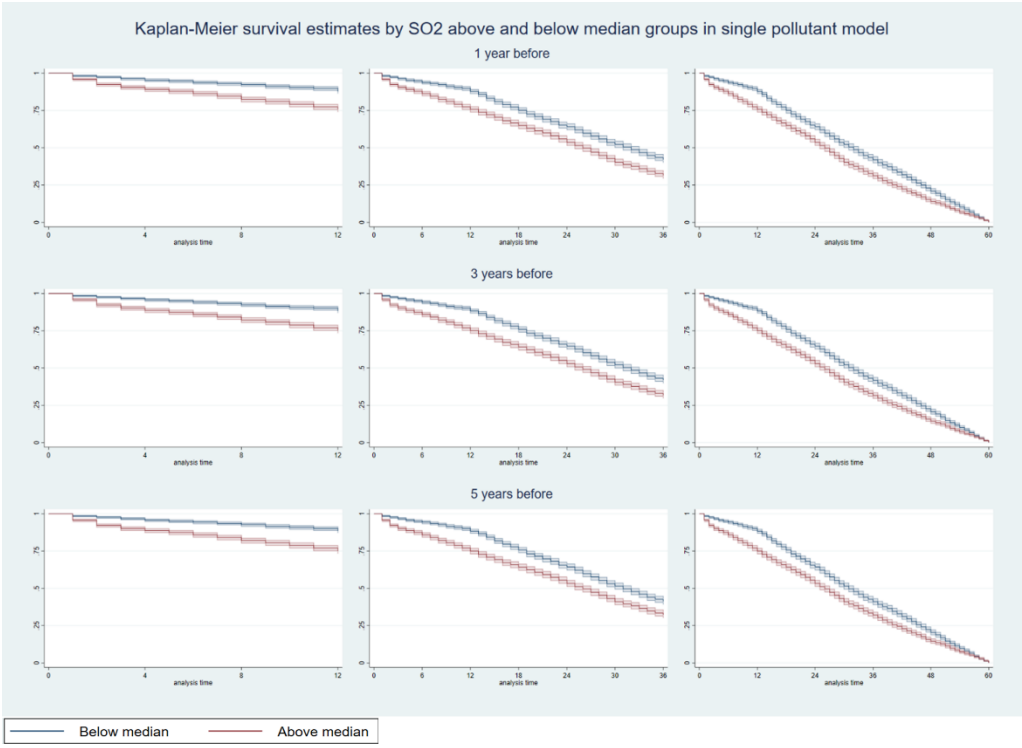


Figure 4.10 Single pollutant Model: Kaplan-Meier survival estimates with 95% confidence interval by SO_2 above and below median groups, maximum 40 miles distance 50% missing

Overall, those exposed to above-median levels of air pollutants had a lower survival probability than those exposed to below-median levels. The single-pollutant model graphs did not seem to have striking differences from their multi-pollutant counterparts, proving the robustness of the results. Similarly, the 30 miles of air pollution nearest station values reflected similar directions of survival probability. Those exposed to above the median had lower OS than those exposed to below the median. As explained in the exposure assignment section, longer periods imply a higher chance of misspecification errors. Hence, the survival probability seems similar for both groups after about 1.5 years for 1-5 years before diagnosis exposures. Additionally, multi-pollutant Kaplan-Meier graphs for the 40 miles nearest air pollution monitor values

reflected a difference in survival probabilities of the two groups up to three years after diagnosis.

4.3.1 HAZARDS OF DEATH ONE YEAR AFTER DIAGNOSIS

The all-cause mortality hazard of death for those exposed to NO_2 increased by 19%, 25%, and 21%, with an average monthly increase of 1 ppb for one, three, and five years before diagnosis exposure, respectively (Table 4.2a). Those exposed to SO_2 had an increase in all-cause mortality hazards by 61%, 86%, and 98%, with an average increase in monthly averages of 1 ppb for one, three, and five years before diagnosis exposure. Those exposed to CO had a 100% increase in all-cause mortality hazards of death, with an average monthly increase of 1 ppb for one, three, and five years before diagnosis exposure, respectively. Death hazards for those exposed to precipitation decreased by 12% and 14% by an average monthly increase of one-tenth of a millimeter for three and five years before diagnosis exposure, respectively. Similarly, the hazard of death for those exposed to snowfall decreased by 26% and 41%, with an average monthly increase of one mm for three and five years before diagnosis exposure, respectively. Death hazards for those exposed to daily minimum temperatures increased modestly by 3-5% with an average increase in monthly $1^{\circ}F$ for one, three, and five years before diagnosis exposure, respectively. The hazard effect modestly changed the effect size for single-pollutant models; however, the estimates remained significant. For the SO_2 single-pollutant model, statistical significance no longer remains a hazard of death due to precipitation. In contrast, the NO_2 and CO single-pollutant models showed no statistical significance for the minimum temperature increase, reflecting biased estimates due to omitted variable bias.

The all-cause mortality hazard of death due to exposure to an average increase of 1 ppb to NO_2 modestly increased by 6% for three and five years before diagnosis, respectively, for patients who underwent limited resection with adjuvant radiotherapy compared to those who received lobectomy. In addition, the hazards of exposure to an average increase of 1 ppb in SO_2 modestly increased by 15% for one year before diagnosis in patients who received limited resection with adjuvant radiotherapy compared with those who received lobectomy. In contrast, the hazards of being exposed to an average increase in CO by 1 ppb decreased by 58% for three years before diagnosis exposure to those patients who received limited resection with adjuvant radiotherapy compared to those who received lobectomy at a 1 % significance level. However, the single-pollutant CO model did not determine a statistically significant interaction with the treatment, which was not true for the SO_2 and NO_2 models.

The all-cause mortality hazards of male deaths increased by 53% and 52% compared to females for one, three, and five years before diagnosis exposure, respectively (Table 4.2b). The hazard of death increased by 33% for those with tumor grade III compared with those with tumor grade II for one, three, and five years before diagnosis exposure, respectively. Similarly, the hazard of death increased by 66%, 67%, and 62% for those with tumor grade IV compared with those with tumor grade II for one, three, and five years before diagnosis exposure, respectively. The increase in age at diagnosis has a moderately higher hazard of death (2%) for all years before diagnosis exposure. Patients with adenoma histology type had all-cause mortality hazards that decreased by approximately 27% compared to those with squamous cell carcinomas.

	Multipollutant			NO2			SO2			CO		
	Hazard of death 1 year after diagnosis			Hazard of death 1 year after diagnosis			Hazard of death 1 year after diagnosis			Hazard of death 1 year after diagnosis		
	Duration of exposure from diagnosis			Duration of exposure from diagnosis			Duration of exposure from diagnosis			Duration of exposure from diagnosis		
	1 yr bf	3 yrs. bf	5 yrs. bf	1 yr bf	3 yrs. bf	5 yrs. bf	1 yr bf	3 yrs. bf	5 yrs. bf	1 yr bf	3 yrs. bf	5 yrs. bf
Air pollutants and weather components												
NO2	1.19***	1.25***	1.21***	1.16***	1.16***	1.13***						
	(1.12 , 1.27)	(1.15 , 1.35)	(1.12 , 1.31)	(1.13 , 5.73)	(1.14 , 5.73)	(1.11 , 4.10)						
SO2	1.61***	1.86***	1.98***				1.41***	1.54***	1.58***			
	(1.44 , 1.80)	(1.66 , 2.08)	(1.76 , 2.22)				(1.31 , 1.52)	(1.42 , 1.67)	(1.47 , 1.71)			
CO	5.79***	10.55***	10.49***							10.82***	17.41***	10.71***
	(2.96 , 11.31)	(4.72 , 23.58)	(4.22 , 26.06)							(7.29 , 16.05)	(11.18 , 27.09)	(7.18 , 15.98)
Precipitation	0.98	0.88***	0.86***	0.98***	0.97***	0.97***	1	0.98***	0.97***	0.97***	0.95***	0.95***
	(0.94 , 1.02)	(0.84 , 0.93)	(0.80 , 0.91)	(0.96 , 1)	(0.95 , 0.98)	(0.96 , 0.99)	(0.98 , 1.02)	(0.96 , 1)	(0.94 , 0.99)	(0.95 , 0.99)	(0.93 , 0.97)	(0.93 , 0.97)
Snow	0.84	0.74***	0.59***	0.81***	0.75***	0.72***	0.86*	0.79***	0.71***	0.79***	0.72***	0.71***
	(0.67 , 1.04)	(0.56 , 0.98)	(0.41 , 0.84)	(0.71 , 0.91)	(0.67 , 0.84)	(0.64 , 0.81)	(0.72 , 1.01)	(0.66 , 0.94)	(0.59 , 0.86)	(0.70 , 0.89)	(0.65 , 0.81)	(0.64 , 0.79)
Daily temperature minimum	1.03***	1.04***	1.05***	1	1	1	0.98***	0.98***	0.98***	1	0.99	1
	(1.01 , 1.05)	(1.01 , 1.06)	(1.02 , 1.09)	(0.99 , 1.01)	(0.99 , 1.01)	(0.99 , 1.01)	(0.97 , 1)	(0.97 , 0.99)	(0.97 , 0.99)	(0.99 , 1.01)	(0.98 , 1)	(0.99 , 1.01)
Treatment options (reference :lobectomy)												
Limited resection with adjuvant radiotherapy	0.48	0.27	0.40	0.92	0.54	0.75	1.08	0.73	0.74	1.62	1.08	1.32
	(0.04 , 6.13)	(0.03 , 2.94)	(0.03 , 5.31)	(.08 , 10.92)	(.06 , 4.93)	(.07 , 7.72)	(.11 , 10.31)	(.09 , 6.05)	(.08 , 7.23)	(.15 , 17.38)	(.12 , 9.7)	(.13 , 13.14)
Treatment interaction with air pollutant and weather components												
NO2 * Treatment	1.05	1.06**	1.06**	1.02**	1.02**	1.02**						
	(1. , 1.11)	(1.01 , 1.12)	(1.01 , 1.11)	(1 , 1.04)	(1 , 1.05)	(1 , 1.04)						
SO2 * Treatment	1.15*	1.10	1.11				1.16**	1.12	1.14*			
	(0.99 , 1.34)	(0.95 , 1.27)	(0.96 , 1.29)				(1.01 , 1.35)	(.98 , 1.29)	(.99 , 1.31)			
CO * Treatment	0.51	0.42*	0.45							1.26	1.24	1.24
	(0.20 , 1.33)	(0.16 , 1.14)	(0.17 , 1.19)							(.83 , 1.91)	(.81 , 1.89)	(.83 , 1.86)
Precipitation * Treatment	1	1.02	1.01	1	1.01	1.01	0.99	1.01	1	1	1.01	1.01
	(0.96 , 1.03)	(0.98 , 1.05)	(0.98 , 1.05)	(.97 , 1.03)	(.99 , 1.04)	(.98 , 1.04)	(.96 , 1.02)	(.98 , 1.04)	(.97 , 1.03)	(.97 , 1.03)	(.98 , 1.04)	(.98 , 1.03)
Snow * Treatment	0.94	0.95	0.92	0.96	0.97	0.95	0.92	0.93	0.93	0.92	0.91	0.91
	(0.68 , 1.29)	(0.69 , 1.32)	(0.63 , 1.36)	(.73 , 1.28)	(.73 , 1.29)	(.68 , 1.32)	(.69 , 1.23)	(.69 , 1.26)	(.66 , 1.33)	(.7 , 1.22)	(.69 , 1.22)	(.65 , 1.25)
Temperature minimum * Treatment	1	1	1	0.99	1	0.99	1	1	1	0.99	0.99	0.99
	(0.97 , 1.02)	(0.98 , 1.02)	(0.97 , 1.02)	(.97 , 1.02)	(.97 , 1.02)	(.97 , 1.02)	(.97 , 1.02)	(.97 , 1.02)	(.97 , 1.02)	(.97 , 1.02)	(.97 , 1.02)	(.97 , 1.02)

pvalue: * <0.1, ** <0.05, *** <0.01

Table 4.2a Hazards of death one year after diagnosis for air pollutants, weather, and treatment type by annual average monthly mean

Note: Air pollution levels and weather conditions are calculated based on the recording of monitors within 40 miles and 20 miles, respectively, of the centroid of the county of residence; missing values for air pollution and weather up to half of the days of a month are allowed

	Multipollutant Hazard of death 1 year after diagnosis Duration of exposure from diagnosis			NO_2 Hazard of death 1 year after diagnosis Duration of exposure from diagnosis			SO_2 Hazard of death 1 year after diagnosis Duration of exposure from diagnosis			CO Hazard of death 1 year after diagnosis Duration of exposure from diagnosis		
	1 yr bf	3 yrs. bf	5 yrs. bf	1 yr bf	3 yrs. bf	5 yrs. bf	1 yr bf	3 yrs. bf	5 yrs. bf	1 yr bf	3 yrs. bf	5 yrs. bf
Race (reference: Black)												
Other	1.06	1.04	1.04	1.04	1.06	1.07	1.03	1.01	1.01	1.07	1.08	1.09
	(.72 , 1.55)	(.71 , 1.53)	(.71 , 1.52)	(.71 , 1.52)	(.73 , 1.55)	(.73 , 1.57)	(.7 , 1.5)	(.69 , 1.49)	(.69 , 1.48)	(.74 , 1.56)	(.74 , 1.57)	(.75 , 1.58)
White	0.91	0.91	0.91	0.9	0.91	0.91	0.89	0.88	0.88	0.89	0.9	0.9
	(.72 , 1.15)	(.72 , 1.15)	(.72 , 1.15)	(.71 , 1.13)	(.72 , 1.15)	(.72 , 1.15)	(.7 , 1.12)	(.7 , 1.12)	(.7 , 1.12)	(.71 , 1.13)	(.71 , 1.14)	(.71 , 1.14)
Sex (reference: Female)												
Male	1.53***	1.53***	1.52***	1.53***	1.54***	1.53***	1.54***	1.55***	1.55***	1.53***	1.53***	1.52***
	(1.31 , 1.79)	(1.31 , 1.79)	(1.3 , 1.78)	(1.31 , 1.79)	(1.32 , 1.8)	(1.31 , 1.79)	(1.31 , 1.8)	(1.32 , 1.81)	(1.32 , 1.81)	(1.31 , 1.79)	(1.31 , 1.79)	(1.3 , 1.78)
Tumor Grade (reference: II)												
Grade III	1.33***	1.33***	1.33***	1.34***	1.34***	1.33***	1.36***	1.36***	1.36***	1.33***	1.33***	1.32***
	(1.13 , 1.57)	(1.13 , 1.57)	(1.13 , 1.57)	(1.14 , 1.58)	(1.14 , 1.58)	(1.13 , 1.57)	(1.16 , 1.61)	(1.16 , 1.61)	(1.16 , 1.61)	(1.13 , 1.57)	(1.13 , 1.56)	(1.12 , 1.56)
Grade IV	1.66*	1.67*	1.62*	1.75**	1.74*	1.7*	1.77**	1.77**	1.75**	1.68*	1.68*	1.61*
	(.97 , 2.85)	(.96 , 2.89)	(.93 , 2.83)	(1. , 3.08)	(.99 , 3.06)	(.97 , 2.99)	(1.02 , 3.07)	(1.02 , 3.07)	(1.01 , 3.02)	(.97 , 2.9)	(.96 , 2.92)	(.92 , 2.82)
Unknown	1.36**	1.36**	1.35**	1.36**	1.37**	1.36**	1.32**	1.31**	1.32**	1.36**	1.36**	1.35**
	(1.04 , 1.79)	(1.04 , 1.78)	(1.03 , 1.77)	(1.04 , 1.78)	(1.04 , 1.79)	(1.04 , 1.79)	(1. , 1.73)	(1. , 1.73)	(1.01 , 1.73)	(1.04 , 1.79)	(1.04 , 1.79)	(1.03 , 1.77)
Grade I	1.04	1.06	1.06	1.06	1.07	1.06	1.05	1.07	1.07	1.06	1.07	1.06
	(.81 , 1.33)	(.83 , 1.36)	(.83 , 1.36)	(.83 , 1.35)	(.84 , 1.37)	(.83 , 1.36)	(.82 , 1.34)	(.83 , 1.37)	(.83 , 1.37)	(.83 , 1.35)	(.84 , 1.37)	(.83 , 1.35)
Marital status (reference: Divorced)												
Married	0.93	0.94	0.93	0.93	0.93	0.92	0.91	0.91	0.9	0.93	0.93	0.94
	(.74 , 1.17)	(.75 , 1.18)	(.74 , 1.17)	(.74 , 1.16)	(.74 , 1.16)	(.73 , 1.15)	(.72 , 1.14)	(.72 , 1.14)	(.72 , 1.13)	(.74 , 1.16)	(.74 , 1.18)	(.74 , 1.18)
Single	1.02	1.05	1.05	1.01	1.02	1.02	0.96	0.97	0.97	1	1.01	1.01
	(.75 , 1.39)	(.77 , 1.42)	(.77 , 1.43)	(.75 , 1.38)	(.75 , 1.39)	(.75 , 1.38)	(.71 , 1.31)	(.71 , 1.32)	(.71 , 1.32)	(.73 , 1.35)	(.74 , 1.37)	(.74 , 1.37)
Unknown	0.86	0.91	0.87	0.88	0.92	0.89	0.89	0.92	0.89	0.86	0.89	0.87
	(.52 , 1.42)	(.55 , 1.5)	(.53 , 1.43)	(.53 , 1.47)	(.55 , 1.52)	(.54 , 1.49)	(.54 , 1.47)	(.56 , 1.51)	(.54 , 1.47)	(.52 , 1.43)	(.54 , 1.48)	(.52 , 1.44)
Widowed	1.06	1.05	1.06	1.05	1.05	1.05	1.03	1.02	1.03	1.06	1.05	1.05
	(.8 , 1.41)	(.79 , 1.39)	(.8 , 1.41)	(.8 , 1.39)	(.79 , 1.39)	(.8 , 1.39)	(.78 , 1.36)	(.77 , 1.35)	(.78 , 1.36)	(.8 , 1.4)	(.79 , 1.38)	(.8 , 1.39)
Tumor size (reference: upto 1cm)												
>1cm & <=2cm	0.95	0.93	0.92	0.99	0.97	0.95	0.94	0.93	0.92	0.99	0.98	0.98
	(.6 , 1.5)	(.59 , 1.46)	(.59 , 1.44)	(.63 , 1.54)	(.62 , 1.52)	(.61 , 1.48)	(.6 , 1.49)	(.59 , 1.46)	(.59 , 1.45)	(.63 , 1.55)	(.63 , 1.54)	(.63 , 1.53)
>2cm	1.07	1.04	1.03	1.1	1.08	1.05	1.03	1.01	1	1.11	1.1	1.09
	(.67 , 1.69)	(.65 , 1.64)	(.65 , 1.62)	(.7 , 1.74)	(.68 , 1.7)	(.67 , 1.66)	(.65 , 1.63)	(.64 , 1.59)	(.64 , 1.58)	(.7 , 1.75)	(.7 , 1.74)	(.69 , 1.72)
Unknown	0.81	0.9	0.8	0.89	0.96	0.82	0.9	0.91	0.86	0.68	0.7	0.67
	(.13 , 5.16)	(.13 , 6.06)	(.12 , 5.17)	(.15 , 5.39)	(.14 , 6.51)	(.11 , 5.81)	(.13 , 6.03)	(.14 , 6.13)	(.13 , 5.61)	(.09 , 5.42)	(.08 , 6.17)	(.08 , 5.61)
Tumor histology (reference: squamous cell)												
Adenomas	0.73***	0.74***	0.73***	0.74***	0.75***	0.74***	0.73***	0.73***	0.73***	0.73***	0.74***	0.73***
	(.63 , .85)	(.63 , .86)	(.62 , .85)	(.64 , .87)	(.64 , .88)	(.64 , .87)	(.63 , .85)	(.63 , .86)	(.63 , .85)	(.63 , .85)	(.63 , .86)	(.63 , .86)
Age at diagnosis	1.02***	1.02***	1.02***	1.02***	1.02***	1.02***	1.02***	1.02***	1.02***	1.02***	1.02***	1.02***
	(1.01 , 1.03)	(1.01 , 1.03)	(1.01 , 1.03)	(1.01 , 1.03)	(1.01 , 1.03)	(1.01 , 1.03)	(1.01 , 1.03)	(1.01 , 1.03)	(1.01 , 1.03)	(1.01 , 1.03)	(1.01 , 1.03)	(1.01 , 1.03)
Insurance type (reference: Only Medicaid)												
Only Medicare	0.59**	0.58**	0.60**	0.59**	0.58**	0.58**	0.63*	0.63*	0.65*	0.57**	0.56**	0.57**
	(.36 , .94)	(.36 , .95)	(.37 , .98)	(.36 , .95)	(.36 , .94)	(.36 , .95)	(.39 , 1.01)	(.39 , 1.02)	(.4 , 1.05)	(.35 , .92)	(.35 , .91)	(.35 , .92)
Only private	0.41***	0.40***	0.43***	0.40***	0.39***	0.39***	0.47***	0.47***	0.50**	0.39***	0.38***	0.39***

	(.23 , .73)	(.22 , .72)	(.24 , .77)	(.22 , .71)	(.22 , .7)	(.22 , .71)	(.26 , .82)	(.27 , .83)	(.28 , .88)	(.22 , .69)	(.21 , .68)	(.22 , .7)
Uninsured	2.54*	2.81**	3.17**	1.99	2.05	2.15	2.22	2.35*	2.43*	2.04	2.11	2.24
	(.91 , 7.08)	(1.02 , 7.72)	(1.14 , 8.82)	(.74 , 5.38)	(.77 , 5.45)	(.8 , 5.78)	(.83 , 5.94)	(.88 , 6.26)	(.91 , 6.51)	(.74 , 5.61)	(.79 , 5.64)	(.83 , 6.08)
Unknown	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***
	(0.00 , 0.00)	(0.00 , 0.00)	(0.00 , 0.00)	(0.00 , 0.00)	(0.00 , 0.00)	(0.00 , 0.00)	(0.00 , 0.00)	(0.00 , 0.00)	(0.00 , 0.00)	(0.00 , 0.00)	(0.00 , 0.00)	(0.00 , 0.00)
Rural-Urban continuum (reference: Large central metro)												
Large fringe metro	0.82	1.05	0.84	0.93	1.31	1.11	0.52	0.47	0.43	0.93	1.2	1.08
	(.07 , 9.34)	(.08 , 13.93)	(.05 , 12.87)	(.08 , 10.21)	(.12 , 14.58)	(.1 , 12.71)	(.05 , 5.81)	(.04 , 5.48)	(.03 , 5.35)	(.09 , 9.82)	(.11 , 13.47)	(.09 , 13.37)
Medium metro	0.32	0.07**	0.08**	0.27	0.09**	0.18	0.45	0.18	0.26	0.31	0.15*	0.34
	(.05 , 2.18)	(.01 , .58)	(.01 , .85)	(.04 , 1.82)	(.01 , .75)	(.02 , 1.79)	(.06 , 3.09)	(.02 , 1.56)	(.02 , 2.77)	(.05 , 2.08)	(.02 , 1.14)	(.04 , 3.25)
Non-metropolitan	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***
	(0.00 , 0.00)	(0.00 , 0.00)	(0.00 , 0.00)	(0.00 , 0.00)	(0.00 , 0.00)	(0.00 , 0.00)	(0.00 , 0.00)	(0.00 , 0.00)	(0.00 , 0.00)	(0.00 , 0.00)	(0.00 , 0.00)	(0.00 , 0.00)

pvalue: * <0.1, ** <0.05, *** <0.01

Table 4.2b Hazards of death one year after diagnosis for study covariates by annual average monthly mean

Note: Air pollution levels and weather conditions are calculated based on the recording of monitors within 40 miles and 20 miles, respectively, of the centroid of the county of residence; missing values for air pollution and weather up to half of the days of a month are allowed

Compared to Medicaid alone, patients with Medicare insurance had decreased death hazards by 41%, 42%, and 40% at one, three, and five years before diagnosis exposure, respectively. Compared with Medicaid alone, patients with only private insurance had decreased hazards of death by 58%, 60%, and 57% at one, three, and five years before diagnosis exposure, respectively. Uninsured patients had 100% higher death hazards for all years before diagnosis. However, the hazard of death for the uninsured was not statistically significant in the NO2 and CO single-pollutant models.

4.3.2 HAZARDS OF DEATH THREE YEARS AFTER DIAGNOSIS

The all-cause mortality hazard of death for those exposed to NO_2 increased by 5%, 7%, and 9% with an average monthly increase of 1 ppb for one, three, and five years before diagnosis exposure, respectively (Table 4.3a). Those exposed to SO_2 had an increase in all-cause mortality hazards by 18% and 28%, with an average increase in monthly averages of 1 ppb for one, three, and five years before diagnosis exposure.

Those exposed to CO had an increase in all-cause mortality hazards of 95%, 98%, and 93%, with an average increase in monthly averages of 1 ppb for one, three, and five years before diagnosis exposure, respectively. Death hazards for those exposed to precipitation decreased by 2% and 4% by an average monthly increase of one-tenth of a millimeter for one, three, and five years before diagnosis exposure, respectively. Similarly, the hazards of death for those exposed to snowfall decreased by 10% and 15%, with an average monthly increase of one mm for three and five years before diagnosis exposure, respectively. Death hazards for those exposed to daily minimum temperature increased modestly by 3%, with an average increase in monthly $1^{\circ}F$ for five years before diagnosis exposure. The hazard effect modestly changed the effect size for single-pollutant models; however, the estimates remained significant. For the SO_2 single-pollutant model, statistical significance no longer remained for hazards of death due to precipitation.

The all-cause mortality hazards of death due to exposure to an average increase of 1 ppb to NO_2 modestly increased by 2% for three and five years before diagnosis exposure, respectively, for patients who received limited resection with adjuvant radiotherapy compared to those who received lobectomy. However, the single-pollutant CO model showed a statistically significant interaction with treatment, with increased mortality hazards.

The all-cause mortality hazards of male deaths increased by approximately 18% compared with females for one, three, and five years before diagnosis exposure (Table 4.3b). The hazard of death increased by approximately 17% for those with tumor grade III compared with those with tumor grade II for one, three, and five years before diagnosis exposure, respectively. The all-cause mortality hazard of death decreased by approximately 12% for patients with tumors up to 2 cm in size compared to those with

tumors up to 1 cm. However, this was not the case for the single-pollutant CO model. The increase in age at diagnosis has a moderately higher hazard of death (1%) for all years before diagnosis exposure. Patients with adenoma histology type had all-cause mortality hazards that decreased by approximately 11% compared to those with squamous cell carcinomas. Compared to Medicaid alone, patients with Medicare insurance had decreased death hazards by approximately 18% one and three years before diagnosis exposure, respectively.

	Multipollutant			NO_2			SO_2			CO		
	Hazard of death 3 year after diagnosis Duration of exposure from diagnosis			Hazard of death 3 year after diagnosis Duration of exposure from diagnosis			Hazard of death 3 year after diagnosis Duration of exposure from diagnosis			Hazard of death 3 year after diagnosis Duration of exposure from diagnosis		
	1 yr bf	3 yrs. bf	5 yrs. bf	1 yr bf	3 yrs. bf	5 yrs. bf	1 yr bf	3 yrs. bf	5 yrs. bf	1 yr bf	3 yrs. bf	5 yrs. bf
Air pollutants and weather components												
NO_2	1.05***	1.07***	1.09***	1.08***	1.11***	1.13***						
	(1.02, 1.07)	(1.04, 1.1)	(1.06, 1.13)	(1.05, 1.51)	(1.08, 1.83)	(1.1, 5.69)						
SO_2	1.18***	1.2***	1.2***				1.18***	1.2***	1.18***			
	(1.13, 1.24)	(1.15, 1.26)	(1.15, 1.26)				(1.12, 1.23)	(1.14, 1.25)	(1.13, 1.24)			
CO	1.95***	1.98***	1.93***							2.52***	2.96***	3.32***
	(1.42, 2.68)	(1.41, 2.78)	(1.36, 2.74)							(1.9, 3.34)	(2.22, 3.94)	(2.52, 4.37)
Precipitation	0.98**	0.96***	0.96***	0.98**	0.97***	0.97**	0.99	0.99	0.98	0.98***	0.97***	0.97**
	(0.96, 1)	(0.94, 0.98)	(0.93, 0.98)	(.96, 1)	(.95, .99)	(.95, .99)	(.97, 1.01)	(.97, 1.01)	(.96, 1.01)	(.96, 1)	(.95, .99)	(.95, 1)
Snow	0.95	0.90*	0.85***	0.91**	0.82***	0.75***	0.96	0.94	0.90*	0.96	0.88**	0.79***
	(0.86, 1.05)	(0.81, 1.01)	(0.76, 0.95)	(.83, .99)	(.74, .91)	(.67, .83)	(.88, 1.05)	(.85, 1.04)	(.81, 1.01)	(.87, 1.05)	(.79, .98)	(.71, .88)
Daily temperature minimum	1.01	1.01	1.03***	1.01	1.01	1.03***	0.99	.99*	1	1.01	1.01	1.03***
	(1, 1.02)	(1, 1.02)	(1.01, 1.05)	(.99, 1.02)	(1., 1.02)	(1.02, 1.05)	(.98, 1.)	(.98, 1.)	(.98, 1.01)	(1., 1.02)	(1., 1.02)	(1.01, 1.04)
Treatment options (reference :lobectomy)												
Limited resection with adjuvant radiotherapy	0.63	0.55	0.52	0.52	0.45	0.40*	1.07	0.94	0.77	0.66	0.57	0.48
	(0.20, 1.95)	(0.17, 1.79)	(0.15, 1.77)	(.18, 1.48)	(.15, 1.32)	(.14, 1.18)	(.38, 3.02)	(.32, 2.8)	(.25, 2.37)	(.23, 1.91)	(.2, 1.66)	(.17, 1.38)
Treatment interaction with air pollutant and weather components												
NO_2 * Treatment	1.02	1.02**	1.02*	1.02***	1.02***	1.02***						
	(1, 1.05)	(1, 1.05)	(1, 1.05)	(1.01, 1.03)	(1.01, 1.03)	(1.01, 1.03)						
SO_2 * Treatment	0.97	0.97	0.97				0.99	0.99	1			
	(0.91, 1.04)	(0.91, 1.04)	(0.90, 1.04)				(.93, 1.06)	(.93, 1.06)	(.94, 1.06)			
CO * Treatment	0.92	0.89	0.88							1.4***	1.36***	1.34***
	(0.58, 1.47)	(0.56, 1.42)	(0.55, 1.40)							(1.1, 1.78)	(1.08, 1.72)	(1.09, 1.65)
Precipitation * Treatment	1	1.01	1.01	1	1	1	1	1.01	1.01	1	1	1
	(0.99, 1.02)	(0.99, 1.02)	(1, 1.02)	(.99, 1.01)	(.99, 1.01)	(.99, 1.01)	(.99, 1.02)	(.99, 1.02)	(.99, 1.02)	(.99, 1.01)	(.99, 1.01)	(.99, 1.01)
Snow * Treatment	1.03	1.03	1.04	1.04	1.04	1.06	0.99	0.99	1.01	1.04	1.04	1.07
	(0.91, 1.16)	(0.90, 1.18)	(0.90, 1.20)	(.92, 1.16)	(.92, 1.18)	(.93, 1.21)	(.88, 1.11)	(.87, 1.12)	(.88, 1.15)	(.91, 1.17)	(.91, 1.19)	(.93, 1.22)
Temperature minimum* Treatment	1	1	1	1	1	1	1	1	1	1	1	1
	(0.99, 1.01)	(0.99, 1.01)	(0.99, 1.01)	(.99, 1.01)	(.99, 1.01)	(.99, 1.02)	(.99, 1.01)	(.99, 1.01)	(.99, 1.01)	(.99, 1.01)	(.99, 1.01)	(.99, 1.02)

pvalue: * <0.1, ** <0.05, *** <0.01

Table 4.3a Hazards of death three years after diagnosis for air pollutants, weather, and treatment type by annual average monthly mean

Note: Air pollution levels and weather conditions are calculated based on the recording of monitors within 40 miles and 20 miles, respectively, of the centroid of the county of residence; missing values for air pollution and weather up to half of the days of a month are allowed

	Multipollutant Hazard of death 3 year after diagnosis Duration of exposure from diagnosis			NO_2 Hazard of death 3 year after diagnosis Duration of exposure from diagnosis			SO_2 Hazard of death 3 year after diagnosis Duration of exposure from diagnosis			CO Hazard of death 3 year after diagnosis Duration of exposure from diagnosis		
	1 yr bf	3 yrs. bf	5 yrs. bf	1 yr bf	3 yrs. bf	5 yrs. bf	1 yr bf	3 yrs. bf	5 yrs. bf	1 yr bf	3 yrs. bf	5 yrs. bf
Race (reference: Black)												
Other	1.06	1.05	1.07	1.07	1.07	1.08	1.03	1.02	1.02	1.07	1.07	1.09
	(.88 , 1.29)	(.87 , 1.27)	(.88 , 1.29)	(.88 , 1.29)	(.88 , 1.29)	(.89 , 1.31)	(.85 , 1.25)	(.84 , 1.23)	(.84 , 1.23)	(.89 , 1.3)	(.88 , 1.29)	(.9 , 1.32)
White	1	1	1	1.01	1.02	1.02	0.99	0.99	0.99	1.01	1	1.01
	(.88 , 1.14)	(.88 , 1.14)	(.88 , 1.14)	(.89 , 1.15)	(.9 , 1.16)	(.9 , 1.16)	(.88 , 1.13)	(.87 , 1.12)	(.87 , 1.12)	(.89 , 1.14)	(.88 , 1.14)	(.89 , 1.14)
Sex (reference: Female)												
Male	1.18***	1.19***	1.19***	1.19***	1.19***	1.19***	1.18***	1.18***	1.18***	1.19***	1.19***	1.18***
	(1.1 , 1.28)	(1.1 , 1.28)	(1.1 , 1.28)	(1.1 , 1.28)	(1.1 , 1.28)	(1.1 , 1.29)	(1.09 , 1.27)	(1.1 , 1.28)	(1.1 , 1.28)	(1.1 , 1.28)	(1.1 , 1.28)	(1.1 , 1.28)
Tumor Grade (reference: II)												
Grade III	1.16***	1.16***	1.17***	1.15***	1.15***	1.16***	1.17***	1.17***	1.17***	1.15***	1.15***	1.16***
	(1.06 , 1.27)	(1.07 , 1.27)	(1.07 , 1.27)	(1.06 , 1.26)	(1.06 , 1.26)	(1.06 , 1.26)	(1.07 , 1.28)	(1.07 , 1.28)	(1.07 , 1.28)	(1.06 , 1.26)	(1.06 , 1.26)	(1.06 , 1.26)
Grade IV	1.07	1.05	1.06	1.08	1.06	1.06	1.09	1.09	1.09	1.06	1.03	1.02
	(.73 , 1.57)	(.71 , 1.54)	(.72 , 1.54)	(.73 , 1.61)	(.72 , 1.58)	(.72 , 1.57)	(.74 , 1.62)	(.73 , 1.61)	(.74 , 1.61)	(.72 , 1.56)	(.7 , 1.52)	(.7 , 1.5)
Unknown	0.97	0.97	0.98	0.98	0.98	0.98	0.96	0.95	0.96	0.96	0.96	0.96
	(.85 , 1.11)	(.85 , 1.11)	(.85 , 1.12)	(.85 , 1.12)	(.86 , 1.12)	(.86 , 1.12)	(.84 , 1.1)	(.83 , 1.09)	(.84 , 1.09)	(.84 , 1.1)	(.84 , 1.1)	(.84 , 1.1)
Grade I	0.94	0.95	0.95	0.95	0.96	0.96	0.94	0.94	0.94	0.95	0.95	0.96
	(.85 , 1.04)	(.86 , 1.05)	(.86 , 1.06)	(.86 , 1.05)	(.86 , 1.06)	(.87 , 1.06)	(.85 , 1.04)	(.85 , 1.04)	(.85 , 1.04)	(.86 , 1.05)	(.86 , 1.05)	(.86 , 1.06)
Marital status (reference: Divorced)												
Married	0.97	0.98	0.97	0.97	0.97	0.97	0.95	0.95	0.95	0.98	0.98	0.99
	(.87 , 1.1)	(.87 , 1.1)	(.87 , 1.1)	(.87 , 1.09)	(.87 , 1.1)	(.87 , 1.09)	(.85 , 1.07)	(.85 , 1.07)	(.85 , 1.07)	(.87 , 1.1)	(.87 , 1.11)	(.88 , 1.11)
Single	1.01	1.02	1.02	1.01	1.02	1.02	0.98	0.98	0.98	1	1	1.01
	(.87 , 1.18)	(.88 , 1.19)	(.88 , 1.19)	(.87 , 1.17)	(.87 , 1.18)	(.87 , 1.18)	(.84 , 1.14)	(.84 , 1.14)	(.84 , 1.14)	(.86 , 1.16)	(.86 , 1.17)	(.87 , 1.17)
Unknown	1.02	1.03	1.03	1.04	1.04	1.04	1.03	1.02	1.03	1.02	1.02	1.02
	(.83 , 1.26)	(.84 , 1.27)	(.83 , 1.27)	(.84 , 1.27)	(.85 , 1.27)	(.84 , 1.28)	(.83 , 1.27)	(.83 , 1.27)	(.83 , 1.27)	(.83 , 1.25)	(.83 , 1.25)	(.83 , 1.25)
Widowed	1.02	1.02	1.03	1.03	1.02	1.03	1	1	1.01	1.02	1.02	1.03
	(.88 , 1.19)	(.88 , 1.19)	(.89 , 1.2)	(.89 , 1.19)	(.89 , 1.19)	(.89 , 1.19)	(.86 , 1.16)	(.86 , 1.16)	(.87 , 1.17)	(.88 , 1.18)	(.88 , 1.18)	(.89 , 1.19)
Tumor size (reference: upto 1cm)												
>1cm & <=2cm	0.89*	0.88*	0.88*	0.89*	0.89	0.88*	0.89*	0.88*	0.88*	0.9	0.9	0.9
	(.77 , 1.02)	(.77 , 1.01)	(.76 , 1.01)	(.78 , 1.02)	(.78 , 1.02)	(.77 , 1.01)	(.78 , 1.02)	(.77 , 1.01)	(.77 , 1.01)	(.79 , 1.03)	(.79 , 1.04)	(.79 , 1.03)
>2cm	1	1	1	1.01	1	1	1	0.99	0.99	1.01	1.01	1.02
	(.87 , 1.16)	(.86 , 1.16)	(.87 , 1.16)	(.87 , 1.16)	(.87 , 1.16)	(.87 , 1.15)	(.87 , 1.15)	(.86 , 1.14)	(.85 , 1.14)	(.88 , 1.16)	(.88 , 1.17)	(.88 , 1.17)

Unknown	0.68	0.68	0.64	0.71	0.71	0.65	0.68	0.66	0.63	0.57	0.56	0.56
	(.3 , 1.57)	(.29 , 1.61)	(.27 , 1.49)	(.31 , 1.62)	(.3 , 1.68)	(.28 , 1.52)	(.29 , 1.58)	(.28 , 1.53)	(.27 , 1.47)	(.24 , 1.36)	(.23 , 1.38)	(.23 , 1.36)
Tumor histology (reference: squamous cell)												
Adenomas	0.89***	0.89***	0.89***	0.89***	0.89***	0.89***	0.89***	0.89***	0.89***	0.88***	0.88***	0.88***
	(.81 , .97)	(.82 , .97)	(.82 , .97)	(.81 , .97)	(.82 , .97)	(.82 , .97)	(.81 , .97)	(.81 , .97)	(.81 , .97)	(.81 , .96)	(.81 , .96)	(.81 , .96)
Age at diagnosis	1.01***	1.01***	1.01***	1.01***	1.01***	1.01***	1.01***	1.01***	1.01***	1.01***	1.01***	1.01***
	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)
Insurance type (reference: Only Medicaid)												
Only Medicare	0.82*	0.82*	0.85	0.83*	0.83*	0.84*	0.84*	0.83*	0.84*	0.81**	0.80**	0.80**
	(.67 , 1.)	(.68 , 1.)	(.69 , 1.04)	(.69 , 1.02)	(.68 , 1.01)	(.69 , 1.03)	(.69 , 1.02)	(.68 , 1.02)	(.69 , 1.03)	(.66 , .98)	(.65 , .97)	(.66 , .98)
Only private	0.91	0.9	0.93	0.9	0.89	0.91	0.94	0.94	0.95	0.87	0.86	0.88
	(.74 , 1.11)	(.74 , 1.1)	(.76 , 1.13)	(.74 , 1.1)	(.73 , 1.09)	(.74 , 1.11)	(.77 , 1.15)	(.77 , 1.14)	(.78 , 1.16)	(.71 , 1.06)	(.71 , 1.06)	(.72 , 1.08)
Uninsured	1.26	1.31	1.35	1.19	1.2	1.22	1.23	1.23	1.24	1.16	1.19	1.2
	(.75 , 2.09)	(.79 , 2.16)	(.82 , 2.23)	(.72 , 1.96)	(.73 , 1.98)	(.74 , 2.01)	(.75 , 1.99)	(.76 , 2.01)	(.77 , 2.01)	(.7 , 1.92)	(.72 , 1.95)	(.73 , 1.96)
Unknown	1.02	1.05	1.1	0.87	0.91	0.96	1.07	1.12	1.16	0.98	1.01	1.05
	(.66 , 1.58)	(.69 , 1.6)	(.72 , 1.68)	(.55 , 1.37)	(.59 , 1.41)	(.63 , 1.47)	(.69 , 1.66)	(.73 , 1.72)	(.75 , 1.79)	(.64 , 1.51)	(.66 , 1.54)	(.69 , 1.61)
Rural-Urban continuum (reference: Large central metro)												
Large fringe metro	0.8	0.91	0.88	0.98	1.25	1.23	0.54	0.52	0.54	0.9	0.99	0.96
	(.23 , 2.84)	(.24 , 3.45)	(.21 , 3.7)	(.31 , 3.1)	(.4 , 3.9)	(.36 , 4.18)	(.16 , 1.84)	(.15 , 1.78)	(.15 , 1.98)	(.28 , 2.96)	(.29 , 3.42)	(.25 , 3.71)
Medium metro	0.08***	0.05***	0.05***	0.09***	0.07***	0.09***	0.12***	0.12***	0.12***	0.09***	0.09***	0.14***
	(.02 , .26)	(.01 , .18)	(.01 , .19)	(.03 , .3)	(.02 , .25)	(.02 , .36)	(.04 , .39)	(.03 , .4)	(.03 , .45)	(.03 , .3)	(.03 , .32)	(.04 , .52)
Non-metropolitan	0.42*	0.4	0.77	0.54	0.61	1.62	0.15***	0.15***	0.15***	0.34**	0.32**	0.58
	(.15 , 1.18)	(.13 , 1.22)	(.21 , 2.75)	(.19 , 1.51)	(.2 , 1.83)	(.46 , 5.69)	(.06 , .4)	(.05 , .44)	(.05 , .5)	(.13 , .89)	(.11 , .91)	(.18 , 1.87)

pvalue: * <0.1, ** <0.05, *** <0.01

Table 4.3b Hazards of death three years after diagnosis for study covariates by annual average monthly mean

Note: Air pollution levels and weather conditions are calculated based on the recording of monitors within 40 miles and 20 miles, respectively, of the centroid of the county of residence; missing values for air pollution and weather up to half of the days of a month are allowed

4.3.3 HAZARDS OF DEATH FIVE YEARS AFTER DIAGNOSIS

The all-cause mortality hazard of death for those exposed to NO_2 increased by 4%, 6%, and 9% with an average monthly increase of 1 ppb for one, three, and five years before diagnosis exposure, respectively (Table 4.4a). Those exposed to SO_2 had an increase in all-cause mortality hazards by 16% and 17%, with an average increase in monthly averages of 1 ppb for one, three, and five years before diagnosis exposure. Those exposed to CO had an increase in all-cause mortality hazards of 53%, 51%, and 42%, with an average increase in monthly averages of 1 ppb for one, three, and five

years before diagnosis exposure, respectively. Death hazards for those exposed to precipitation decreased by 2% and 3% by an average monthly increase of one-tenth of a millimeter for one, three, and five years before diagnosis exposure, respectively. Similarly, the hazards of death for those exposed to snowfall decreased by 10%, with an average monthly increase of one mm for five years before diagnosis exposure. Death hazards for those exposed to daily minimum temperature increased modestly by 1% and 3%, with an average increase in monthly $1^{\circ}F$ for three and five years before diagnosis exposure, respectively. The hazard effect modestly changed the effect size for single-pollutant models; however, the estimates remained significant. For the SO_2 single-pollutant model, statistical significance no longer remained for hazards of death due to precipitation.

The all-cause mortality hazards of death due to exposure to an average increase of 1 ppb to NO_2 modestly increased by 2% for three and five years before diagnosis exposure, respectively, for patients who received limited resection with adjuvant radiotherapy compared to those who received lobectomy. However, the single-pollutant CO model showed a statistically significant interaction with treatment, with increased mortality hazards. For example, those who received limited resection with adjuvant radiotherapy had higher hazards of death by 10%, 14%, and 11%, respectively, with an increase in average monthly snow by 1 mm. However, this was not the case for single-pollutant NO_2 and CO models.

The all-cause mortality hazards of male deaths increased by approximately 13% compared with females for one, three, and five years before diagnosis exposure (Table 4.4b). The hazard of death increased by approximately 10% for those with tumor grade III compared with those with tumor grade II for one, three, and five years before diagnosis exposure, respectively. Death hazards decreased by approximately 8% for

those with grade I tumors compared with those with grade II tumors. The all-cause mortality hazard of death decreased by approximately 12% for patients with tumors up to 2 cm in size compared to those with tumors up to 1 cm. Compared with Medicaid alone, uninsured patients have increased all-cause mortality hazards of death by approximately 35% for five and three years before diagnosis. However, statistical significance was not achieved in single-pollutant models.

The sensitivity analysis determined a similar effect direction, size, and statistical significance, except for one year after diagnosis. The hazards for the average maximum exposure values for NO_2 and daily minimum temperature were no longer significant (Appendix Tables 4.3-4.8).

	Multipollutant			NO_2			SO_2			CO		
	Hazard of death 5 year after diagnosis Duration of exposure from diagnosis			Hazard of death 5 year after diagnosis Duration of exposure from diagnosis			Hazard of death 5 year after diagnosis Duration of exposure from diagnosis			Hazard of death 5 year after diagnosis Duration of exposure from diagnosis		
	1 yr bf	3 yrs. bf	5 yrs. bf	1 yr bf	3 yrs. bf	5 yrs. bf	1 yr bf	3 yrs. bf	5 yrs. bf	1 yr bf	3 yrs. bf	5 yrs. bf
Air pollutants and weather components												
NO_2	1.04***	1.06***	1.09***	1.06***	1.08***	1.11***						
	(1.02 , 1.06)	(1.04 , 1.08)	(1.06 , 1.12)	(1.04 , 1.29)	(1.06 , 1.68)	(1.08 , 5.82)						
SO_2	1.16***	1.17***	1.17***				1.15***	1.16***	1.15***			
	(1.12 , 1.21)	(1.13 , 1.22)	(1.12 , 1.21)				(1.11 , 1.2)	(1.12 , 1.21)	(1.1 , 1.19)			
CO	1.53***	1.51***	1.42**							1.90***	2.07***	2.32***
	(1.19 , 1.97)	(1.16 , 1.96)	(1.08 , 1.86)							(1.52 , 2.38)	(1.65 , 2.6)	(1.86 , 2.9)
Precipitation	0.98**	0.97***	0.97**	.98**	.98***	0.98	1	1	1	.99*	.98**	0.99
	(0.97 , 1)	(0.95 , 0.99)	(0.95 , 1)	(.97 , 1.)	(.96 , .99)	(.96 , 1.01)	(.98 , 1.01)	(.98 , 1.01)	(.98 , 1.02)	(.97 , 1.)	(.96 , 1.)	(.97 , 1.01)
Snow	0.99	0.96	0.90**	0.94	.88***	.82***	1	1.01	0.99	1	0.94	.88***
	(0.92 , 1.07)	(0.88 , 1.05)	(0.82 , 0.99)	(.87 , 1.01)	(.81 , .96)	(.75 , .89)	(.93 , 1.08)	(.93 , 1.1)	(.9 , 1.08)	(.93 , 1.07)	(.87 , 1.03)	(.8 , .96)
Daily temperature minimum	1.01	1.01**	1.03***	1.01	1.01**	1.03***	.99**	.99**	1	1.01	1.01	1.02***
	(1 , 1.02)	(1 , 1.02)	(1.02 , 1.04)	(1. , 1.01)	(1. , 1.02)	(1.02 , 1.05)	(.99 , 1.)	(.98 , 1.)	(.99 , 1.01)	(1. , 1.01)	(1. , 1.02)	(1.01 , 1.03)
Treatment options (reference :lobectomy)												
Limited resection with adjuvant radiotherapy	0.95	0.89	0.97	0.70	0.63	0.67	1.34	1.24	1.14	0.93	0.79	0.75
	(0.39 , 2.32)	(0.35 , 2.22)	(0.37 , 2.52)	(.31 , 1.57)	(.28 , 1.43)	(.29 , 1.54)	(.55 , 3.22)	(.5 , 3.08)	(.45 , 2.88)	(.4 , 2.16)	(.34 , 1.83)	(.32 , 1.72)
Treatment interaction with air pollutant and weather components												
NO_2 * Treatment	1.01	1.02*	1.02*	1.01*	1.01*	1.01***						
	(1 , 1.03)	(1 , 1.03)	(1 , 1.03)	(1 , 1.02)	(1 , 1.02)	(1 , 1.02)						
SO_2 * Treatment	0.99	0.98	0.99				1.02	1.02	1.02			
	(0.93 , 1.04)	(0.93 , 1.04)	(0.93 , 1.05)				(0.97 , 1.07)	(0.98 , 1.07)	(0.97 , 1.06)			
CO * Treatment	0.94	0.85	0.86							1.16	1.24**	1.36***
	(0.68 , 1.29)	(0.60 , 1.21)	(0.60 , 1.22)							(0.95 , 1.43)	(1.04 , 1.48)	(1.16 , 1.60)
Precipitation * Treatment	1	1.01	1.01*	1.00	1.00	1.00	1	1.01*	1.01**	1.00	1.00	1.00

	(0.99 , 1.01)	(1 , 1.02)	(1 , 1.02)	(0.99 , 1)	(0.99 , 1.01)	(0.99 , 1.01)	(1 , 1.01)	(1 , 1.02)	(1 , 1.02)	(0.99 , 1.01)	(0.99 , 1.01)	(0.99 , 1)
Snow * Treatment	1.10**	1.14***	1.11**	1.03	1.04	1.00	1.09**	1.10**	1.06	1.03	1.06	1.05
	(1 , 1.2)	(1.03 , 1.25)	(1.01 , 1.23)	(0.96 , 1.10)	(0.97 , 1.12)	(0.93 , 1.07)	(1 , 1.18)	(1 , 1.2)	(0.97 , 1.17)	(0.95 , 1.12)	(0.98 , 1.14)	(0.97 , 1.13)
Temperature minimum * Treatment	1.00	1.01*	1.01*	1	1	1	1.01**	1.01**	1.01*	1	1	1
	(1 , 1.01)	(1 , 1.02)	(1 , 1.02)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)

pvalue: * <0.1, ** <0.05, *** <0.01

Table 4.4a Hazards of death five years after diagnosis from air pollution, weather, and treatment type by annual average monthly *mean*

Note: Air pollution levels and weather conditions are calculated based on the recording of monitors within 40 miles and 20 miles, respectively, of the centroid of the county of residence; missing values for air pollution and weather up to half of the days of a month are allowed

	Multipollutant Hazard of death 5 year after diagnosis Duration of exposure from diagnosis			NO_2 Hazard of death 5 year after diagnosis Duration of exposure from diagnosis			SO_2 Hazard of death 5 year after diagnosis Duration of exposure from diagnosis			CO Hazard of death 5 year after diagnosis Duration of exposure from diagnosis		
	1 yr bf	3 yrs. bf	5 yrs. bf	1 yr bf	3 yrs. bf	5 yrs. bf	1 yr bf	3 yrs. bf	5 yrs. bf	1 yr bf	3 yrs. bf	5 yrs. bf
Race (reference: Black)												
Other	1	1	1.01	1.02	1.02	1.03	0.99	0.98	0.98	1.02	1.01	1.02
	(.87 , 1.16)	(.86 , 1.15)	(.87 , 1.16)	(.88 , 1.18)	(.88 , 1.18)	(.89 , 1.19)	(.86 , 1.14)	(.85 , 1.13)	(.85 , 1.13)	(.88 , 1.17)	(.88 , 1.17)	(.88 , 1.18)
White	0.97	0.96	0.97	0.98	0.98	0.99	0.96	0.95	0.95	0.97	0.97	0.97
	(.88 , 1.07)	(.88 , 1.06)	(.88 , 1.07)	(.89 , 1.08)	(.89 , 1.08)	(.9 , 1.09)	(.87 , 1.06)	(.87 , 1.05)	(.86 , 1.05)	(.88 , 1.07)	(.88 , 1.07)	(.88 , 1.07)
Sex (reference: Female)												
Male	1.12***	1.12***	1.13***	1.12***	1.12***	1.13***	1.11***	1.11***	1.11***	1.12***	1.12***	1.12***
	(1.05 , 1.19)	(1.05 , 1.19)	(1.06 , 1.2)	(1.05 , 1.19)	(1.06 , 1.19)	(1.06 , 1.2)	(1.04 , 1.17)	(1.04 , 1.18)	(1.04 , 1.18)	(1.05 , 1.19)	(1.05 , 1.19)	(1.05 , 1.19)
Tumor Grade (reference: II)												
Grade III	1.1***	1.1***	1.1***	1.09**	1.09**	1.09**	1.12***	1.12***	1.12***	1.09**	1.09**	1.1**
	(1.02 , 1.19)	(1.02 , 1.19)	(1.02 , 1.19)	(1.01 , 1.18)	(1.01 , 1.17)	(1.01 , 1.18)	(1.04 , 1.2)	(1.04 , 1.2)	(1.04 , 1.2)	(1.01 , 1.18)	(1.02 , 1.18)	(1.02 , 1.18)
Grade IV	1	0.99	1	0.98	0.97	0.97	1.01	1.01	1.02	0.96	0.95	0.95
	(.72 , 1.39)	(.71 , 1.37)	(.72 , 1.39)	(.7 , 1.38)	(.69 , 1.36)	(.68 , 1.37)	(.72 , 1.41)	(.72 , 1.42)	(.72 , 1.42)	(.68 , 1.37)	(.67 , 1.35)	(.67 , 1.34)
Unknown	0.95	0.94	0.94	0.95	0.95	0.95	0.94	0.94	0.94	0.94	0.94	0.94
	(.85 , 1.06)	(.85 , 1.05)	(.85 , 1.06)	(.85 , 1.06)	(.85 , 1.06)	(.85 , 1.06)	(.84 , 1.05)	(.84 , 1.05)	(.84 , 1.04)	(.85 , 1.05)	(.84 , 1.05)	(.84 , 1.04)
Grade I	0.92**	0.92**	0.93**	0.93*	0.93*	0.93*	0.92**	0.93**	0.93*	0.93*	0.93*	0.93*
	(.85 , 1.)	(.86 , 1.)	(.86 , 1.)	(.86 , 1.)	(.86 , 1.)	(.86 , 1.)	(.86 , 1.)	(.86 , 1.)	(.86 , 1.)	(.86 , 1.)	(.86 , 1.)	(.86 , 1.)
Marital status (reference: Divorced)												
Married	0.96	0.96	0.96	0.96	0.96	0.96	0.95	0.95	0.95	0.96	0.96	0.97
	(.88 , 1.05)	(.88 , 1.06)	(.88 , 1.06)	(.88 , 1.05)	(.88 , 1.06)	(.88 , 1.06)	(.86 , 1.04)	(.86 , 1.04)	(.86 , 1.04)	(.88 , 1.05)	(.88 , 1.06)	(.88 , 1.06)
Single	0.98	0.98	0.98	0.98	0.98	0.99	0.96	0.95	0.95	0.96	0.97	0.97
	(.87 , 1.1)	(.87 , 1.1)	(.87 , 1.1)	(.87 , 1.1)	(.87 , 1.11)	(.88 , 1.11)	(.85 , 1.08)	(.85 , 1.07)	(.84 , 1.07)	(.85 , 1.08)	(.86 , 1.09)	(.86 , 1.09)
Unknown	0.98	0.99	0.99	0.99	1	1	0.98	0.98	0.98	0.97	0.98	0.98
	(.84 , 1.15)	(.85 , 1.16)	(.85 , 1.16)	(.84 , 1.16)	(.85 , 1.16)	(.85 , 1.16)	(.84 , 1.15)	(.83 , 1.15)	(.83 , 1.14)	(.83 , 1.14)	(.84 , 1.14)	(.84 , 1.14)
Widowed	0.98	0.98	0.99	0.98	0.99	0.99	0.96	0.96	0.96	0.97	0.98	0.98

	(.87 , 1.1)	(.88 , 1.11)	(.88 , 1.12)	(.88 , 1.11)	(.88 , 1.11)	(.88 , 1.11)	(.85 , 1.08)	(.85 , 1.08)	(.86 , 1.08)	(.87 , 1.09)	(.87 , 1.1)	(.87 , 1.1)
Tumor size (reference: upto 1cm)												
>1cm & <=2cm	0.99	0.99	0.99	1	1	0.99	0.99	0.99	0.99	1.01	1.01	1
	(.89 , 1.1)	(.89 , 1.1)	(.89 , 1.1)	(.9 , 1.11)	(.9 , 1.11)	(.89 , 1.1)	(.9 , 1.1)	(.89 , 1.09)	(.89 , 1.1)	(.91 , 1.12)	(.91 , 1.12)	(.9 , 1.11)
>2cm	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.01	1.02	1.03	1.02	1.03
	(.91 , 1.14)	(.91 , 1.14)	(.91 , 1.15)	(.91 , 1.14)	(.91 , 1.14)	(.91 , 1.14)	(.91 , 1.14)	(.91 , 1.13)	(.91 , 1.13)	(.92 , 1.15)	(.92 , 1.15)	(.92 , 1.15)
Unknown	0.87	0.85	0.81	0.88	0.86	0.8	0.84	0.82	0.81	0.76	0.75	0.75
	(.49 , 1.55)	(.47 , 1.57)	(.45 , 1.47)	(.49 , 1.56)	(.47 , 1.57)	(.45 , 1.45)	(.48 , 1.47)	(.47 , 1.44)	(.46 , 1.42)	(.42 , 1.38)	(.41 , 1.39)	(.41 , 1.36)
Tumor histology (reference: squamous cell)												
Adenomas	0.94	0.95	0.94	0.94*	0.94	0.94	0.94*	0.94	0.94	0.93*	0.93*	0.93*
	(.87 , 1.01)	(.88 , 1.02)	(.88 , 1.02)	(.87 , 1.01)	(.87 , 1.01)	(.87 , 1.01)	(.87 , 1.01)	(.87 , 1.01)	(.87 , 1.01)	(.87 , 1.01)	(.87 , 1.01)	(.87 , 1.01)
Age at diagnosis	1.01***	1.01***	1.01***	1.01***	1.01***	1.01***	1.01***	1.01***	1.01***	1.01***	1.01***	1.01***
	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)	(1 , 1.01)
Insurance type (reference: Only Medicaid)												
Only Medicare	0.93	0.93	0.95	0.94	0.94	0.95	0.94	0.94	0.95	0.92	0.92	0.92
	(.81 , 1.07)	(.81 , 1.08)	(.82 , 1.1)	(.82 , 1.08)	(.82 , 1.08)	(.82 , 1.09)	(.82 , 1.09)	(.82 , 1.09)	(.83 , 1.1)	(.8 , 1.06)	(.79 , 1.06)	(.8 , 1.06)
Only private	0.97	0.97	0.99	0.97	0.96	0.97	1	0.99	1	0.95	0.94	0.95
	(.84 , 1.12)	(.84 , 1.12)	(.85 , 1.14)	(.84 , 1.11)	(.84 , 1.11)	(.84 , 1.12)	(.87 , 1.15)	(.86 , 1.14)	(.87 , 1.15)	(.82 , 1.09)	(.82 , 1.09)	(.82 , 1.1)
Uninsured	1.27	1.31*	1.35*	1.17	1.19	1.22	1.22	1.23	1.22	1.15	1.17	1.2
	(.92 , 1.76)	(.95 , 1.81)	(.98 , 1.86)	(.85 , 1.61)	(.87 , 1.64)	(.89 , 1.67)	(.89 , 1.68)	(.9 , 1.69)	(.89 , 1.68)	(.83 , 1.58)	(.86 , 1.61)	(.87 , 1.64)
Unknown	1.05	1.07	1.11	0.93	0.96	0.99	1.07	1.09	1.11	0.98	0.98	1.02
	(.8 , 1.37)	(.82 , 1.4)	(.84 , 1.45)	(.71 , 1.21)	(.73 , 1.25)	(.76 , 1.29)	(.83 , 1.39)	(.84 , 1.41)	(.86 , 1.44)	(.74 , 1.28)	(.75 , 1.29)	(.77 , 1.34)
Rural-Urban continuum (reference: Large central metro)												
Large fringe metro	0.84	0.93	0.99	0.98	1.16	1.24	0.57	0.55	0.62	0.86	0.91	0.95
	(.26 , 2.67)	(.28 , 3.12)	(.28 , 3.56)	(.34 , 2.84)	(.4 , 3.38)	(.41 , 3.78)	(.19 , 1.72)	(.18 , 1.66)	(.2 , 1.89)	(.29 , 2.58)	(.3 , 2.8)	(.29 , 3.12)
Medium metro	0.10***	0.07***	0.09***	0.11***	0.10***	0.14***	0.16***	0.15***	0.20***	0.12***	0.12***	0.20***
	(.04 , .27)	(.03 , .2)	(.03 , .27)	(.04 , .3)	(.03 , .28)	(.05 , .42)	(.06 , .4)	(.06 , .4)	(.07 , .56)	(.05 , .31)	(.05 , .34)	(.07 , .58)
Non-metropolitan	0.45*	0.46	1.14	0.53	0.64	1.92	0.23***	0.24***	0.32**	0.37**	0.36**	0.70
	(.18 , 1.1)	(.17 , 1.21)	(.37 , 3.53)	(.21 , 1.29)	(.24 , 1.68)	(.64 , 5.82)	(.1 , .56)	(.09 , .61)	(.11 , .95)	(.15 , .87)	(.14 , .91)	(.24 , 2.02)

pvalue: * <0.1, ** <0.05, *** <0.01

Table 4.4b Hazards of death five years after diagnosis for study covariates by annual average monthly mean

Note: Air pollution levels and weather conditions are calculated based on the recording of monitors within 40 miles and 20 miles, respectively, of the centroid of the county of residence; missing values for air pollution and weather up to half of the days of a month are allowed

4.4 DISCUSSION

The present study found that patients exposed to higher concentrations of NO_2, SO_2, and CO ambient air pollution before diagnosis had decreased survival after diagnosis. The results are consistent with the results of a study by Eckel et al. 2016 regarding estimate direction for air pollutant NO_2.[6] It has also been determined that snowfall and precipitation decrease death events after diagnosis, which aligns with the

logic that ambient air pollution concentration is lower during snowfall. Finally, patients who underwent limited resection had modest significant higher hazards with increased ambient air pollution, possibly due to early air pollution exposure after surgery, than those who underwent lobectomy due to longer indoor post-surgery recovery time due to invasive procedure. Those] Those who underwent limited resection with adjuvant radiotherapy were less physically fit than those who underwent lobectomy. It might be the case that these patients already have higher comorbidities, so future studies might want to adjust for key confounders such as patient functional status, comorbidity score, and surgeon expertise to confirm the treatment interaction hazards with air pollutants. Although our study is the first of its kind and no relevant studies exist, other studies examining different health outcomes in the presence of air pollution exposure and survival outcomes in the absence of air pollution are present. These findings align with the existing literature[100,101] and claim that lobectomy has increased surgery-associated morbidity post-lobectomy, translating into longer hospital/indoor stays and less exposure to ambient air pollutants. Higher ambient air pollutants also affect lung function, as per a recent study which translates into increased death hazards.[22] The survival outcome difference between lobectomy and limited resection is similar as per some literature, while other studies find no statistical difference.[54,55]

The present study has several strengths as it utilizes key primary air pollutants such as SO_2, CO, and weather components such as precipitation, snowfall, and daily minimum temperature to account for the confounding effects. Ozone and daily temperature maximum pose multicollinearity problems due to their inherent correlation with primary pollutants,[96] so they were excluded from the analysis. This exclusion aligns with Eckel et al. 2016 study[6] findings that ozone had a non-significant effect on survival outcomes, possibly because of multicollinearity. In addition, the study

evaluated the effects of air pollutants and weather components before and after diagnosis exposure assignments to determine their cumulative effects.

The key limitations of this study include significant missing data for primary air pollutants, Particulate Matter (PM), benzene, lead, wind, and daily average temperature. For the same reason, we could not determine their interaction with other pollutants, their overall effect on survival outcomes, and their interaction with treatment types. Moreover, data availability for air pollutant stations less than 30 miles, weather stations less than 10 miles, and less than 33.33% missing values in a month was significantly lower. This led to use of the most lenient distance and missing percentage model. However, it can be argued that the present results might not change with respect to the hazard rate in the presence of the nearest stations, as measurement errors are randomly distributed and not spatially correlated.

Another limitation is the insufficient sample size for radiotherapy and limited resection. Therefore, determining the actual hazard rate using these treatment categories is difficult. In addition, the AHRF had significant missing values for area-level information relevant to the study, which could not be controlled for in the analysis. However, the county level and year of diagnosis dummy variables address these limitations for time-invariant unobserved variables. Some of the missing contextual variables that could help reduce estimation bias were comorbidity score, cardiopulmonary function, lung function, smoking status, hospital region, patient's overall functional status, occupation, and surgeon expertise. Patients who smoked survived significantly fewer years than non-smokers, and smokers tended to live in areas with higher air pollution. The biased estimates derived in the absence of this variable could lead to underestimation. Similarly, hospital regions seem to be negatively correlated with air pollution and weather exposure, given the greater

likelihood of high-volume hospital regions being present in metropolitan areas. The hospital region seems to be positively correlated with survival outcomes, given the higher likelihood of hospitals in metropolitan areas with advanced medical technologies/expertise leading to better survival. The biased estimates determined in the absence of the variable could lead to underestimation. This could lead to an underestimation of the determined effect in the absence of the variable. Patient functional status and cardiopulmonary function are other variables that seem to be negatively correlated with air pollution and weather exposure. However, they appear to be positively correlated with survival outcomes. In the absence of these variables, the derived biased estimates are underestimated. Hence, it was vital to account for key confounders in the present study. For the same reason, our study only measured associational relationships because we did not account for these identified unobserved confounders in the analysis, nor was the study designed to be a randomized control trial or natural experiment.

Moreover, the results of the current study are only generalizable to the population representative of the sub-sample. As most monitors are present in metropolitan areas, potentially due to higher pollution levels, the results from the present study cannot be generalized to population outcomes in rural areas. The lack of sufficient monitoring stations in non-metropolitan areas necessitates future studies to focus on these areas. Ecological fallacy persists as county level exposure values are assignment to inviduals. Additionally, the sub-sample lacks sufficient Black individuals, potentially because of their higher presence in non-metropolitan areas. Hence, the study results are highly generalizable to white groups. Finally, the SEER 18 Research Plus registry data lack information on key confounders, such as comorbidity status, lung function, patient's overall functioning status, timing and dosage of

treatment, migration status, and tumor recurrence. Furthermore, the study results are only generalizable to a population with stage 1A NSCLC first primary only, with age at diagnosis between 18 and 80 years, specifically with tumor histology adenomas, adenocarcinomas, and squamous cell carcinomas.

4.5 IMPLICATIONS FOR PRACTICE AND POLICY

The survival of patients with stage 1A NSCLC is negatively associated with increased concentrations of ambient air pollutants such as NO_2, SO_2, CO, and daily minimum temperature. Hence, it is vital to implement environmental policies that control the emission or source of emission to reduce preventable deaths in stage 1A NSCLC patients with adenocarcinoma or squamous cell carcinoma histology types and other cardiopulmonary patients residing in metropolitan areas. Furthermore, future treatment decision-making might also consider treatment modalities, given the presence of local air pollution and weather factors, to determine the best patient survival outcomes and provide more patient-centric care if cancer registry data limitations are overcome.

4.6 FUTURE RESEARCH DIRECTIONS

Future studies might want to consider the interaction of particulate matter like $PM_{2.5}$ and PM_{10}, weather components such as wind, existing primary pollutants, weather components, benzene, and lead, and their overall effect on survival outcomes to determine their interaction with one another in a multi-pollutant model, as well as the overall exposure effect on survival outcomes. Another data technique could be the use of kriging or accurate spatiotemporal interpolation to compensate for significant missing exposure values, along with mobility data to determine close-to-true exposure. In addition, a similar relationship can be tested using other national cancer registry

databases, such as the National Cancer Database, or medical claims data, such as each state Medicare/Medicaid claims data, that provide information on key confounders, such as frailty, treatment timing, hospital region, hospital type, treatment dosage, patient migration status, and overall patient functional status.

CHAPTER 5: CONCLUSION

5.1 SUMMARY

This dissertation attempts to fill the empirical research gap in the topic context. The lack of empirical studies determining comprehensive causal pathways through evidence synthesis Directed Acyclic Graphs (DAG) for factors affecting treatment receipt and survival outcomes among patients with stage 1A Non-Small Cell Lung Cancer (NSCLC) in the U.S. to understand the front and back door pathways introduces the need for this study. DAGs inform us of all possible factors that act as confounders, mediators, or colliders in a given study context that one could evaluate in statistical modeling for a given study context. Moreover, advanced data techniques could facilitate the inclusion of certain key variables absent in one database that helps determine the comprehensive associational relationship between exposure and outcome in an attempt to reduce selection bias. This study attempts to account for this fact and to reduce selection bias by including key factors affecting treatment receipt and survival outcomes. Although not all potential confounders in the study topic were accounted for owing to data limitations of Survival, Epidemiology, and End Results (SEER), Area Health Resource Files (AHRF), and the United States (U.S.) environmental data, the dissertation results can help future research build on these results when data limitations are overcome. The advanced data techniques developed in this study for multiple longer

period big data merging and exposure assignment could help future epidemiological studies build on the present study's techniques and evaluate other health outcomes.

The first chapter helped to identify factors that affect treatment receipt and survival outcomes among patients with stage 1A TN0M0 NSCLC first primary tumor in the U.S.. The chapter employed an evidence-based synthesis protocol to construct integrated DAGs (iDAG) informed by a thorough systematic review and further proposed DAGs that could inform model decision-making for Chapters Two and Three. The identified confounders affecting treatment receipt were comorbidity score, education, age, insurance status, tumor size, sex, patient preference, physician preference, presence of Chronic Obstructive Pulmonary Disease (COPD), marital status, income, health status, coronary artery disease, race, geographic region, and type of treatment facility. The identified confounders affecting survival outcomes were age, sex, tumor histology, tumor size, access to care, comorbidities, lung function, cardiopulmonary function, tumor grade, quality of life, adjuvant therapy, treatment facility type, patient functional status, smoking status, race, year of diagnosis, insurance status, tumor markers, and surgeon expertise. Based on the respective iDAG, we proposed factors that affect treatment receipt and survival outcomes in the given study context. The factors that affected lobectomy treatment receipt were treatment guideline revision years, patient preference, doctor recommendation, race, tumor grade, age, sex, insurance status, tumor size, marital status, and rural-urban continuum. The factors that affected survival outcomes were air pollution, weather, treatment type, marital status, age, tumor histology, sex, insurance status, tumor grade, tumor size, race, and rural-urban continuum. Both iDAGs helped inform the statistical model decision-making in Chapters Two and Three.

The second chapter identified factors affecting treatment receipt among cohort-assigned environmental exposure values within 40 miles of air pollution, 20 miles of weather nearest monitor stations (sub-sample), and up to 50% monthly missing exposure values and without an exposure assignment larger study sample. The statistical model for the chapter was adjusted for the variables identified in Chapter One to determine the associational relationship. However, due to data limitations not, all the identified confounders were accounted for as identified in iDAG. The data analysis attempted to account for time-invariant confounders from the year of diagnosis and county-specific factors, adjusting for standard errors at the state level. Overall, the study confirms that the factors identified in the iDAG of Chapter One affect lobectomy treatment receipt, including race, tumor grade, marital status, insurance status, rural-urban continuum, treatment guideline revision years, age, and tumor size. Black individuals are less likely to receive guideline-concordant care compared to white individuals. Similarly, Medicaid beneficiaries are less likely to receive guideline-concordant care (*i.e.,* lobectomy) than Medicare beneficiaries. Patients living in large fringe metros are less likely to undergo lobectomy than those living in a large central metro. Hence, it is evident that there is a difference in guideline-concordant care receipt among stage 1A TN0M0 NSCLC patients receiving either lobectomy or limited resection with adjuvant radiotherapy, for whom lobectomy is the first recommended treatment among medically fit candidates.

This chapter also analyzes the study sub-sample used in Chapter Three to determine if a compositional effect affects study estimates, although the selection of the sub-sample was completely outside investigator control, as it was solely guided by the presence of exposure values for the cohort. The sub-sample differed from the study sample in several aspects, including differences in the proportion of Black and reduced

non-metropolitan observations in the sub-sample. The sub-sample represented metropolitan residing groups; hence, the results from Chapter Three are only generalizable to the population residing in metropolitan areas because of the higher presence of monitoring stations in those areas than in non-metropolitan areas.

The third chapter identifies factors that affect survival outcomes in the presence of air pollution and weather by analyzing the sub-sample cohort utilized in Chapter Two. We accounted for the key factors that determined the factors affecting survival outcomes. Air pollution and weather affect the survival of patients with stage 1A TN0M0 NSCLC in the U.S., which aligns with the results of a similar study in the study area. Those with exposure assignment from within 40 miles air pollution, 20 miles weather nearest monitoring stations, and up to 50% average monthly mean missing exposure values comprised the analysis sample for Chapter Three. Through descriptive statistics, it was determined that because of the higher presence of monitoring stations in metropolitan areas than in non-metropolitan areas, the study results of Chapter Three are only generalizable to them, given the included observations represented it. Population migration might logically affect environmental exposure; hence, the study analysis assumed that the study observations did not migrate during the study period. For the same reason, the rationale behind restricting the sub-sample to ten years of exposure period *(i.e.,* five years before diagnosis exposure and five years after diagnosis exposure) helped us reduce the higher chances of migration had the sample inclusion longer than ten years. The results from the analysis seem to be robust and confirm that air pollutants and weather affect the survival outcomes of stage 1A TN0M0 NSCLC first primary patients. Additionally, among the factors identified in the proposed DAG (Figure 3b), sex, tumor grade, histology, age, and insurance status affected survival outcomes in the given study context.

The three chapters help identify the clinical and environmental factors affecting treatment receipt and survival outcomes among stage 1A TN0M0 NSCLC patients in the U.S.. The factors that affect treatment receipt and survival outcome are determined by evidence-based iDAGs and developed DAGs in Chapter One, utilizing Andersen and Aday's model for Chapter Two, and Shi and Steven's model for Chapter Three. The variables for the statistical models in Chapters Two and Three were informed by the respective iDAGs proposed in Chapter One. The chapters jointly determine treatment and survival outcome difference among stage 1A TN0M0 NSCLC patients in the U.S. when nationally representative population cancer registry SEER 18 Research Plus data were utilized for analysis along with AHRF and environmental data. The results of this study align with the existing literature in the field.

5.2 POLICY IMPLICATIONS

Although clinical trials have tested stage 1A NSCLC guideline-concordant treatment options, there is a translational gap in real-world effectiveness. It includes differences in treatment and survival outcomes that remain unaccounted for in studies such as clinical trials that test the efficacy of treatments. Moreover, as reflected in Chapter One, past literature lacks accounting for all key exposure factors, specifically for stage 1A NSCLC primary tumors. One of the future policy implications could be an effort to improve existing monitors' functional efficiency and install more monitors in non-metropolitan areas to help determine the dose–response relationship of air pollution and weather on health outcomes. Future policies might be targeted towards the control of ambient air pollution, specifically near residential areas where people are treated/diagnosed with stage 1A NSCLC first primary tumors and have compromised lung function with higher comorbidity scores. In addition, regarding stage 1A NSCLC

treatment guideline policies, future revisions might consider inculcating air pollution trends in treatment decision-making to improve survival outcomes and patient-centric care. Finally, public databases such as the national cancer registry and weather and air pollution databases might consider improving data limitations by providing information on individual patient migration, treatment guideline revision years, and fewer missing air pollution values to avoid measurement errors.

5.3 FUTURE RESEARCH

This study has several limitations that future research should address. For example, patient-relevant information, such as migration information, tumor recurrence, lymph node-associated information (although for the N0 stage, it might be irrelevant), treatment receipt timings, treatment dosage, travel distance to treatment receipt center, physician-patient preferences, and patient functional status/scores might be some of the key variables. Air pollution data accuracy and the availability of nearest monitoring data might help obtain more accurate results.

The air pollution exposure assignments are averaged at the population level, and a difference exists between population and personal exposure levels. However, these measurement errors are random and consistent, and their effects undermine the estimates towards the null hypothesis. Future studies might consider exposure assignment at a more granular level (*i.e.*, personal or zip code) and account for events that cause an increase in local air pollution levels.

Studies that want to build on existing study results might want to consider exposure assignments for nearest miles less than 40 miles and fewer missing value percentages (*i.e.*, < 50% for air pollution monitors and weather stations) if and when accurate data are available for longer study periods, as in the present study. Although

the average monitoring distance ranges from to 4-15 miles, it is vital to corroborate the result robustness if further nearest monitor (probably ≤ 1 mi) values are utilized. Studies might also want to inculcate other pollutants such as particulate matter, benzene, lead, and weather components such as wind, and build on existing results for the interaction between pollutants and treatment types. Finally, the studies might want to consider estimating the effect by including other air pollutants and weather measures, such as standard deviation and interquartile range, to confirm the robustness of existing findings.

REFERENCES

1. Aday L, Begley C, Lairson D, Slater C. *Evaluating the Healthcare System: Effectiveness, Efficiency, and Equity*. 3rd ed. Health Administration Press; 2004.
2. Shi L. *Health Services Research Methods*. Thomson/Delmar Learning; 2008.
3. Slatore CG, Au DH, Gould MK. An Official American Thoracic Society Systematic Review: Insurance Status and Disparities in Lung Cancer Practices and Outcomes. *Am J Respir Crit Care Med*. 2010;182(9):1195-1205. doi:10.1164/rccm.2009-038ST
4. Goldstraw P, Chansky K, Crowley J, et al. The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. *J Thorac Oncol*. 2016;11(1):39-51. doi:10.1016/j.jtho.2015.09.009
5. Abdel-Rahman O. Role of postoperative radiotherapy in the management of malignant pleural mesothelioma. *Strahlentherapie und Onkol*. 2017;193(4):276-284. doi:10.1007/s00066-016-1092-7
6. Eckel SP, Cockburn M, Shu YH, et al. Air pollution affects lung cancer survival. *Thorax*. 2016;71(10):891-898. doi:10.1136/thoraxjnl-2015-207927
7. Huesch MD, Ong MK. Lung Cancer Care Before and After Medicare Eligibility. *J Heal Care Organ Provision, Financ*. 2016;53(1):004695801664730. doi:10.1177/0046958016647301
8. Boyer MJ, Williams CD, Harpole DH, Onaitis MW, Kelley MJ, Salama JK. Improved Survival of Stage I Non–Small Cell Lung Cancer: A VA Central Cancer Registry Analysis. *J Thorac Oncol*. 2017;12(12):1814-1823. doi:10.1016/j.jtho.2017.09.1952
9. Sheppard L, Burnett RT, Szpiro AA, et al. Confounding and exposure measurement error in air pollution epidemiology. *Air Qual Atmos Heal*. 2012;5(2):203-216. doi:10.1007/s11869-011-0140-9
10. Bowe B, Xie Y, Yan Y, Al-Aly Z. Burden of Cause-Specific Mortality Associated With PM 2.5 Air Pollution in the United States. *JAMA Netw Open*. 2019;2(11):e1915834. doi:10.1001/jamanetworkopen.2019.15834
11. Chao X, Yi L, Lan LL, Wei HY, Wei D. Long-term PM2.5 exposure increases the risk of non-small cell lung cancer (NSCLC) progression by enhancing interleukin-17a (IL-17a)-regulated proliferation and metastasis. *Aging (Albany NY)*. 2020;12(12):11579-11602. doi:10.18632/aging.103319
12. Baxter LK, Dionisio KL, Burke J, et al. Exposure prediction approaches used in air pollution epidemiology studies: Key findings and future recommendations. *J Expo Sci Environ Epidemiol*. 2013;23(6):654-659. doi:10.1038/jes.2013.62
13. American Lung Association. State Data - Kentucky . American Lung Association. Published 2021. Accessed December 12, 2021. https://www.lung.org/research/state-of-lung-cancer/states/kentucky
14. Lung Cancer Key Findings. American Lung Association. Accessed June 10, 2023. https://www.lung.org/research/state-of-lung-cancer/key-findings
15. Kullgren JT, McLaughlin CG, Mitra N, Armstrong K. Nonfinancial barriers and access to care for U.S. adults. *Health Serv Res*. 2012;47(1 PART 2):462-485. doi:10.1111/j.1475-6773.2011.01308.x
16. Dawes AJ, Louie R, Nguyen DK, et al. The impact of continuous medicaid enrollment on diagnosis, treatment, and survival in six surgical cancers. *Health*

Serv Res. 2014;49(6):1787-1811. doi:10.1111/1475-6773.12237
17. State of Lung Cancer Key Findings . American Lung Association. Published 2021. Accessed January 1, 2022. https://www.lung.org/research/state-of-lung-cancer/key-findings
18. Evans WK, Wolfson MC, Flanagan WM, et al. Canadian cancer risk management model: Evaluation of cancer control. *Int J Technol Assess Health Care*. 2013;29(2):131-139. doi:10.1017/S0266462313000044
19. Jiang X, Lin G, Islam KMM. Socioeconomic factors related to surgical treatment for localized, non-small cell lung cancer. *Soc Sci Med*. 2017;175:52-57. doi:10.1016/j.socscimed.2016.12.042
20. Vest MT, Herrin J, Soulos PR, et al. Use of new treatment modalities for non-small cell lung cancer care in the medicare population. *Chest*. 2013;143(2):429-435. doi:10.1378/chest.12-1149
21. Keenan RJ, Landreneau RJ, Maley RH, et al. Segmental resection spares pulmonary function in patients with stage I lung cancer. *Ann Thorac Surg*. 2004;78(1):228-233. doi:10.1016/j.athoracsur.2004.01.024
22. Lee KK, Bing R, Kiang J, et al. Articles Adverse health effects associated with household air pollution : a systematic review , meta-analysis , and burden estimation study. *Lancet Glob Heal*. 2020;8(11):e1427-e1434. doi:10.1016/S2214-109X(20)30343-0
23. Dalwadi SM, Zhang J, Bernicker EH, Butler EB, Teh BS, Farach AM. Socioeconomic Factors Associated with Lack of Treatment in Early Stage Non-Small Cell Lung Cancer. *Cancer Invest*. 2019;37(9):506-511. doi:10.1080/07357907.2019.1666136
24. Dalwadi S, Teh BS, Bernicker E, Butler EB, Farach AM. Community-based Disparities in the Treatment and Outcomes of Early-stage Non-small-cell Carcinoma. *Cureus*. 2019;11(10):e5889. doi:10.7759/cureus.5889
25. Haque W, Szeja S, Tann A, Kalra S, Teh BS. Changes in Treatment Patterns and Overall Survival in Patients with Early-Stage Non-Small Cell Lung Cancer in the United States after the Incorporation of Stereotactic Ablative Radiation Therapy. *Am J Clin Oncol Cancer Clin Trials*. 2018;41(3):259-266. doi:10.1097/COC.0000000000000265
26. Fossum CC, Ding L, Atay SM, et al. Stereotactic body radiation therapy for early-stage non-small cell lung cancer in the USA: patterns of adoption and potential healthcare disparities. *J Radiat Oncol*. 2020;9(3-4):225-234. doi:10.1007/s13566-020-00442-y
27. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10(1):37-48. doi:10.1097/00001648-199901000-00008
28. Connected Papers | Find and explore academic papers.
29. Home - ClinicalTrials.gov.
30. Eden J, Levit L, Berg A, Morton S, eds. Standards for Finding and Assessing Individual Studies | Finding What Works in Health Care: Standards for Systematic Reviews |The National Academies Press. In: *Finding What Works in Health Care: Standards for Systematic Reviews*. THE NATIONAL ACADEMIES PRESS ; 2011. Accessed February 3, 2023. https://nap.nationalacademies.org/read/13059/chapter/5
31. Textor J, Van Der Zander B, Gilthorpe MS, Li Skiewicz M, Ellison GT. Software Application Profile Robust causal inference using directed acyclic graphs: the R package "dagitty." doi:10.1093/ije/dyw341
32. Ferguson KD, McCann M, Katikireddi SV, et al. Evidence synthesis for

constructing directed acyclic graphs (ESC-DAGs): a novel and systematic method for building directed acyclic graphs. *Int J Epidemiol*. 2020;49(1):322-329. doi:10.1093/IJE/DYZ150
33. Balekian AA, Wisnivesky JP, Gould MK. Surgical Disparities Among Patients With Stage I Lung Cancer in the National Lung Screening Trial. *Chest*. 2019;155(1):44-52. doi:10.1016/j.chest.2018.07.011
34. Berry MF, Canchola AJ, Gensheimer MF, Gomez SL, Cheng I. Factors Associated With Treatment of Clinical Stage I Non–Small-cell Lung Cancer: A Population-based Analysis. *Clin Lung Cancer*. 2018;19(5):e745-e758. doi:10.1016/j.cllc.2018.05.009
35. Dalwadi SM, Zhang J, Bernicker EH, Butler EB, Teh BS, Farach AM. Socioeconomic Factors Associated with Lack of Treatment in Early Stage Non-Small Cell Lung Cancer. *Cancer Invest*. 2019;37(9):506-511. doi:10.1080/07357907.2019.1666136
36. Dalwadi S, Teh BS, Bernicker E, Butler EB, Farach AM. Community-based Disparities in the Treatment and Outcomes of Early-stage Non-small-cell Carcinoma. *Cureus*. 2019;11(10):e5889. doi:10.7759/cureus.5889
37. Dezube AR, Mazzola E, Cooper L, et al. Geographic differences in therapy for stage I non-small-cell lung cancer in older adults. *J Surg Oncol*. 2022;125(6):1053-1060. doi:10.1002/jso.26809
38. Fossum CC, Ding L, Atay SM, et al. Stereotactic body radiation therapy for early-stage non-small cell lung cancer in the USA: patterns of adoption and potential healthcare disparities. *J Radiat Oncol*. 2020;9(3-4):225-234. doi:10.1007/s13566-020-00442-y
39. Ganesh A, Korpics M, Pasquinelli M, Feldman L, Spiotto M, Koshy M. Increased Utilization of Stereotactic Body Radiotherapy is Associated with Decreased Disparities and Improved Survival for Early-Stage NSCLC. *Clin Lung Cancer*. 2023;24(1):60-71. doi:10.1016/j.cllc.2022.09.006
40. Iezzoni LI, Ngo LH, Li D, Roetzheim RG, Drews RE, McCarthy EP. Treatment Disparities for Disabled Medicare Beneficiaries With Stage I Non-Small Cell Lung Cancer. *Arch Phys Med Rehabil*. 2008;89(4):595-601. doi:10.1016/j.apmr.2007.09.042
41. Chang JY, Mehran RJ, Feng L, et al. Stereotactic ablative radiotherapy for operable stage I non-small-cell lung cancer (revised STARS): long-term results of a single-arm, prospective trial with prespecified comparison to surgery. *Lancet Oncol*. 2021;22(10):1448-1457. doi:10.1016/S1470-2045(21)00401-0
42. Dai C, Shen J, Ren Y, et al. Choice of Surgical Procedure for Patients With Non-Small-Cell Lung Cancer ≤ 1 cm or > 1 to 2 cm Among Lobectomy, Segmentectomy, and Wedge Resection: A Population-Based Study. *J Clin Oncol*. 2016;34(26):3175-3182. doi:10.1200/JCO.2015.64.6729
43. Dalwadi SM, Lewis GD, Bernicker EH, Butler EB, Teh BS, Farach AM. Disparities in the Treatment and Outcome of Stage I Non–Small-Cell Lung Cancer in the 21st Century. *Clin Lung Cancer*. 2019;20(3):194-200. doi:10.1016/j.cllc.2018.11.004
44. Dalwadi SM, Szeja SS, Bernicker EH, Butler EB, Teh BS, Farach AM. Practice Patterns and Outcomes in Elderly Stage I Non–Small-cell Lung Cancer: A 2004 to 2012 SEER Analysis. *Clin Lung Cancer*. 2018;19(2):e269-e276. doi:10.1016/j.cllc.2017.11.004
45. Hao B, Li F, Wan X, et al. Squamous cell carcinoma predicts worse prognosis than adenocarcinoma in stage IA lung cancer patients: A population-based

propensity score matching analysis. *Front Surg*. 2022;9:944032. doi:10.3389/fsurg.2022.944032
46. Huang L, Peng S, Sun C, et al. Impact of marital status on survival in patients with stage 1A NSCLC. *Aging (Albany NY)*. 2022;14(2):770-779. doi:10.18632/aging.203838
47. Li G, Xie S, Hu F, Tan M, Fan L, Wang C. Segmentectomy or Wedge Resection in Stage IA Lung Squamous Cell Carcinoma and Adenocarcinoma? *J Cancer*. 2021;12(6):1708-1714. doi:10.7150/jca.49683
48. Li M, Qin Y, Mei A, Wang C, Fan L. Effectiveness of radiofrequency ablation therapy for patients with unresected Stage IA non-small cell lung cancer. *J Cancer Res Ther*. 2020;16(5):1007. doi:10.4103/jcrt.JCRT_1040_19
49. Li M, Xu X, Qin Y, et al. Radiofrequency ablation vs. stereotactic body radiotherapy for stage IA non-small cell lung cancer in nonsurgical patients. *J Cancer*. 2021;12(10):3057-3066. doi:10.7150/jca.51413
50. Liang L, Li G, Xie S, et al. Choice of Treatment for Stage IA Non-small Cell Lung Cancer Patients Ineligible for Surgery: Ablation or Stereotactic Body Radiotherapy? *J Cancer*. 2020;11(6):1634-1640. doi:10.7150/jca.39465
51. Lin G, Liu H, Li J. Lobectomy versus sub-lobar resection in patients with stage IA right middle lobe non-small cell lung cancer: A propensity score matched analysis. *J Thorac Dis*. 2019;11(6):2523-2534. doi:10.21037/jtd.2019.05.48
52. Ling L, Zheng H, Lin H, et al. Adequate Number of Lymph Nodes Sampled May Determine Appropriate Surgical Modality for Early-Stage NSCLC: A Population-Based Real-World Study. *Clin Lung Cancer*. Published online December 2022. doi:10.1016/j.cllc.2022.12.011
53. Ni L, Lin G, Zhang Z, Sun D, Liu Z, Liu X. Surgery versus radiotherapy in octogenarians with stage Ia non-small cell lung cancer: propensity score matching analysis of the SEER database. *BMC Pulm Med*. 2022;22(1):411. doi:10.1186/s12890-022-02177-7
54. Razi SS, John MM, Sainathan S, Stavropoulos C. Sublobar resection is equivalent to lobectomy for T1a non–small cell lung cancer in the elderly: a Surveillance, Epidemiology, and End Results database analysis. *J Surg Res*. 2016;200(2):683-689. doi:10.1016/j.jss.2015.08.045
55. Wang L, Ge L, You S, Liu Y, Ren Y. Lobectomy versus segmentectomy in patients with stage T (> 2 cm and ≤ 3 cm) N0M0 non-small cell lung cancer: a propensity score matching study. *J Cardiothorac Surg*. 2022;17(1):110. doi:10.1186/s13019-022-01867-x
56. Wang W, Sun Y, Li H, et al. Surgical modality for stage IA non-small cell lung cancer among the elderly: Analysis of the surveillance, epidemiology, and end results database. *J Thorac Dis*. 2020;12(11):6731-6742. doi:10.21037/jtd-20-2221
57. Wu J, Bai HX, Chan L, et al. Sublobar resection compared with stereotactic body radiation therapy and ablation for early stage non–small cell lung cancer: A National Cancer Database study. *J Thorac Cardiovasc Surg*. 2020;160(5):1350-1357.e11. doi:10.1016/j.jtcvs.2019.11.132
58. Wu L, Zhao W, Chen T, Yang Y. Surgical choice for patients with stage I non-small-cell lung cancer ≤2 cm: an analysis from surveillance, epidemiology, and end results database. *J Cardiothorac Surg*. 2021;16(1):191. doi:10.1186/s13019-021-01568-x
59. Yendamuri S, Dhillon SS, Groman A, et al. Effect of the number of lymph nodes examined on the survival of patients with stage I non–small cell lung

cancer who undergo sublobar resection. *J Thorac Cardiovasc Surg*. 2018;156(1):394-402. doi:10.1016/j.jtcvs.2018.03.113
60. Yu X, Zhang R, Zhang M, et al. Segmental resection is associated with decreased survival in patients with stage IA non-small cell lung cancer with a tumor size of 21-30 mm. *Transl Lung Cancer Res*. 2021;10(2):900-913. doi:10.21037/tlcr-20-1217
61. Zeng C, Lu J, Tian Y, Fu X. Thermal Ablation Versus Wedge Resection for Stage I Non-small Cell Lung Cancer Based on the Eighth Edition of the TNM Classification: A Population Study of the US SEER Database. *Front Oncol*. 2020;10:571684. doi:10.3389/fonc.2020.571684
62. Chang MY, Mentzer SJ, Colson YL, et al. Factors predicting poor survival after resection of stage IA non–small cell lung cancer. *J Thorac Cardiovasc Surg*. 2007;134(4):850-856. doi:10.1016/j.jtcvs.2007.03.044
63. Kates M, Swanson S, Wisnivesky JP. Survival Following Lobectomy and Limited Resection for the Treatment of Stage I Non-small Cell Lung Cancer ≤ 1 cm in Size. *Chest*. 2011;139(3):491-496. doi:10.1378/chest.09-2547
64. Ludwig MS, Goodman M, Miller DL, Johnstone PAS. Postoperative Survival and the Number of Lymph Nodes Sampled During Resection of Node-Negative Non-Small Cell Lung Cancer. *Chest*. 2005;128(3):1545-1550. doi:10.1378/chest.128.3.1545
65. Whitson BA, Groth SS, Andrade RS, Habermann EB, Maddaus MA, D'Cunha J. T1/T2 non–small-cell lung cancer treated by lobectomy: Does tumor anatomic location matter? *J Surg Res*. 2012;177(2):185-190. doi:10.1016/j.jss.2012.05.022
66. Lung Cancer STARS Trial - STARS Revised Clinical Trial Protocol: Stereotactic Ablative Radiotherapy (SABR) in Stage I Non-small Cell Lung Cancer Patients Who Can Undergo Lobectomy - Full Text View - ClinicalTrials.gov. Accessed January 22, 2023. https://clinicaltrials.gov/ct2/show/NCT02357992?term=NCT02357992&draw=2&rank=1
67. Radiofrequency Ablation in Treating Patients With Stage I Non-Small Cell Lung Cancer - Full Text View - ClinicalTrials.gov. Accessed January 22, 2023. https://clinicaltrials.gov/ct2/show/study/NCT00109876?term=NCT00109876&draw=2&rank=1
68. Ettinger D, Wood D, Aisner D, et al. *NCCN Clinical Practice Guidelines in Oncology - Non-Small Cell Lung*.; 2023.
69. Sineshaw HM, Sahar L, Osarogiagbon RU, Dana Flanders W, Robin Yabroff K, Jemal A. County-Level Variations in Receipt of Surgery for Early-Stage Non-small Cell Lung Cancer in the United States. doi:10.1016/j.chest.2019.09.016
70. Balekian AA, Wisnivesky JP, Gould MK. Surgical Disparities Among Patients With Stage I Lung Cancer in the National Lung Screening Trial. *Chest*. 2019;155(1):44-52. doi:10.1016/j.chest.2018.07.011
71. Dalwadi SM, Zhang J, Bernicker EH, Butler EB, Teh BS, Farach AM. Socioeconomic Factors Associated with Lack of Treatment in Early Stage Non-Small Cell Lung Cancer. *Cancer Invest*. 2019;37(9):506-511. doi:10.1080/07357907.2019.1666136
72. Dalwadi S, Teh BS, Bernicker E, Butler EB, Farach AM. Community-based Disparities in the Treatment and Outcomes of Early-stage Non-small-cell Carcinoma. *Cureus*. Published online October 11, 2019.

doi:10.7759/cureus.5889
73. Fossum CC, Ding L, Atay SM, et al. Stereotactic body radiation therapy for early-stage non-small cell lung cancer in the USA: patterns of adoption and potential healthcare disparities. *J Radiat Oncol*. 2020;9(3-4):225-234. doi:10.1007/s13566-020-00442-y
74. Overview of the SEER Program. Accessed March 3, 2023. https://seer.cancer.gov/about/overview.html
75. Area Health Resources Files. Accessed March 3, 2023. https://data.hrsa.gov/topics/health-workforce/ahrf
76. *Social Determinants of Health (SDOH) Beta Data Files Data Source Documentation*.; 2020. Accessed March 3, 2023. https://www.ahrq.gov/sites/default/files/wysiwyg/sdohchallenge/data/sdoh_data_file_documentation.pdf
77. Stata Statistical Software: Release 16. Published online 2019.
78. Bei Y, Chen X, Raturi VP, et al. Treatment patterns and outcomes change in early-stage non-small cell lung cancer in octogenarians and older: a SEER database analysis. *Aging Clin Exp Res*. 2021;33(1):147-156. doi:10.1007/s40520-020-01517-z
79. Kreinbrink P, Blumenfeld P, Tolekidis G, Sen N, Sher D, Marwaha G. Lung stereotactic body radiation therapy (SBRT) for early-stage non-small cell lung cancer in the very elderly (≥ 80 years old): Extremely safe and effective. *J Geriatr Oncol*. 2017;8(5):351-355. doi:10.1016/j.jgo.2017.07.002
80. NIH National Cancer Institute. Non-Small Cell Lung Cancer Treatment (PDQ®): Health Professional Version. NIH National Cancer Institute . Published 2020. Accessed May 10, 2023. https://www.cancer.gov/types/lung/hp/non-small-cell-lung-treatment-pdq
81. Kann BH, Johnson SB, Aerts HJWL, Mak RH, Nguyen PL. Changes in Length and Complexity of Clinical Practice Guidelines in Oncology, 1996-2019. *JAMA Netw Open*. 2020;3(3):2019-2021. doi:10.1001/jamanetworkopen.2020.0841
82. American Cancer Society. Non-small Cell Lung Cancer Staging | Stages of Lung Cancer. American Cancer Society. Accessed April 28, 2023. https://www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/staging-nsclc.html
83. Urban and Rural. United States Census Bureau. Accessed April 29, 2023. https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural.html
84. Wink KCJ, van Baardwijk A, Troost EGC, De Ruysscher D. Nodal recurrence after stereotactic body radiotherapy for early stage non-small cell lung cancer: Incidence and proposed risk factors. *Cancer Treat Rev*. 2017;56:8-15. doi:10.1016/j.ctrv.2017.04.001
85. Nakagawa M, Tanaka H, Tsukuma H, Kishi Y. Relationship Between the Duration of the Preoperative Smoke-Free Period and the Incidence of Postoperative Pulmonary Complications After Pulmonary Surgery. *Chest*. 2001;120(3):705-710. doi:10.1378/chest.120.3.705
86. Lugg ST, Tikka T, Agostini PJ, et al. Smoking and timing of cessation on postoperative pulmonary complications after curative-intent lung cancer surgery. *J Cardiothorac Surg*. 2017;12(1):1-8. doi:10.1186/s13019-017-0614-4
87. Samson P, Patel A, Garrett T, et al. Effects of Delayed Surgical Resection on Short-Term and Long-Term Outcomes in Clinical Stage I Non-Small Cell Lung

Cancer. *Ann Thorac Surg*. 2015;99(6):1906-1913. doi:10.1016/j.athoracsur.2015.02.022
88. Pandit AA, Gressler LE, Halpern MT, Kamel M, Payakachat N, Li C. Racial/Ethnic Disparities in Patient Care Experiences among Prostate Cancer Survivors: A SEER-CAHPS Study. *Curr Oncol*. 2022;29(11):8357-8373. doi:10.3390/curroncol29110659
89. Shugarman LR, MacK K, Sorbero MES, et al. Race and sex differences in the receipt of timely and appropriate lung cancer treatment. *Med Care*. 2009;47(7):774-781. doi:10.1097/MLR.0b013e3181a393fe
90. Nadpara P, Madhavan SS, Tworek C. Guideline-concordant timely lung cancer care and prognosis among elderly patients in the United States: A population-based study. *Cancer Epidemiol*. 2015;39(6):1136-1144. doi:10.1016/j.canep.2015.06.005
91. Aksoy M. Different types of malignancies due to occupational exposure to benzene: a review of recent observations in Turkey. *Environ Res*. 1980;23 1:181-190. doi:10.1016/0013-9351(80)90104-8
92. McKeon TP, Anil Vachani, Penning TM, Hwang WT. Air pollution and lung cancer survival in Pennsylvania. *Lung Cancer*. 2022;170(June):65-73. doi:10.1016/j.lungcan.2022.06.004
93. Pyo JS, Kim NY, Kang DW. Impacts of Outdoor Particulate Matter Exposure on the Incidence of Lung Cancer and Mortality. *Med*. 2022;58(9):1-9. doi:10.3390/medicina58091159
94. Lamichhane DK, Kim H, Choi C, et al. Lung Cancer Risk and Residential Exposure to Air Pollution : A Korean Population-Based Case-Control Study. 2017;58(6):1111-1118.
95. Jerrett M, Burnett RT, Ma R, et al. Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiology*. 2005;16(6):727-736. doi:10.1097/01.ede.0000181630.15826.7d
96. Ngarambe J, Joen SJ, Han CH, Yun GY. Exploring the relationship between particulate matter, CO, SO2, NO2, O3 and urban heat island in Seoul, Korea. *J Hazard Mater*. 2021;403(2):123615. doi:10.1016/j.jhazmat.2020.123615
97. Wei Y, Qiu X, Yazdi MD, et al. The Impact of Exposure Measurement Error on the Estimated Concentration-Response Relationship between Long-Term Exposure to PM2.5 and Mortality. *Environ Health Perspect*. 2022;130(7):77006. doi:10.1289/EHP10389
98. Kates M, Swanson S, Wisnivesky JP. Survival Following Lobectomy and Limited Resection for the Treatment of Stage I Non-small Cell Lung Cancer ≤ 1 cm in Size: A Review of SEER Data. *Chest*. 2011;139(3):491-496. doi:10.1378/chest.09-2547
99. Mery CM, Pappas AN, Bueno R, et al. Similar long-term survival of elderly patients with non-small cell lung cancer treated with lobectomy or wedge resection within the surveillance, epidemiology, and end results database. *Chest*. 2005;128(1):237-245. doi:10.1378/chest.128.1.237
100. Altorki NK, Wang X, Wigle D, et al. Perioperative mortality and Morbidity after Lobar versus Sublobar Resection for early stage lung cancer: A post-hoc analysis of an international randomized phase III trial (CALGB/ Alliance 140503). 2019;6(12):915-924. doi:10.1016/S2213-2600(18)30411-9
101. Rueth NM, Parsons HM, Habermann EB, et al. Surgical treatment of lung cancer: predicting postoperative morbidity in the elderly population. *J Thorac Cardiovasc Surg*. 2012;143(6):1314-1323. doi:10.1016/j.jtcvs.2011.09.072

APPENDICES

Search Strategy for Aim 1 across each database

PubMed: n=753

("Carcinoma, Non-Small-Cell Lung"[Mesh] OR "non-small-cell lung carcinoma" OR "non-small-cell lung carcinomas" OR "non small cell lung carcinoma" OR "non small cell lung carcinomas" OR "non-small cell lung carcinoma" OR "non-small cell lung carcinomas" OR "non-small cell lung cancer" OR "non small cell lung cancer" OR "non-small-cell lung cancer" OR "Adenocarcinoma of Lung"[Mesh] OR "squamous cell carcinoma of the lung") AND ("Carcinoma, Non-Small-Cell Lung/surgery"[Mesh] OR "Surgical Procedures, Operative"[Mesh] OR surgery OR "operative procedure" OR "operative procedures" OR "surgical procedure" OR "surgical procedures" OR resection* OR "surgical treatment" OR "Carcinoma, Non-Small-Cell Lung/radiotherapy"[Mesh] OR radiother* OR "radiation therapy" OR "radiation therapies" OR "radiation treatment" OR "radiation treatments" OR irradiation OR Survival[Mesh] OR Mortality[Mesh] OR "Survival Rate"[Mesh] OR outcome* OR mortality OR surviv*) AND ("SEER Program"[Mesh] OR "SEER program" OR SEER OR "Surveillance, Epidemiology, and End Results Program" OR "Surveillance, Epidemiology and End Results Program" OR "National Cancer Registry" OR " US National Cancer Database") AND (2002:2023[pdat])

Embase: n=1762

('non small cell lung cancer'/exp OR 'bronchial non small cell cancer' OR 'bronchial non small cell carcinoma' OR 'carcinoma, non-small-cell lung' OR 'lung cancer, non small cell' OR 'lung non small cell cancer' OR 'lung non small cell carcinoma' OR 'non small cell bronchial cancer' OR 'non small cell cancer, lung' OR

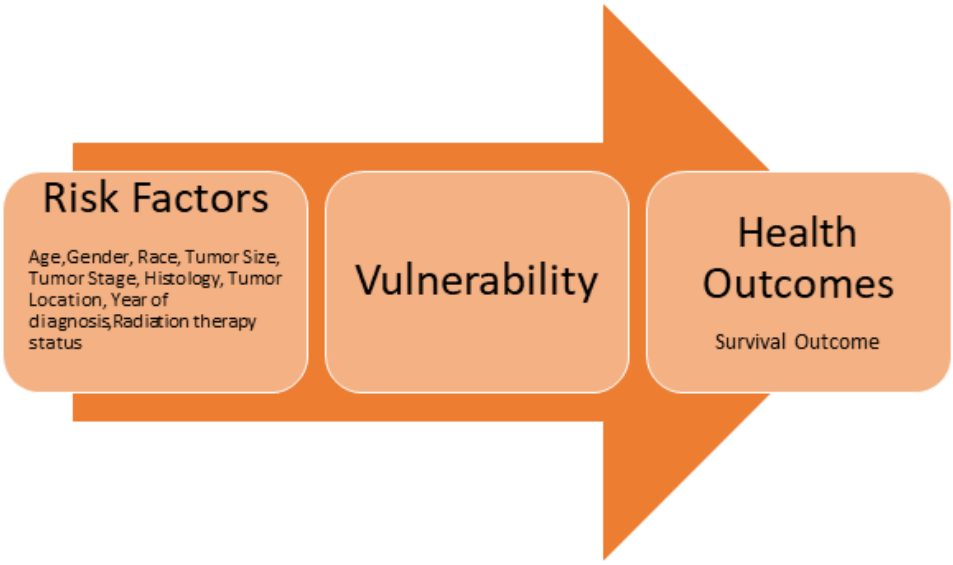
'non small cell lung cancer' OR 'non small cell lung carcinoma' OR 'non small cell pulmonary cancer' OR 'non small cell pulmonary carcinoma' OR 'non-small-cell lung carcinoma' OR 'pulmonary non small cell cancer' OR 'pulmonary non small cell carcinoma' OR 'adenocarcinoma of lung' OR 'squamous cell carcinoma of the lung') AND ('cancer registry'/exp OR 'cdc-npcr' OR 'centers for disease control and prevention national program of cancer registries' OR 'npcr' OR 'national program of cancer registries' OR 'seer program' OR 'seer programme' OR 'united states national program of cancer registries' OR 'cancer register' OR 'cancer registration' OR 'cancer registry') AND ('surgery'/exp OR 'diagnosis, surgical' OR 'diagnostic techniques, surgical' OR 'operation' OR 'operation care' OR 'operative intervention' OR 'operative repair' OR 'operative restoration' OR 'operative surgery' OR 'operative surgical procedure' OR 'operative surgical procedures' OR 'operative treatment' OR 'research surgery' OR 'resection' OR 'specialties, surgical' OR 'surgery' OR 'surgery, operative' OR 'surgical care' OR 'surgical correction' OR 'surgical diagnosis' OR 'surgical diagnostic techniques' OR 'surgical exposure' OR 'surgical intervention' OR 'surgical management' OR 'surgical operation' OR 'surgical practice' OR 'surgical procedures, operative' OR 'surgical repair' OR 'surgical research' OR 'surgical restoration' OR 'surgical service' OR 'surgical speciality' OR 'surgical specialties' OR 'surgical specialty' OR 'surgical therapy' OR 'surgical treatment' OR 'radiotherapy'/exp OR 'bioradiant therapy' OR 'bucky irradiation' OR 'bucky radiation' OR 'bucky radiotherapy' OR 'bucky ray' OR 'bucky ray radiation' OR 'bucky therapy' OR 'fractionated radiotherapy' OR 'hemibody irradiation' OR 'hypophysectomy, radiation' OR 'hypophysis irradiation' OR 'hypophysis radiation' OR 'irradiation therapy' OR 'irradiation treatment' OR 'irradiation, hypophysis' OR 'lymphatic irradiation' OR 'pituitary irradiation' OR 'radiation beam centration' OR 'radiation repair' OR 'radiation therapy' OR 'radiation

treatment' OR 'radio therapy' OR 'radio treatment' OR 'radiohypophysectomy' OR 'radiology, therapeutic' OR 'radiotherapy' OR 'radiotherapy setup errors' OR 'radiotreatment' OR 'roentgen irradiation, therapeutic' OR 'roentgen therapy' OR 'roentgen treatment' OR 'rontgen therapy' OR 'therapeutic radiology' OR 'therapy, irradiation' OR 'therapy, radiation' OR 'therapy, roentgen' OR 'treatment, irradiation' OR 'treatment, radiation' OR 'treatment, roentgen' OR 'x radiotherapy' OR 'x ray therapy' OR 'x ray treatment' OR 'x-ray therapy' OR 'survival'/exp OR 'survival' OR 'mortality'/exp OR 'excess mortality' OR 'mortality' OR 'mortality model' OR 'treatment outcome'/exp OR 'medical futility' OR 'outcome and process assessment (health care)' OR 'outcome and process assessment, health care' OR 'outcome management' OR 'patient outcome' OR 'therapeutic outcome' OR 'therapy outcome' OR 'treatment outcome') AND (2002:py OR 2003:py OR 2004:py OR 2005:py OR 2006:py OR 2007:py OR 2008:py OR 2009:py OR 2010:py OR 2011:py OR 2012:py OR 2013:py OR 2014:py OR 2015:py OR 2016:py OR 2017:py OR 2018:py OR 2019:py OR 2020:py OR 2021:py OR 2022:py OR 2023:py)

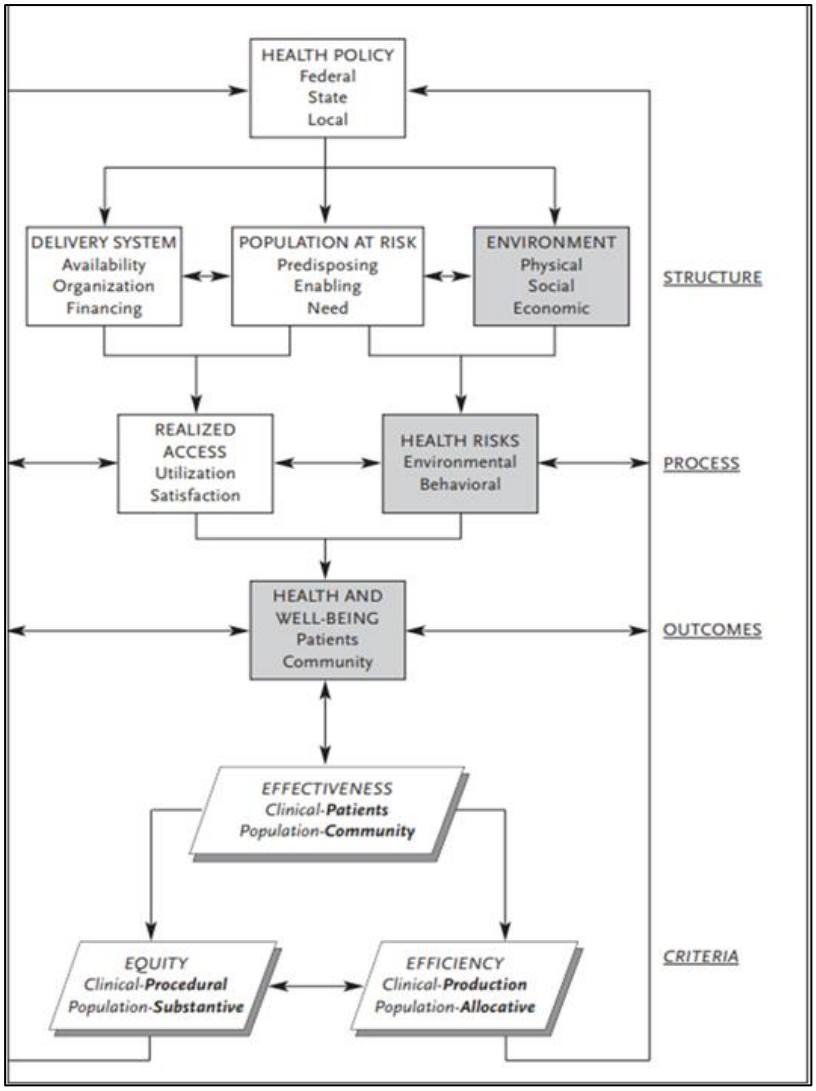
Web of Science: n=6,906

(((((((((((((((TI=(“non-small-cell lung carcinoma” OR “non-small-cell lung carcinomas” OR “non small cell lung carcinoma” OR “non small cell lung carcinomas” OR “non-small cell lung carcinoma” OR “non-small cell lung carcinomas” OR “non-small cell lung cancer” OR “non small cell lung cancer” OR “non-small-cell lung cancer” OR "adenocarcinoma of lung" OR "squamous cell carcinoma of the lung")) OR AB=(“non-small-cell lung carcinoma” OR “non-small-cell lung carcinomas” OR “non small cell lung carcinoma” OR “non small cell lung carcinomas” OR “non-small cell lung carcinoma” OR “non-small cell lung carcinomas” OR “non-small cell lung cancer” OR “non small cell lung cancer” OR “non-small-cell lung cancer” "adenocarcinoma of

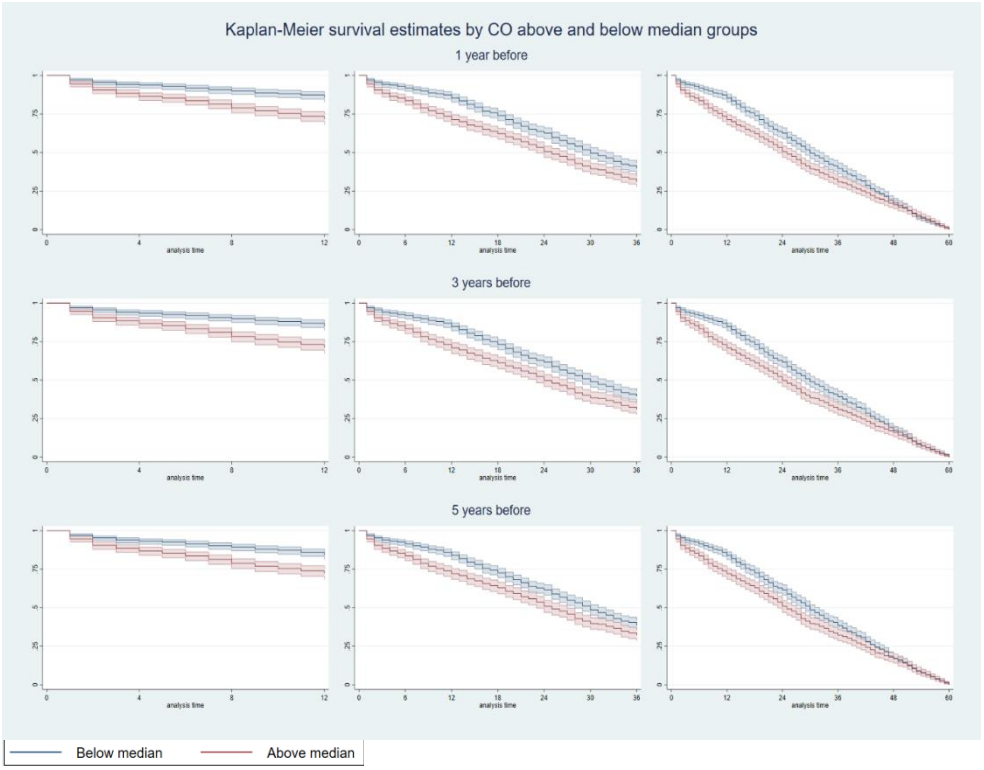
lung" OR "squamous cell carcinoma of the lung"))) AND TI=(surgery OR “operative procedure” OR “operative procedures” OR “surgical procedure” OR “surgical procedures” OR resection* OR “surgical treatment” OR radiother* OR “radiation therapy” OR “radiation therapies” OR “radiation treatment” OR “radiation treatments” OR irradiation)) OR AB=(surgery OR “operative procedure” OR “operative procedures” OR “surgical procedure” OR “surgical procedures” OR resection* OR “surgical treatment” OR radiother* OR “radiation therapy” OR “radiation therapies” OR “radiation treatment” OR “radiation treatments” OR irradiation)) OR TI=(outcome* OR mortality OR surviv*)) OR AB=(outcome* OR mortality OR surviv*)) AND TI=(“SEER program” OR SEER OR “Surveillance, Epidemiology, and End Results Program” OR “Surveillance, Epidemiology and End Results Program” OR "national cancer registry")) OR AB=(“SEER program” OR SEER OR “Surveillance, Epidemiology, and End Results Program” OR “Surveillance, Epidemiology and End Results Program” OR "national cancer registry")))))))) AND (PY==("2002" OR "2003" OR "2004" OR "2005" OR "2006" OR "2007" OR "2008" OR "2009" OR "2010" OR "2011" OR "2012" OR "2013" OR "2014" OR "2015" OR "2016" OR "2017" OR "2018" OR "2019" OR "2020" OR "2021" OR "2022" OR "2023") AND SILOID==("WOS") AND CU==("USA") AND LA==("ENGLISH") AND DT==("ARTICLE"))))



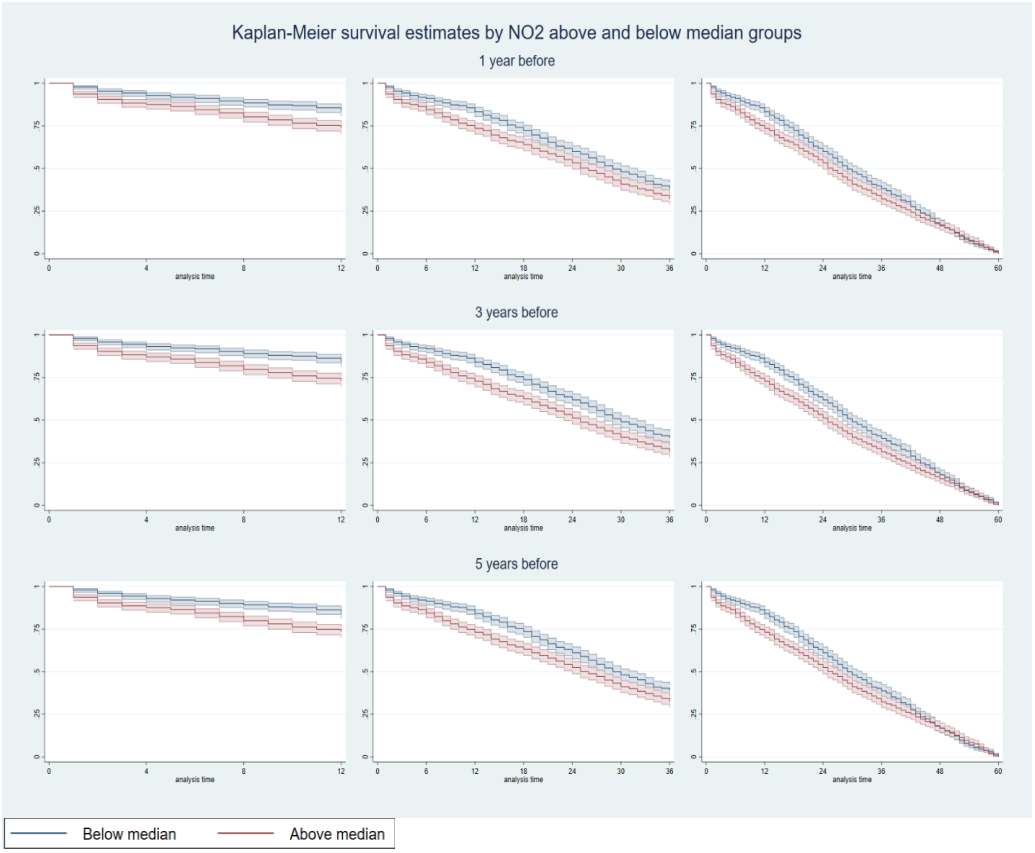
Appendix Figure 2.1 Shi and Steven conceptual model for vulnerable population



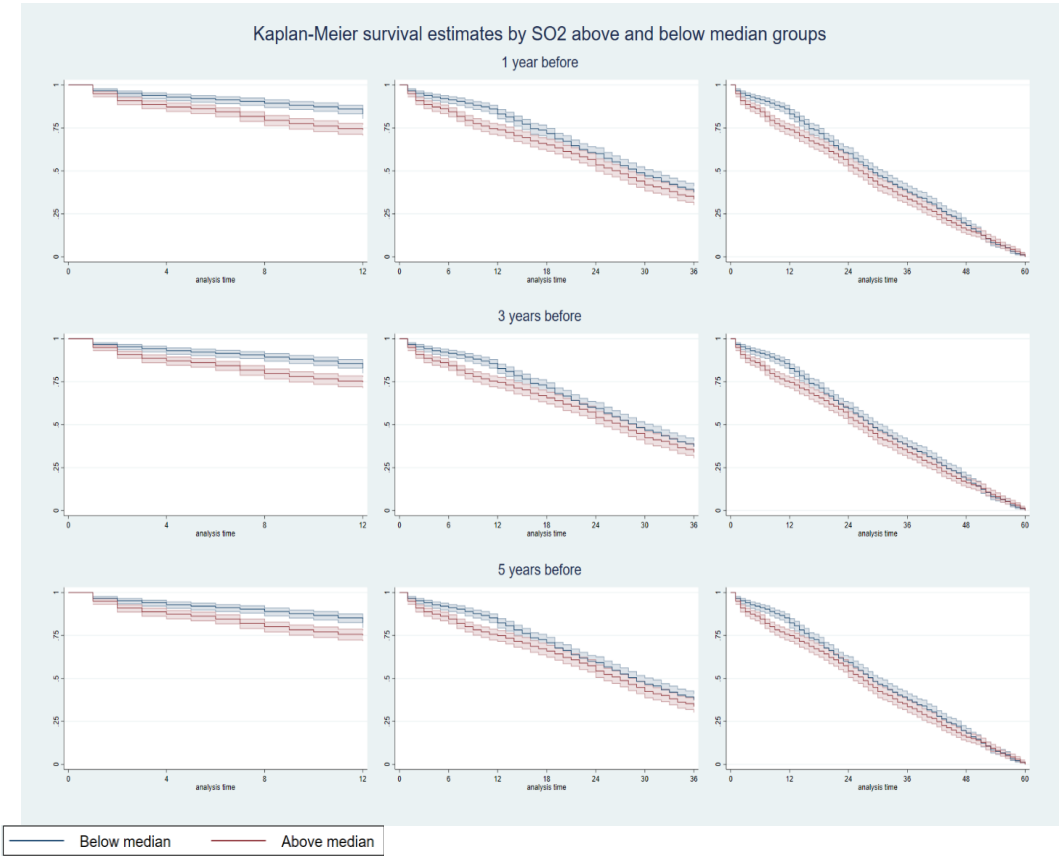
Appendix Figure 2.2 Andersen and Aday health services research model



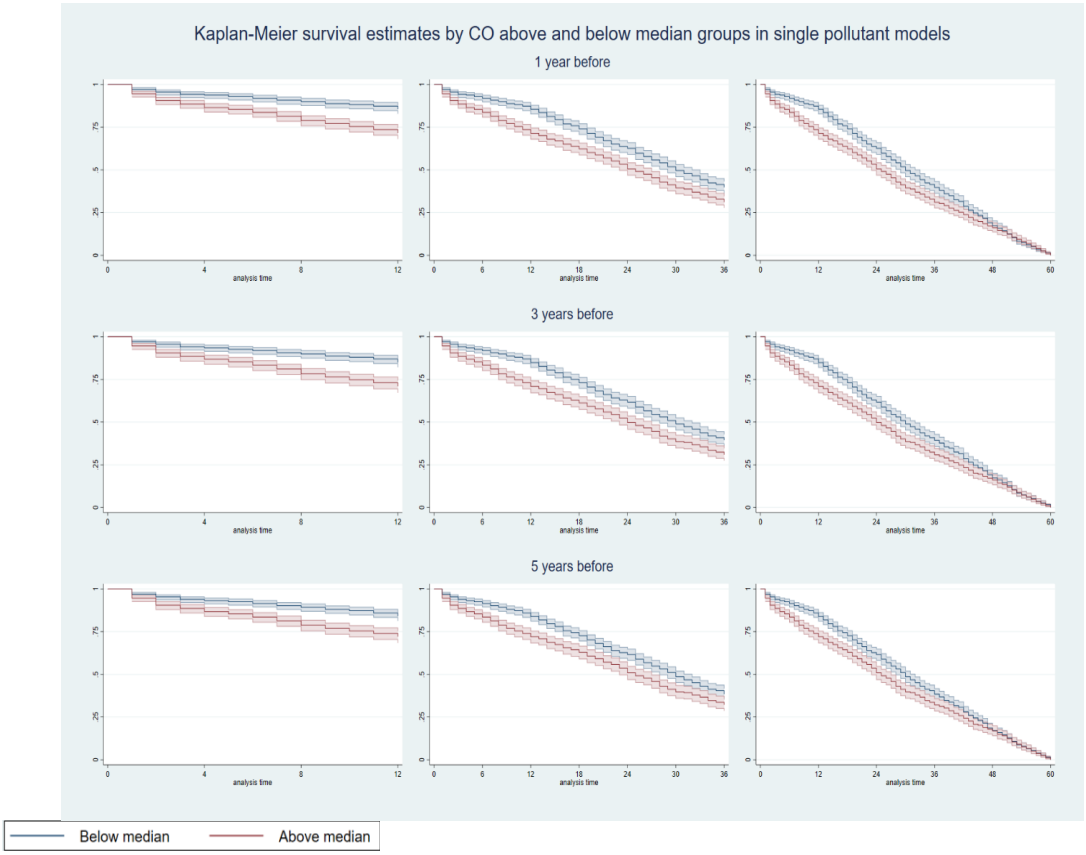
Appendix Figure 4.1 Multi-pollutant Model: Kaplan-Meier survival estimates with 95% confidence interval by CO above and below median groups (N=1500)



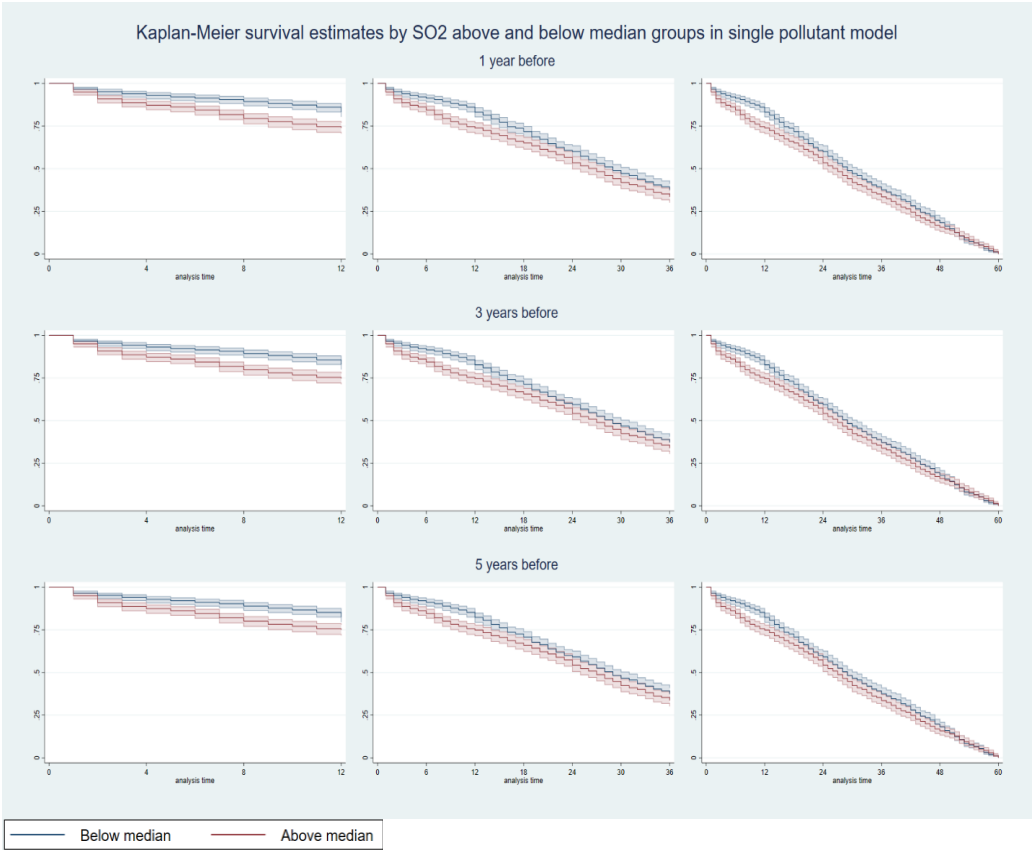
Appendix Figure 4.2 Multi-pollutant Model: Kaplan-Meier survival estimates with 95% confidence interval by NO_2 above and below median groups (N=1500)



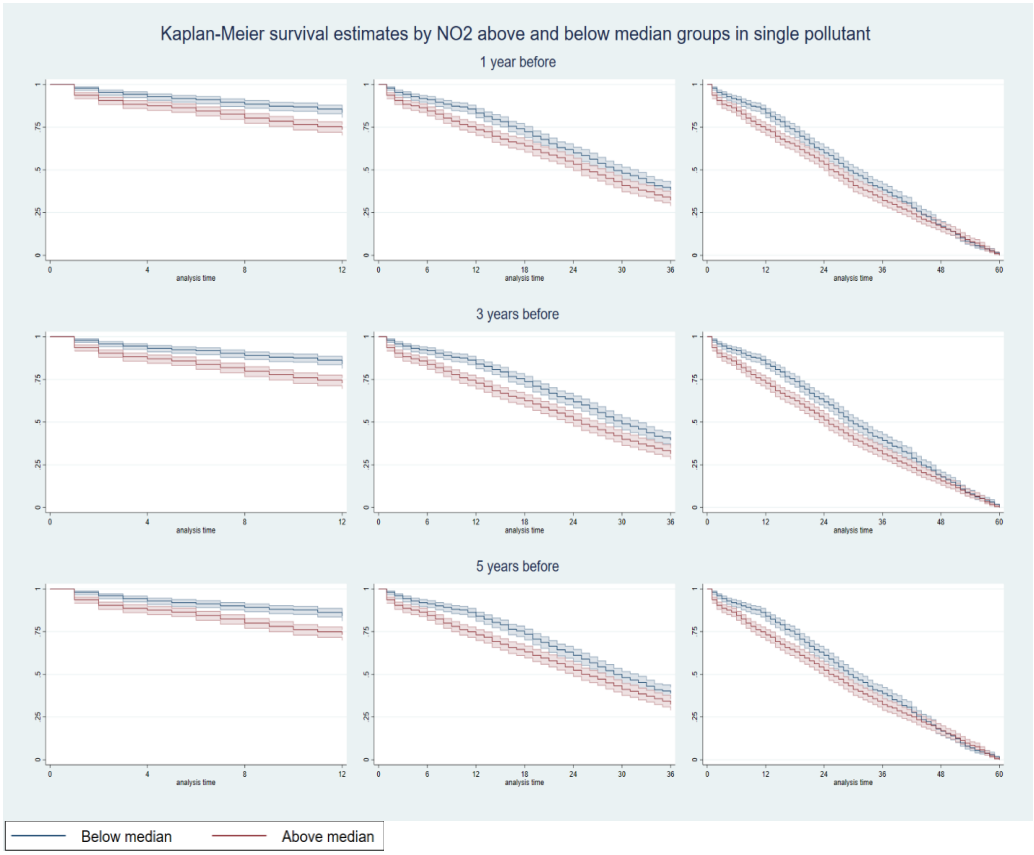
Appendix Figure 4.3 Multi-pollutant Model: Kaplan-Meier survival estimates with 95% confidence interval by SO_2 above and below median groups (N=1500)



Appendix Figure 4.4 Single-pollutant Model: Kaplan-Meier survival estimates by CO pollutant above and below median groups (N=1500)



Appendix Figure 4.5 Single-pollutant Model: Kaplan-Meier survival estimates with 95% confidence interval by SO_2 above and below median groups (N=1500)

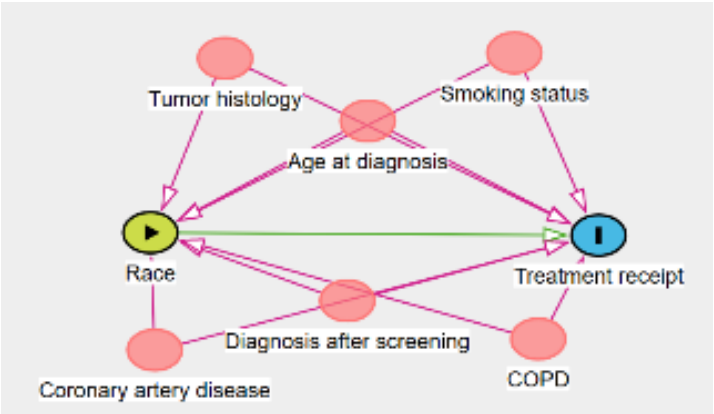
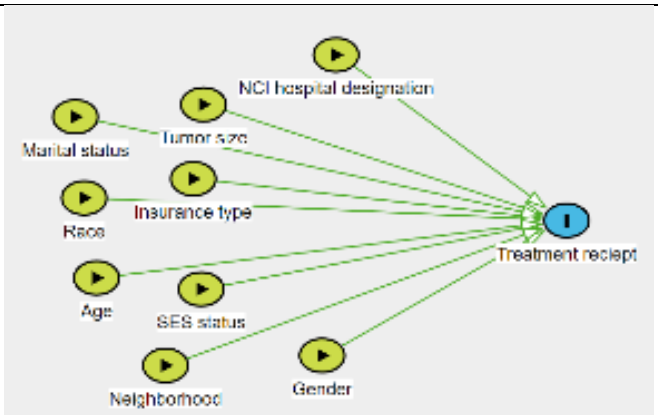


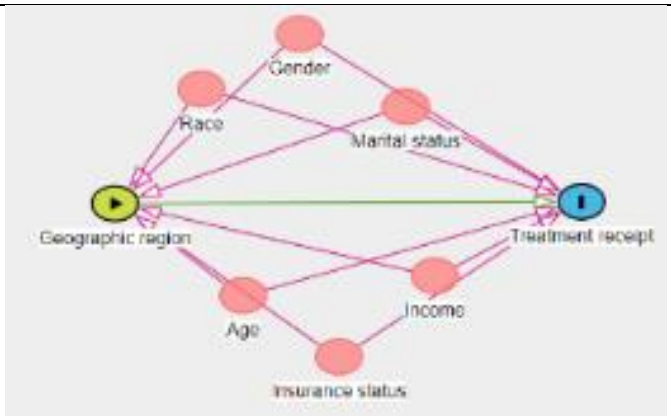
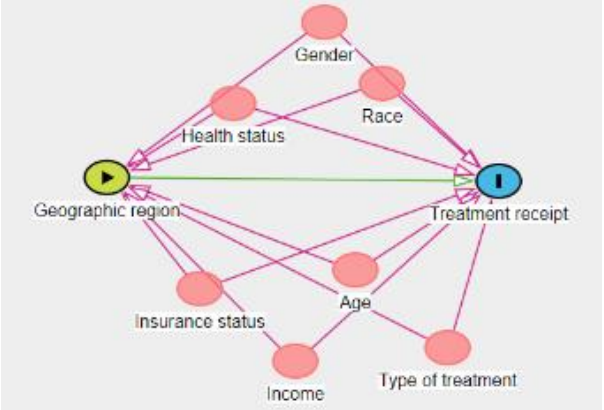
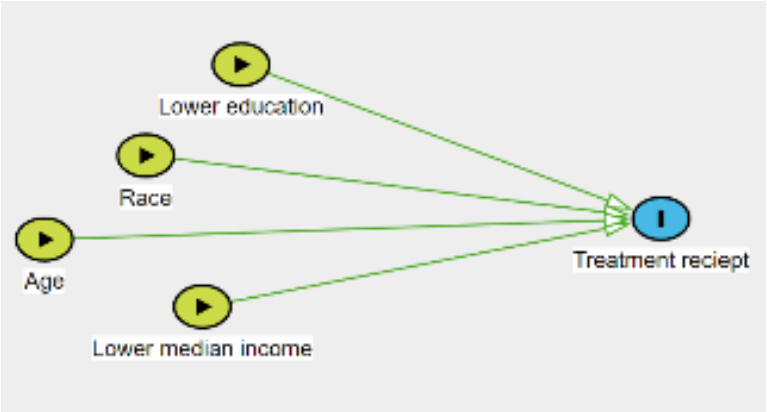
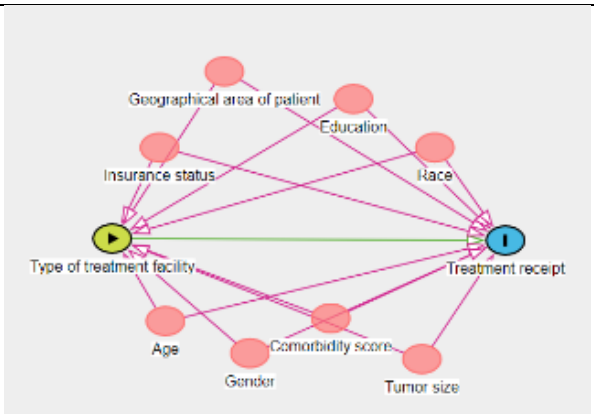
Appendix Figure 4.6 Single-pollutant Model: Kaplan-Meier survival estimates with 95% confidence interval by NO_2 above and below median groups (N=1500)

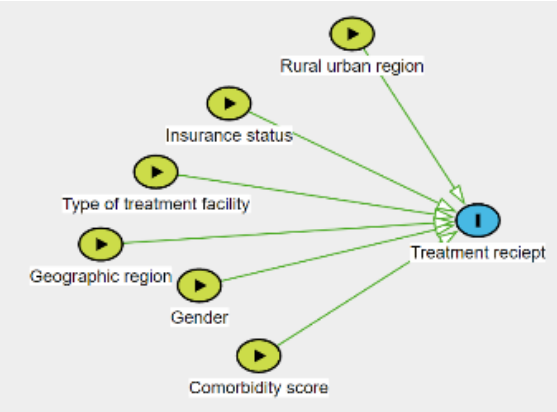
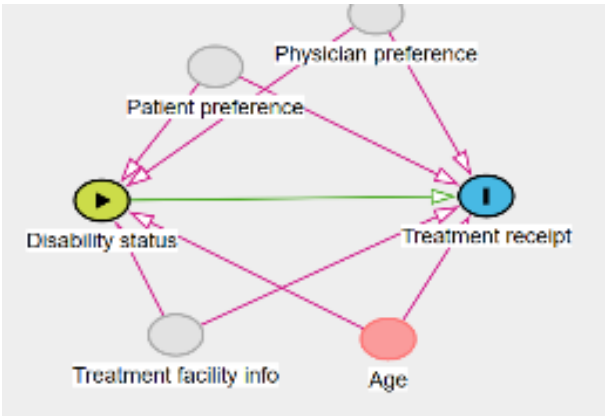
Studies	Study period	Data Registry	Age (Years)	Sample size	Intervention	Outcome	AJCC staging version	Factor component (Treatment/Survival)
Balekian et al.[33]	2002-2004	National Lung Cancer Screening Trial (NLST)	55-74	723	Race	Treatment receipt	6th	Treatment receipt
Berry et al. [34]	2003-2014	California Cancer Registry	>=18	19,893	Factors associated with therapy receipt	Treatment receipts	Not mentioned	Treatment receipt
Chang et al.[41]	2015-2020	STARS trial University of Texas	>=18	80	VATS vs L-MLND	Survival	7th	Survival
Dai et al.[42]	2000-2012	SEER 18	<= & > 65	15,760	Lobectomy vs Sub lobectomy	Survival	Not mentioned	Survival
Dalwadi et al.[43]	2002-2012	SEER 18	>=60	62,213	Surgery/Radiation/Neither	Survival	6th	Survival
Dalwadi et al.[44]	2002-2012	SEER 18	>=60	62,213	Surgery/Radiation/Neither	Survival	6th	Survival
Dalwadi et al.[35]	2002-2012	SEER 18	>=60	62,213	Rural/Urban/Metropolitan	Treatment receipts	7th	Treatment receipt
Dalwadi et al.[36]	2004-2012	SEER 18	>60	62,213	Rural/Urban/Metropolitan	Treatment receipts	6th	Treatment receipt
Dezube et al.[37]	2004-2012	SEER 18	>=60	43,387	Factors associated with therapy receipt	Treatment receipt	8th	Treatment receipt
Fossum et al.[38]	2004-2016	NCDB	>18	65,376	Academic/Community/Comprehensive centre * Year of diagnosis	Treatment receipt	6th or 7th	Treatment receipt
Ganesh et al.[39]	2004-2017	NCDB	Not mentioned	337,594	Factors associated with treatment receipt	Treatment receipt	8th	Treatment receipt
Hao et al.[45]	2004-2013	SEER	<=69 & >69	27,398	Adenocarcinoma/Squamous cell carcinoma histology	Survival	Not mentioned	Survival
Haque et al.[25]	2004-2012	SEER 18	<=50 - >=75	32,249	Surgery/Radiation/Neither	Survival	6th	Survival
Huang et al.[46]	1995-2015	SEER	<=60->=75	55,207	Marital Status	Survival	Not mentioned	Survival
Li et al.[47]	2004-2015	SEER	<=45=>=75	5,599	Wedge resection/Segmentectomy	Survival	Not mentioned	Survival
Li et al.[48]	2004-2015	SEER	<=55->=75	5,268	Radiofrequency ablation/ No treatment	Survival	Not mentioned	Survival
Li et al.[49]	2004-2015	SEER 18	<=44->=75	6,195	Radiofrequency ablation/ Stereotactic body radiotherapy	Survival	Not mentioned	Survival
Liang et al.[50]	2004-2015	SEER	<=44->=75	6,395	Ablation/ Stereotactic body radiotherapy	Survival	Not mentioned	Survival
Lin et al.[51]	2005-2015	SEER	<=67 &>67	1,104	Lobectomy/Sublobectomy	Survival	6th	Survival
Ling et al.[52]	1998-2017	SEER 18	20-80	6,150	Lobectomy/Sublobectomy	Survival	Not mentioned	Survival
Ni et al.[53]	2012-2017	SEER 18	>=80	1,641	Surgery/Radiotherapy	Survival	8th	Survival
Razi et al.[54]	1998-2007	SEER	>=75	1,640	Lobectomy/Sublobectomy	Survival	7th	Survival
Wang et al.[55]	2004-2015	SEER	<=60 & >=80	5,783	Lobectomy/Sublobectomy	Survival	8th	Survival
Wang et al.[56]	1998-2016	SEER	>=70	6,197	Lobectomy/Sublobectomy	Survival	8th	Survival
Wu et al.[57]	2004-2014	NCDB	Not mentioned	53,973	Sublobar resection/ Ablation/ Stereotactic body radiotherapy	Survival	8th	Survival
Wu et al.[58]	2004-2015	SEER 18	<60- & >=75	16,511	Lobectomy/Sublobectomy	Survival	Not mentioned	Survival
Yendamuri et al.[59]	2004-2013	SEER	Not mentioned	3,916	Wedge/ Segmentectomy	Survival	Not mentioned	Survival
Yu et al.[60]	1998-2016	SEER 18	>=18	9,580	Lobectomy/Sublobectomy	Survival	Not mentioned	Survival
Zeng et al.[61]	2004-2014	SEER	<&>=75	4,372	Thermal ablation/Wedge resection	Survival	8th	Survival
Chang et al.[62]	1988-1997	SEER	<& >=67	10,761	Lobectomy/Sublobectomy	Survival	Not mentioned	Survival
Iezzoni et al.[40]	1988-1999	SEER11-Medicare	21-64	9,500	Disability status	Treatment receipt	Not mentioned	Treatment receipt
Kates et al.[63]	1988-2005	SEER	<&>=60	2,090	Limited resection/Lobectomy	Survival	Not mentioned	Survival
Ludwig et al.[64]	1990-2000	SEER	<&>=45	16,800	Number of lymph nodes sampled during surgery	Survival	Not mentioned	Survival

Whitson et al.[65]	1988-2007	SEER	>=40	13,650	Treatment type	Survival	Not mentioned	Survival
STAR trial[66]	2010-2021	Clinical trial study	>=18	122	Surgery/ Stereotactic Body Radiation Therapy (SBRT)	Survival	Not mentioned	Survival
Clinical trial NCT00109876[67]	2006-2013	Clinical trial study	>=18	51	Radiofrequency Ablation	Survival	Not mentioned	Survival

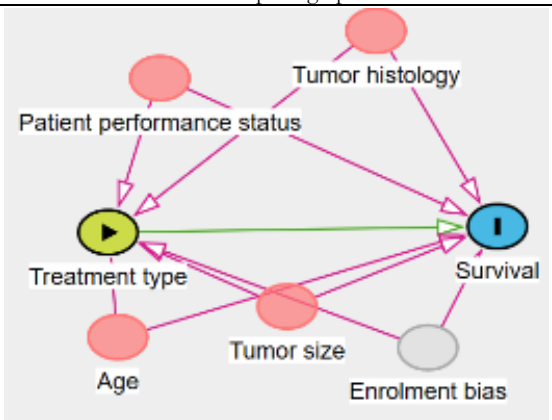
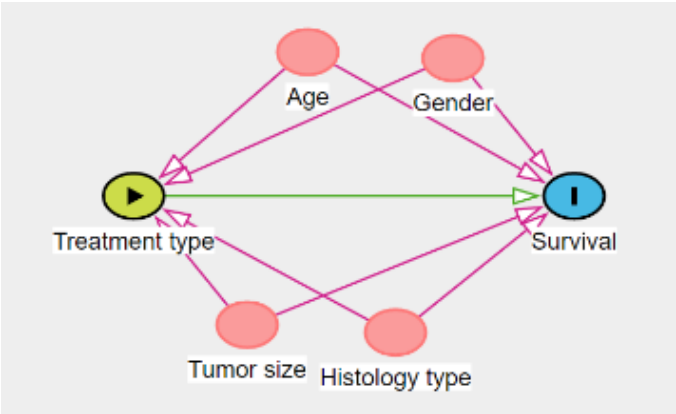
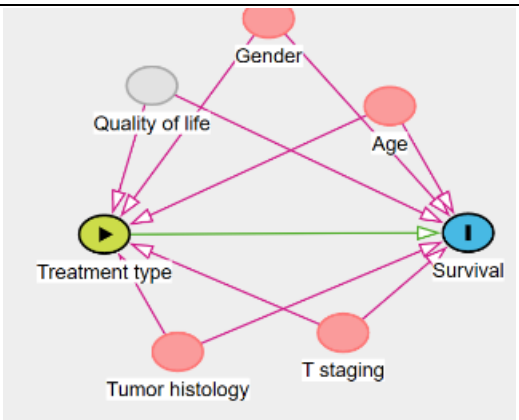
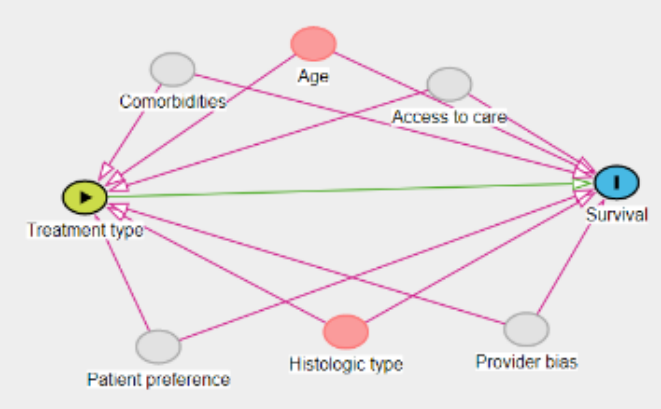
Appendix Table 2.1 Study characteristics

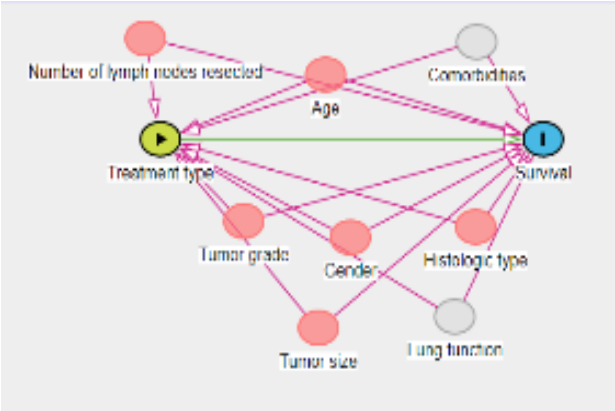
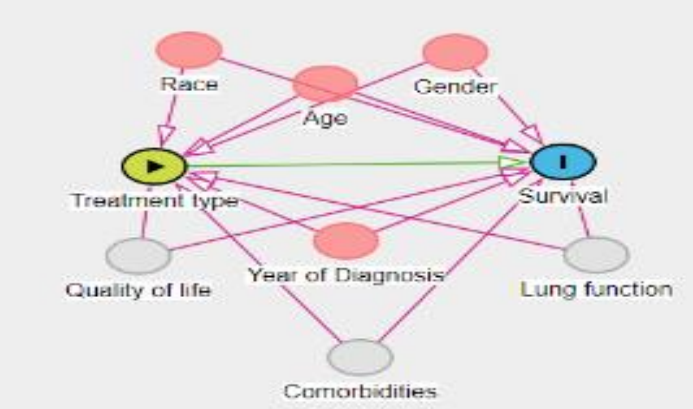
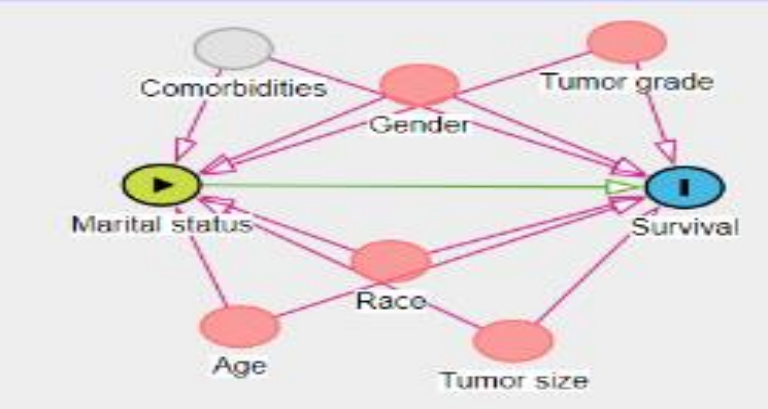
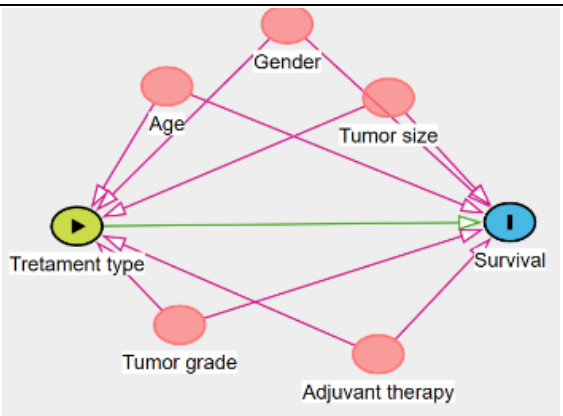
	Implied graph
1[33]	
2[34]	

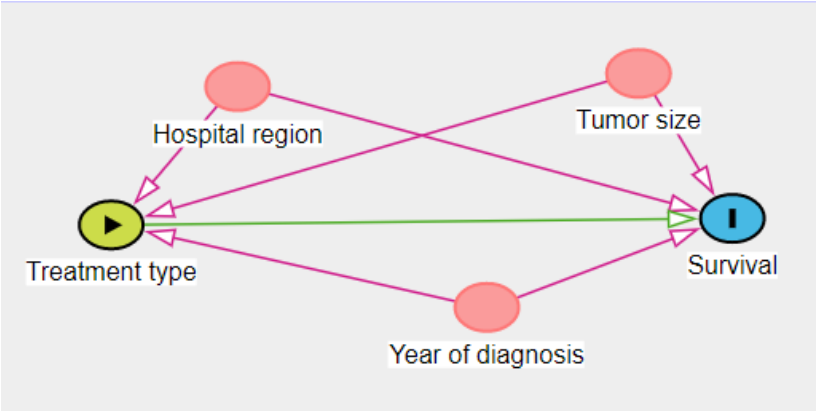
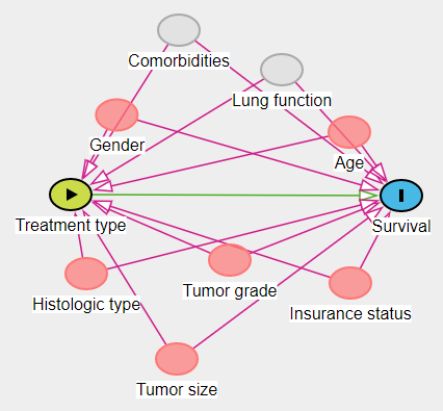
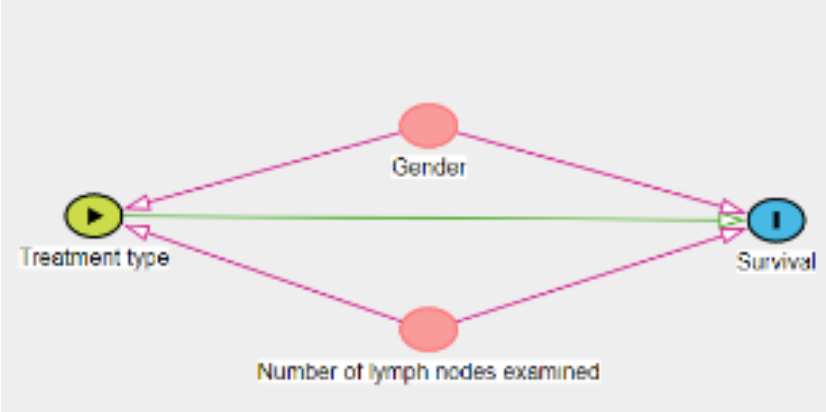
3[35]	
4[36]	
5[37]	
6[38]	

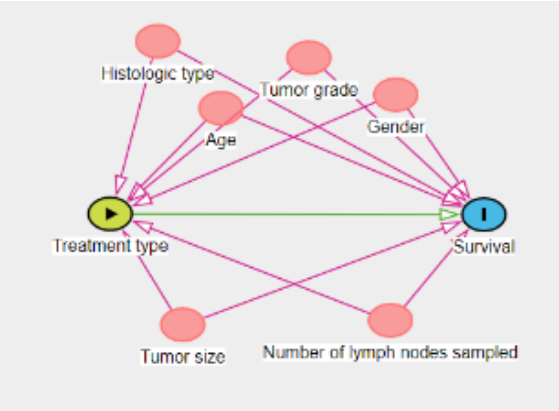
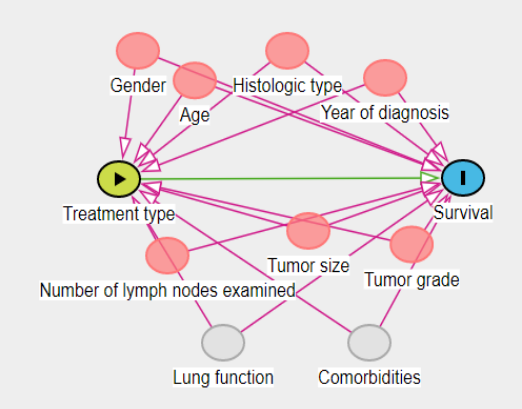
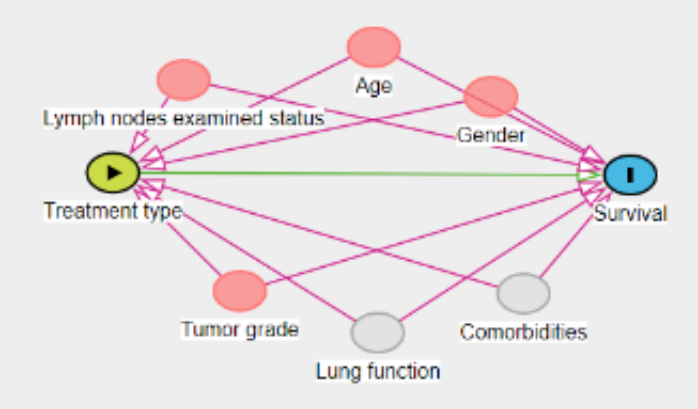
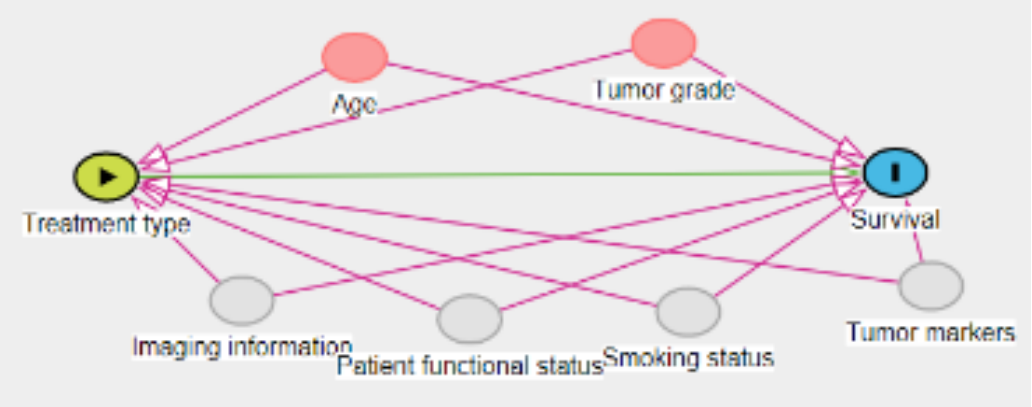
7[39]	
8[40]	

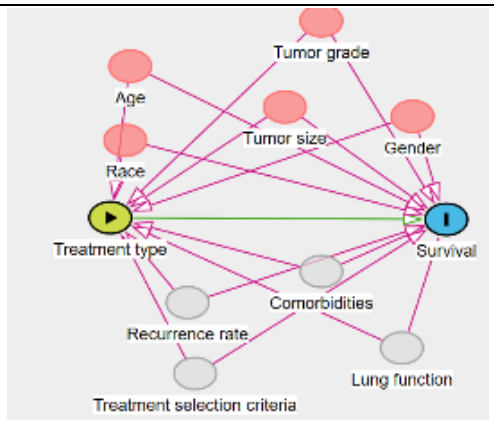
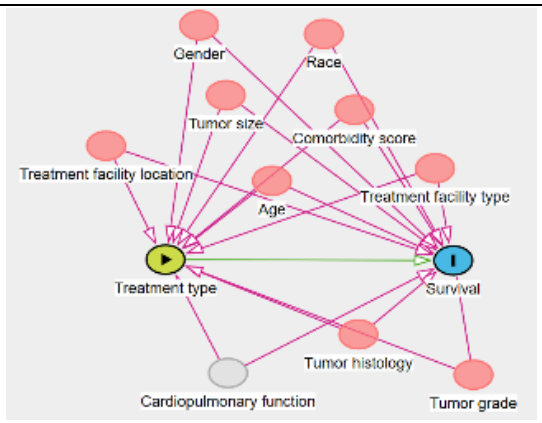
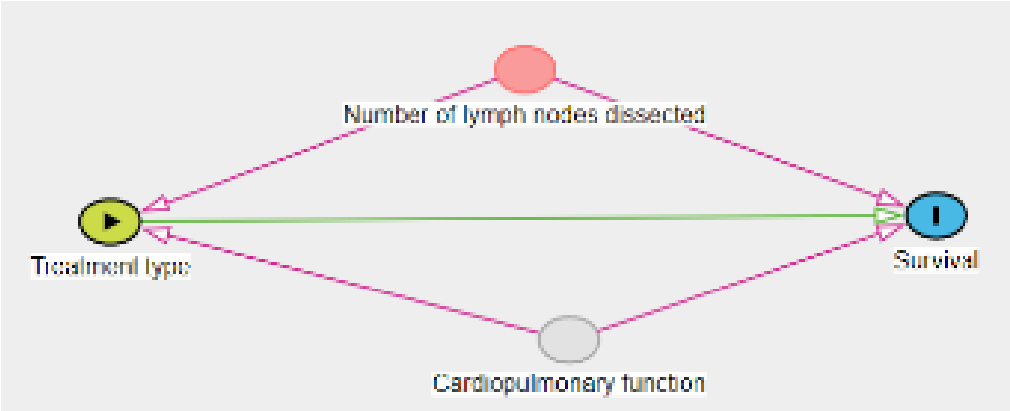
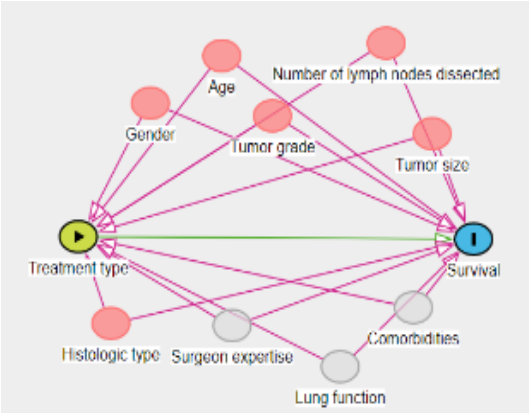
Appendix Table 2.2 Factors affecting treatment receipt mapping stage of ESC-DAG

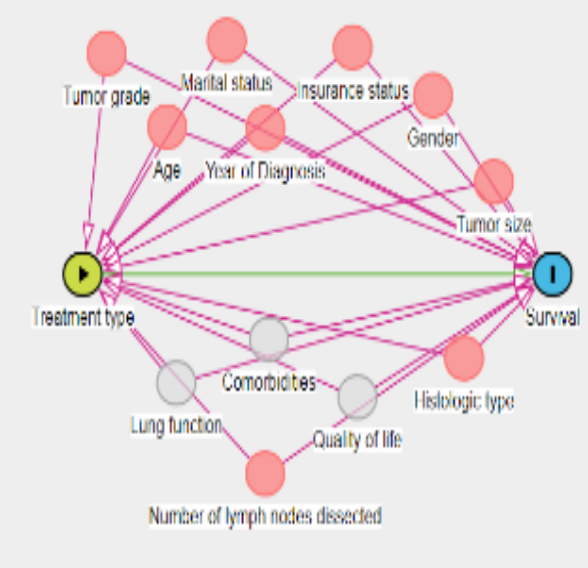
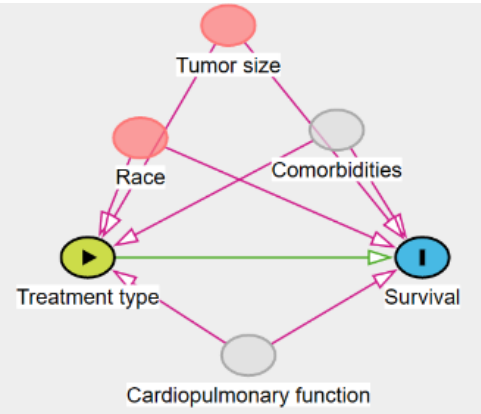
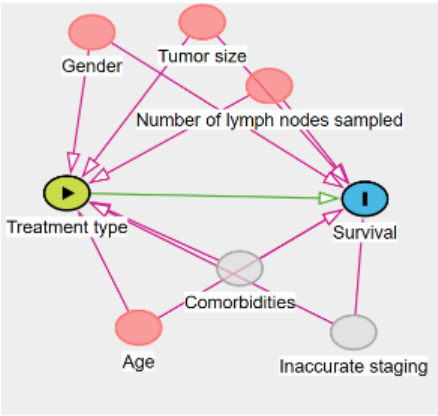
	Implied graph
1[41]	
2[42]	
3[43]	
4[44]	

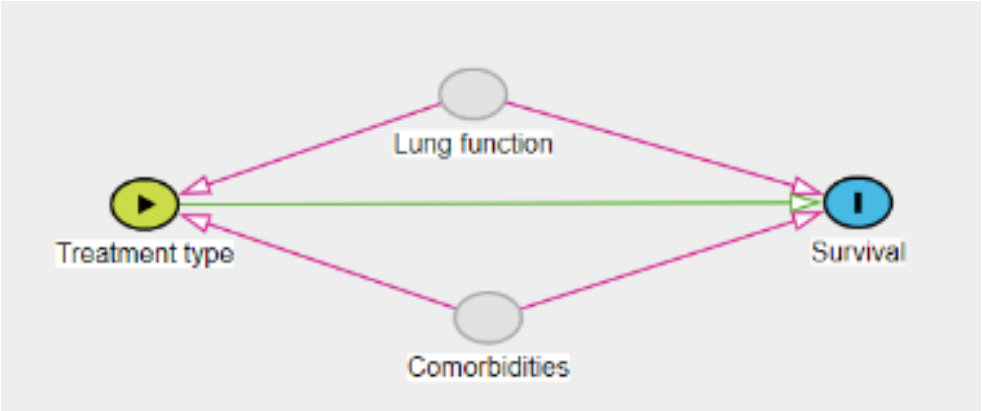
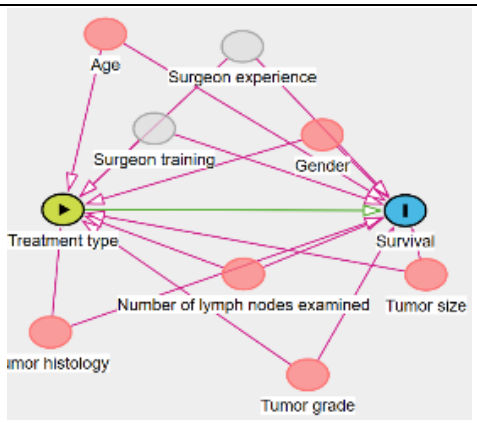
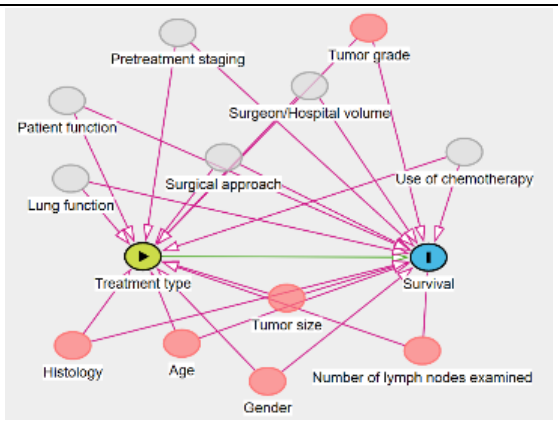
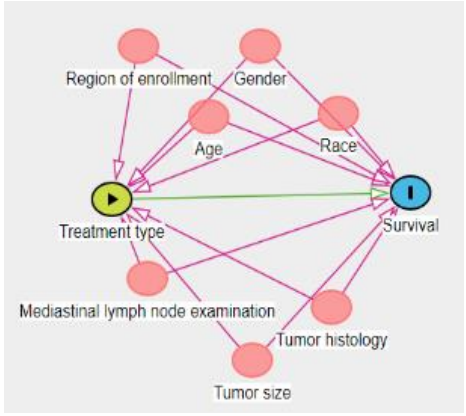
5[45]	
6[25]	
7[46]	
8[47]	

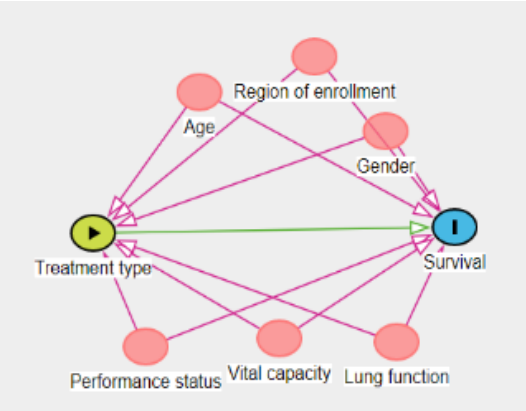
948	
1049	
1150	
1251	

135 2	
145 3	
155 4	
165 5	

175 6	
185 7	
195 8	
205 9	

216[0]	
226[1]	
236[2]	

24[6] [3]	
25[6] [4]	
26[6] [5]	
27[6] [6]	

28[6,7]	

Appendix Table 2.3 Factors affecting survival outcomes mapping stage of ESC-DAG

Study	Edge originates from	Edge terminates at	Bi-directional	Decision based on theory framework
Fossum et al. [38]	Comorbidity score	Treatment receipt	No	Retain
	Comorbidity score	Type of treatment facility	No	Retain
	Geographical area of patient	Treatment receipt	No	Retain
	Geographical area of patient	Type of treatment facility	No	Retain
	Insurance status	Treatment receipt	No	Retain
	Insurance status	Type of treatment facility	No	Retain
	Tumor size	Treatment receipt	No	Retain
	Tumor size	Type of treatment facility	No	Retain
	Type of treatment facility	Treatment receipt	No	Retain
	Age	Treatment receipt	No	Retain
	Age	Type of treatment facility	No	Retain
	Education	Treatment receipt	No	Retain
	Education	Type of treatment facility	No	Retain
	Gender	Treatment receipt	No	Retain
	Gender	Type of treatment facility	No	Retain
	Race	Treatment receipt	No	Retain
Dalwadi et al. [35]	Geographic region	Treatment receipt	No	Retain
	Insurance status	Geographic region	No	Retain
	Insurance status	Treatment receipt	No	Retain
	Marital status	Geographic region	No	Retain

	Marital status	Treatment receipt	No	Retain
	Age	Geographic region	No	Retain
	Age	Treatment receipt	No	Retain
	Gender	Geographic region	No	Retain
	Gender	Treatment receipt	No	Retain
	Income	Geographic region	No	Retain
	Income	Treatment receipt	No	Retain
	Race	Geographic region	No	Retain
	Race	Treatment receipt	No	Retain
Dalwadi et al. [36]	Geographic region	Treatment receipt	No	Retain
	Health status	Geographic region	No	Retain
	Health status	Treatment receipt	No	Retain
	Insurance status	Geographic region	No	Retain
	Insurance status	Treatment receipt	No	Retain
	Age	Geographic region	No	Retain
	Age	Treatment receipt	No	Retain
	Gender	Geographic region	No	Retain
	Gender	Treatment receipt	No	Retain
	Income	Geographic region	No	Retain
	Income	Treatment receipt	No	Retain
Balekian et al.[33]	Age at diagnosis	Treatment receipt	No	Retain
	Age at diagnosis	Race	No	Retain
	Coronary artery disease	Treatment receipt	No	Retain
	Coronary artery disease	Race	No	Retain
	Diagnosis after screening	Treatment receipt	No	Retain
	Diagnosis after screening	Race	No	Retain
	Smoking status	Treatment receipt	No	Retain
	Smoking status	Race	No	Retain
	Tumor histology	Treatment receipt	No	Retain
	Tumor histology	Race	No	Retain
	COPD	Treatment receipt	No	Retain
	COPD	Race	No	Retain
	Race	Treatment receipt	No	Retain
Berry et al.[34]	Factors associated with Rx receipt	Treatment receipt	No	Retain
	Insurance type	Factors associated with Rx receipt	No	Remove
	Insurance type	Treatment receipt	No	Retain
	Marital status	Factors associated with Rx receipt	No	Remove
	Marital status	Treatment receipt	No	Retain
	NCI hospital designation	Factors associated with Rx receipt	No	Remove
	NCI hospital designation	Treatment receipt	No	Retain
	SES status	Factors associated with Rx receipt	No	Remove
	SES status	Treatment receipt	No	Retain
	Tumor size	Factors associated with Rx receipt	No	Remove

	Tumor size	Treatment receipt	No	Retain
	Age	Factors associated with Rx receipt	No	Remove
	Age	Treatment receipt	No	Retain
	Gender	Factors associated with Rx receipt	No	Remove
	Gender	Treatment receipt	No	Retain
	Neighbourhood	Factors associated with Rx receipt	No	Remove
	Neighbourhood	Treatment receipt	No	Retain
	Race	Factors associated with Rx receipt	No	Remove
	Race	Treatment receipt	No	Retain
Dezube et al.[37]	Factors associated with therapy receipt	Treatment receipt	No	Retain
	Lower education	Factors associated with therapy receipt	No	Remove
	Lower education	Treatment receipt	No	Retain
	Lower median income	Factors associated with therapy receipt	No	Remove
	Lower median income	Treatment receipt	No	Retain
	Specialist availability in area	Factors associated with therapy receipt	No	Remove
	Specialist availability in area	Treatment receipt	No	Retain
	Age	Factors associated with therapy receipt	No	Remove
	Age	Treatment receipt	No	Retain
	Comorbidities	Factors associated with therapy receipt	No	Remove
	Comorbidities	Treatment receipt	No	Retain
	Frailty	Factors associated with therapy receipt	No	Remove
	Frailty	Treatment receipt	No	Retain
	Race	Factors associated with therapy receipt	No	Remove
	Race	Treatment receipt	No	Retain
Ganesh et al.[39]	Comorbidity score	Factors associated with treatment receipt	No	Remove
	Comorbidity score	Treatment receipt	No	Retain
	Factors associated with treatment receipt	Treatment receipt	No	Retain
	Geographic region	Factors associated with treatment receipt	No	Remove
	Geographic region	Treatment receipt	No	Retain
	Insurance status	Factors associated with treatment receipt	No	Remove
	Insurance status	Treatment receipt	No	Retain
	Rural/Urban region	Factors associated with treatment receipt	No	Remove
	Rural/Urban region	Treatment receipt	No	Retain
	Type of treatment facility	Factors associated with treatment receipt	No	Remove
	Type of treatment facility	Treatment receipt	No	Retain
	Gender	Factors associated with treatment receipt	No	Remove
	Gender	Treatment receipt	No	Retain
Iezzoni et al.[40]	Disability status	Treatment receipt	No	Retain
	Patient preference	Disability status	No	Retain
	Patient preference	Treatment receipt	No	Retain
	Physician preference	Disability status	No	Retain
	Physician preference	Treatment receipt	No	Retain
	Treatment facility info	Disability status	No	Retain

	Treatment facility info	Treatment receipt	No	Retain
	Age	Disability status	No	Retain
	Age	Treatment receipt	No	Retain

Appendix Table 2.4 Directed Edge index translation stage for factors affecting treatment receipt

Study	Edge originates from	Edge terminates at	Bi-directional	Decision based on theory framework
Dai et al.[42]	Histology type	Treatment type	No	Retain
	Histology type	Survival	No	Retain
	Treatment type	Survival	No	Retain
	Tumor size	Treatment type	No	Retain
	Tumor size	Survival	No	Retain
	Age	Treatment type	No	Retain
	Age	Survival	No	Retain
	Gender	Treatment type	No	Retain
	Gender	Survival	No	Retain
Dalwadi et al.[44]	Access to care	Treatment type	No	Retain
	Access to care	Survival	No	Retain
	Histologic type	Treatment type	No	Retain
	Histologic type	Survival	No	Retain
	Patient preference	Treatment type	No	Retain
	Patient preference	Survival	No	Retain
	Provider bias	Treatment type	No	Retain
	Provider bias	Survival	No	Retain
	Treatment type	Survival	No	Retain
	Age	Treatment type	No	Retain
	Age	Survival	No	Retain
	Comorbidities	Treatment type	No	Retain
	Comorbidities	Survival	No	Retain
Hao et al.[45]	Histologic type	Treatment type	No	Retain
	Histologic type	Survival	No	Retain
	Lung function	Treatment type	No	Retain
	Lung function	Survival	No	Retain
	Number of lymph nodes resected	Treatment type	No	Retain
	Number of lymph nodes resected	Survival	No	Retain
	Treatment type	Survival	No	Retain
	Tumor grade	Treatment type	No	Retain
	Tumor grade	Survival	No	Retain

	Tumor size	Treatment type	No	Retain
	Tumor size	Survival	No	Retain
	Age	Treatment type	No	Retain
	Age	Survival	No	Retain
	Comorbidities	Treatment type	No	Retain
	Comorbidities	Survival	No	Retain
	Gender	Treatment type	No	Retain
	Gender	Survival	No	Retain
Haque et al.[25]	Lung function	Treatment type	No	Retain
	Lung function	Survival	No	Retain
	Quality of life	Treatment type	No	Retain
	Quality of life	Survival	No	Retain
	Treatment type	Survival	No	Retain
	Year of Diagnosis	Treatment type	No	Retain
	Year of Diagnosis	Survival	No	Retain
	Age	Treatment type	No	Retain
	Age	Survival	No	Retain
	Comorbidities	Treatment type	No	Retain
	Comorbidities	Survival	No	Retain
	Gender	Treatment type	No	Retain
	Gender	Survival	No	Retain
	Race	Treatment type	No	Retain
	Race	Survival	No	Retain
Huang et al.[46]	Marital status	Survival	No	Retain
	Tumor grade	Marital status	No	Retain
	Tumor grade	Survival	No	Retain
	Tumor size	Marital status	No	Retain
	Tumor size	Survival	No	Retain
	Age	Marital status	No	Retain
	Age	Survival	No	Retain
	Comorbidities	Marital status	No	Retain
	Comorbidities	Survival	No	Retain
	Gender	Marital status	No	Retain
	Gender	Survival	No	Retain
	Race	Marital status	No	Retain
	Race	Survival	No	Retain
Kates et al.[63]	Lung function	Treatment type	No	Retain
	Lung function	Survival	No	Retain
	Treatment type	Survival	No	Retain
	Comorbidities	Treatment type	No	Retain
	Comorbidities	Survival	No	Retain
Li et al..[48]	Hospital region	Treatment type	No	Retain

	Hospital region	Survival	No	Retain
	Treatment type	Survival	No	Retain
	Tumor size	Treatment type	No	Retain
	Tumor size	Survival	No	Retain
	Year of diagnosis	Treatment type	No	Retain
	Year of diagnosis	Survival	No	Retain
Liang et al.[50]	Histologic type	Treatment type	No	Retain
	Histologic type	Survival	No	Retain
	Insurance status	Treatment type	No	Retain
	Insurance status	Survival	No	Retain
	Lung function	Treatment type	No	Retain
	Lung function	Survival	No	Retain
	Treatment type	Survival	No	Retain
	Tumor grade	Treatment type	No	Retain
	Tumor grade	Survival	No	Retain
	Tumor size	Treatment type	No	Retain
	Tumor size	Survival	No	Retain
	Age	Treatment type	No	Retain
	Age	Survival	No	Retain
	Comorbidities	Treatment type	No	Retain
	Comorbidities	Survival	No	Retain
	Gender	Treatment type	No	Retain
	Gender	Survival	No	Retain
Lin et al.[51]	Number of lymph nodes examined	Treatment type	No	Retain
	Number of lymph nodes examined	Survival	No	Retain
	Treatment type	Survival	No	Retain
	Gender	Treatment type	No	Retain
	Gender	Survival	No	Retain
Ni et al.[53]	Histologic type	Treatment type	No	Retain
	Histologic type	Survival	No	Retain
	Lung function	Treatment type	No	Retain
	Lung function	Survival	No	Retain
	Number of lymph nodes examined	Treatment type	No	Retain
	Number of lymph nodes examined	Survival	No	Retain
	Treatment type	Survival	No	Retain
	Tumor grade	Treatment type	No	Retain
	Tumor grade	Survival	No	Retain
	Tumor size	Treatment type	No	Retain
	Tumor size	Survival	No	Retain
	Year of diagnosis	Treatment type	No	Retain
	Year of diagnosis	Survival	No	Retain
	Age	Treatment type	No	Retain
	Age	Survival	No	Retain

	Comorbidities	Treatment type	No	Retain
	Comorbidities	Survival	No	Retain
	Gender	Treatment type	No	Retain
	Gender	Survival	No	Retain
Razi et al.[54]	Lung function	Treatment type	No	Retain
	Lung function	Survival	No	Retain
	Lymph nodes examined status	Treatment type	No	Retain
	Lymph nodes examined status	Survival	No	Retain
	Treatment type	Survival	No	Retain
	Tumor grade	Treatment type	No	Retain
	Tumor grade	Survival	No	Retain
	Age	Treatment type	No	Retain
	Age	Survival	No	Retain
	Comorbidities	Treatment type	No	Retain
	Comorbidities	Survival	No	Retain
	Gender	Treatment type	No	Retain
	Gender	Survival	No	Retain
Wang et al.[55]	Imaging information	Treatment type	No	Retain
	Imaging information	Survival	No	Retain
	Patient functional status	Treatment type	No	Retain
	Patient functional status	Survival	No	Retain
	Smoking status	Treatment type	No	Retain
	Smoking status	Survival	No	Retain
	Treatment type	Survival	No	Retain
	Tumor grade	Treatment type	No	Retain
	Tumor grade	Survival	No	Retain
	Tumor markers	Treatment type	No	Retain
	Tumor markers	Survival	No	Retain
	Age	Treatment type	No	Retain
	Age	Survival	No	Retain
Wang et al.[56]	Lung function	Treatment type	No	Retain
	Lung function	Survival	No	Retain
	Recurrence rate	Treatment type	No	Retain
	Recurrence rate	Survival	No	Retain
	Treatment selection criteria	Treatment type	No	Retain
	Treatment selection criteria	Survival	No	Retain
	Treatment type	Survival	No	Retain
	Tumor grade	Treatment type	No	Retain
	Tumor grade	Survival	No	Retain
	Tumor size	Treatment type	No	Retain
	Tumor size	Survival	No	Retain
	Age	Treatment type	No	Retain
	Age	Survival	No	Retain

	Comorbidities	Treatment type	No	Retain
	Comorbidities	Survival	No	Retain
	Gender	Treatment type	No	Retain
	Gender	Survival	No	Retain
	Race	Treatment type	No	Retain
	Race	Survival	No	Retain
Wu et al.[58]	Cardiopulmonary function	Treatment type	No	Retain
	Cardiopulmonary function	Survival	No	Retain
	Number of lymph nodes dissected	Treatment type	No	Retain
	Number of lymph nodes dissected	Survival	No	Retain
	Treatment type	Survival	No	Retain
Yendamuri et al.[59]	Histologic type	Treatment type	No	Retain
	Histologic type	Survival	No	Retain
	Lung function	Treatment type	No	Retain
	Lung function	Survival	No	Retain
	Number of lymph nodes dissected	Treatment type	No	Retain
	Number of lymph nodes dissected	Survival	No	Retain
	Surgeon expertise	Treatment type	No	Retain
	Surgeon expertise	Survival	No	Retain
	Treatment type	Survival	No	Retain
	Tumor grade	Treatment type	No	Retain
	Tumor grade	Survival	No	Retain
	Tumor size	Treatment type	No	Retain
	Tumor size	Survival	No	Retain
	Age	Treatment type	No	Retain
	Age	Survival	No	Retain
	Comorbidities	Treatment type	No	Retain
	Comorbidities	Survival	No	Retain
	Gender	Treatment type	No	Retain
	Gender	Survival	No	Retain
Yu et al.[60]	Histologic type	Treatment type	No	Retain
	Histologic type	Survival	No	Retain
	Insurance status	Treatment type	No	Retain
	Insurance status	Survival	No	Retain
	Lung function	Treatment type	No	Retain
	Lung function	Survival	No	Retain
	Marital status	Treatment type	No	Retain
	Marital status	Survival	No	Retain
	Number of lymph nodes dissected	Treatment type	No	Retain
	Number of lymph nodes dissected	Survival	No	Retain
	Quality of life	Treatment type	No	Retain
	Quality of life	Survival	No	Retain
	Treatment type	Survival	No	Retain

	Tumor grade	Treatment type	No	Retain
	Tumor grade	Survival	No	Retain
	Tumor size	Treatment type	No	Retain
	Tumor size	Survival	No	Retain
	Year of Diagnosis	Treatment type	No	Retain
	Year of Diagnosis	Survival	No	Retain
	Age	Treatment type	No	Retain
	Age	Survival	No	Retain
	Comorbidities	Treatment type	No	Retain
	Comorbidities	Survival	No	Retain
	Gender	Treatment type	No	Retain
	Gender	Survival	No	Retain
Ling et al.[52]	Histologic type	Treatment type	No	Retain
	Histologic type	Survival	No	Retain
	Number of lymph nodes sampled	Treatment type	No	Retain
	Number of lymph nodes sampled	Survival	No	Retain
	Treatment type	Survival	No	Retain
	Tumor grade	Treatment type	No	Retain
	Tumor grade	Survival	No	Retain
	Tumor size	Treatment type	No	Retain
	Tumor size	Survival	No	Retain
	Age	Treatment type	No	Retain
	Age	Survival	No	Retain
	Gender	Treatment type	No	Retain
	Gender	Survival	No	Retain
Clinical trial study NCT # NCT00109876[67]	Lung function	Treatment type	No	Retain
	Lung function	Survival	No	Retain
	Performance status	Treatment type	No	Retain
	Performance status	Survival	No	Retain
	Region of enrolment	Treatment type	No	Retain
	Region of enrolment	Survival	No	Retain
	Treatment type	Survival	No	Retain
	Vital capacity	Treatment type	No	Retain
	Vital capacity	Survival	No	Retain
	Age	Treatment type	No	Retain
	Age	Survival	No	Retain
	Gender	Treatment type	No	Retain
	Gender	Survival	No	Retain
Clinical trial NCT # NCT02357992[66]	Mediastinal lymph node examination	Treatment type	No	Retain
	Mediastinal lymph node examination	Survival	No	Retain
	Region of enrolment	Treatment type	No	Retain
	Region of enrolment	Survival	No	Retain
	Treatment type	Survival	No	Retain

	Tumour histology	Treatment type	No	Retain
	Tumour histology	Survival	No	Retain
	Tumour size	Treatment type	No	Retain
	Tumour size	Survival	No	Retain
	Age	Treatment type	No	Retain
	Age	Survival	No	Retain
	Gender	Treatment type	No	Retain
	Gender	Survival	No	Retain
	Race	Treatment type	No	Retain
	Race	Survival	No	Retain
Chang et al.[66]	Enrolment bias	Treatment type	No	Retain
	Enrolment bias	Survival	No	Retain
	Patient performance status	Treatment type	No	Retain
	Patient performance status	Survival	No	Retain
	Treatment type	Survival	No	Retain
	Tumor histology	Treatment type	No	Retain
	Tumor histology	Survival	No	Retain
	Tumor size	Treatment type	No	Retain
	Tumor size	Survival	No	Retain
	Age	Treatment type	No	Retain
	Age	Survival	No	Retain
Dalwadi et al.[43]	Quality of life	Treatment type	No	Retain
	Quality of life	Survival	No	Retain
	T staging	Treatment type	No	Retain
	T staging	Survival	No	Retain
	Treatment type	Survival	No	Retain
	Tumor histology	Treatment type	No	Retain
	Tumor histology	Survival	No	Retain
	Age	Treatment type	No	Retain
	Age	Survival	No	Retain
	Gender	Treatment type	No	Retain
	Gender	Survival	No	Retain
Li et al.[47]	Adjuvant therapy	Treatment type	No	Retain
	Adjuvant therapy	Survival	No	Retain
	Treatment type	Survival	No	Retain
	Tumor grade	Treatment type	No	Retain
	Tumor grade	Survival	No	Retain
	Tumor size	Treatment type	No	Retain
	Tumor size	Survival	No	Retain
	Age	Treatment type	No	Retain
	Age	Survival	No	Retain
	Gender	Treatment type	No	Retain
	Gender	Survival	No	Retain

Wang et al.[56]	Lung function	Treatment type	No	Retain
	Lung function	Survival	No	Retain
	Recurrence rate	Treatment type	No	Retain
	Recurrence rate	Survival	No	Retain
	Treatment selection criteria	Treatment type	No	Retain
	Treatment selection criteria	Survival	No	Retain
	Treatment type	Survival	No	Retain
	Tumor grade	Treatment type	No	Retain
	Tumor grade	Survival	No	Retain
	Tumor size	Treatment type	No	Retain
	Tumor size	Survival	No	Retain
	Age	Treatment type	No	Retain
	Age	Survival	No	Retain
	Comorbidities	Treatment type	No	Retain
	Comorbidities	Survival	No	Retain
	Gender	Treatment type	No	Retain
	Gender	Survival	No	Retain
	Race	Treatment type	No	Retain
	Race	Survival	No	Retain
Wu et al.[57]	Cardiopulmonary function	Treatment type	No	Retain
	Cardiopulmonary function	Survival	No	Retain
	Comorbidity score	Treatment type	No	Retain
	Comorbidity score	Survival	No	Retain
	Treatment facility location	Treatment type	No	Retain
	Treatment facility location	Survival	No	Retain
	Treatment facility type	Treatment type	No	Retain
	Treatment facility type	Survival	No	Retain
	Treatment type	Survival	No	Retain
	Tumor grade	Treatment type	No	Retain
	Tumor grade	Survival	No	Retain
	Tumor histology	Treatment type	No	Retain
	Tumor histology	Survival	No	Retain
	Tumor size	Treatment type	No	Retain
	Tumor size	Survival	No	Retain
	Age	Treatment type	No	Retain
	Age	Survival	No	Retain
	Gender	Treatment type	No	Retain
	Gender	Survival	No	Retain
	Race	Treatment type	No	Retain
	Race	Survival	No	Retain
Zeng et al.[61]	Cardiopulmonary function	Treatment type	No	Retain
	Cardiopulmonary function	Survival	No	Retain
	Treatment type	Survival	No	Retain

	Tumor size	Treatment type	No	Retain
	Tumor size	Survival	No	Retain
	Comorbidities	Treatment type	No	Retain
	Comorbidities	Survival	No	Retain
	Race	Treatment type	No	Retain
	Race	Survival	No	Retain
Chang et al.[62]	Inaccurate staging	Treatment type	No	Retain
	Inaccurate staging	Survival	No	Retain
	Number of lymph nodes sampled	Treatment type	No	Retain
	Number of lymph nodes sampled	Survival	No	Retain
	Treatment type	Survival	No	Retain
	Tumor size	Treatment type	No	Retain
	Tumor size	Survival	No	Retain
	Age	Treatment type	No	Retain
	Age	Survival	No	Retain
	Comorbidities	Treatment type	No	Retain
	Comorbidities	Survival	No	Retain
	Gender	Treatment type	No	Retain
	Gender	Survival	No	Retain
Ludwig et al.[64]	Number of lymph nodes examined	Treatment type	No	Retain
	Number of lymph nodes examined	Survival	No	Retain
	Surgeon experience	Treatment type	No	Retain
	Surgeon experience	Survival	No	Retain
	Surgeon training	Treatment type	No	Retain
	Surgeon training	Survival	No	Retain
	Treatment type	Survival	No	Retain
	Tumor grade	Treatment type	No	Retain
	Tumor grade	Survival	No	Retain
	Tumor histology	Treatment type	No	Retain
	Tumor histology	Survival	No	Retain
	Tumor size	Treatment type	No	Retain
	Tumor size	Survival	No	Retain
Whitson et al.[65]	Lung function	Treatment type	No	Retain
	Lung function	Survival	No	Retain
	Number of lymph nodes examined	Treatment type	No	Retain
	Number of lymph nodes examined	Survival	No	Retain
	Patient function	Treatment type	No	Retain
	Patient function	Survival	No	Retain
	Pre-treatment staging	Treatment type	No	Retain
	Pre-treatment staging	Survival	No	Retain
	Surgeon/Hospital volume	Treatment type	No	Retain
	Surgeon/Hospital volume	Survival	No	Retain
	Surgical approach	Treatment type	No	Retain

	Surgical approach	Survival	No	Retain
	Treatment type	Survival	No	Retain
	Tumor grade	Treatment type	No	Retain
	Tumor grade	Survival	No	Retain
	Tumor size	Treatment type	No	Retain
	Tumor size	Survival	No	Retain
	Use of chemotherapy	Treatment type	No	Retain
	Use of chemotherapy	Survival	No	Retain
	Age	Treatment type	No	Retain
	Age	Survival	No	Retain
	Gender	Treatment type	No	Retain
	Gender	Survival	No	Retain
	Histology	Treatment type	No	Retain
	Histology	Survival	No	Retain

Appendix Table 2.5 Directed Edge index translation stage for factors affecting survival outcomes

N	28,509	
	Frequency	Percentage
Tumor Grade		
Grade I	6,077	21.32
Grade II	10,769	37.77
Grade III	7,917	27.77
Grade IV	152	0.53
Unknown	3,594	12.61
Tumor size		
Upto 1cm	3,135	11.00
> 1cm & <=2cm	8,501	29.82
>2cm	6,359	22.31
Unknown size	10,514	36.88
Rural Urban Continum		
Large central metro	7,975	27.97
Large fringe metro	7,403	25.97
Medium metro	6,442	22.60
Non-metropolitan	6,689	23.46
Insurance type		
Only Medicaid	2,913	10.22
Only Medicare	8,021	28.13
Only Private	3,330	11.68
Uninsured	169	0.59
Unknown	14,076	49.37
Race		
Black	4,133	14.50
White	20,755	72.80
Unknown	3,621	12.70
Sex		
Female	15,127	53.06
Male	13,382	46.94
Marital Status		
Married	14,404	50.52
Widowed	4,807	16.86
Divorced	4,483	15.72
Single	2,053	7.20
Unknown	2,762	9.69
Treatment guideline revision years		
Pre 1996	4,065	14.26
1996 post	8,475	29.73
2005 post	625	2.19
2006 post	768	2.69
2007 post	6,298	22.09
2010 post	2,473	8.67
2012 post	1,142	4.01
2013 post	2,662	9.34
2015 post	2,000	7.02

Appendix Table 3.1a Excluded sample frequency statistics

N		28,509	
	Median	Mean	SD
Survival months	55	71.20	20.17
Age at diagnosis	66	65.15	9.62

Appendix Table 3.1b Excluded sample descriptive statistics

Step #	Description
1	Created data files from wide to long format for 6 components temperature average, snow, wind, precipitation, temperature minimum, temperature maximum
2	Keeping only U.S. weather data from previous step for each year
3	(1) Copied the contents of the file "ghcnd-stations.txt" (link: https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/ghcnd-stations.txt) and pasted in an excel file. (2) In the excel file, used "Text to Column" operator under "Data" menu and created 7 variables according to the explanation provided at this link: https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/readme.txt (3) Named four variables: "stationid" , "lat" , "lon" , "elevation". Deleted the rest. (4) Saved that excel file as Stata file
4	(1) Downloaded U.S. counties' geographical information from: "https://www.weather.gov/gis/Counties" from 5th April 2022 valid date column at bottom box where all zip folders are present and updated with a valid date from section (2) Among the 4 downloaded files in the package, opened "c_22mr22.dbf" in Microsoft Access. (3) Copied the contents of the file and pasted in an excel file.
5	(1) Formatting the data of source file by giving count number to each date by station. (2) Upon checking realized: There are many stations with incomplete set of dates. (3) Assigned a count number to each date under a station and numdate is the merging criterion for next step (4) Kept only one of counts per station ID to have an unrepeated list of stations (5) Each station ID is repeated exactly 365 times for non-leap years and 366 times otherwise (6) Assigning a count number to each repeat of station ID (7) There was exactly the same frequency under for each numDATE (8) Assigning a specific date to each "numDATE". (9) Renamed some latitude and longitude variables in the generated files
6	(1) Each FIPS is repeated exactly 365 times (2) Assign a count number to each repeat of FIPS (3) Assign a specific date to each "numDATE".
7	(1) Find nearest 3 weather grids by using geonearing command in STATA (2) Extract weather variables for the station IDs listed under variable "nid1", "nid2", "nid3". (3) Assigning weather info to FIPS for nearest stations 1-3 (4) Merging all 3 nearest stations into one file
8	Assigning population/county level resource information to the weather files corresponding each county FIPS
9	(1) Drop air pollution and AHRF variables (2) Generate 10-20 miles stations from country centroid with corresponding arithmetic mean weather component values (3) Generating monthly values from daily values. Calculating percentage missing, for each mile: 50% , 33.33%, 25%, and 20% (4) For each mile and each % missing four monthly measures are calculated: mean, median, max and iqr (5) Collapsing all miles, all components, all % missing, and all measures to assign corresponding only one value per month per FIPS
10	Appending all year weather files into one
11	Merging air pollution with weather files
12	After renaming variables the file is reshaped into wide format from long to achieve only one FIPS per row.

	The weather variables are separated from merged file to generate reshaped file and save separately for each mile, each component, and each %.

Appendix Table 4.1 Data cleaning steps for weather components

Step #	Description
1	(1) Rename, and clean raw files by generating date, day, year and month variables. (2) Keeping only one sample duration, and observations with non-zero latitude and longitude values (3) Generating unique site ID's by grouping corresponding latitude and longitude (4) Generating a variable for site monitors which allots unique site monitor, a unique day number for poc numbers (5) Generating a variable for site monitors which allots same number to different poc's per unique site ID with same day observation (6) Excluding observations with excluded even type (7) For a unique site ID only one observation is present as we keep only one poc per unique site ID
2	(1) Renaming and cleaning pollutant/weather data files to prep for merging (2) Assigning 3 nearest pollutant station monitor to the county centroid (3) Merging 1-3 nearest site values into one file
3	(1) Drop weather and AHRF variables (2) Generate 10-40 miles stations from country centroid with corresponding arithmetic mean pollutant values (3) Generating monthly values from daily values. Calculating percentage missing, for each mile: 50% , 33.33%, 25%, and 20% (4) For each mile and each % missing four monthly measures are calculated: mean, median, max and iqr (5) Collapsing all miles, all % missing, and all measures to assign corresponding only one value per month per FIPS
4	Appending all years, all pollutants files into one and assigning
5	Merging Air pollution with Weather files
6	After renaming variables the file is reshaped into wide format from long to achieve only one FIPS per row. The Air pollutant variables are separated from merged file to generate reshaped file and save separately for each pollutant each mile, each component, and each %.

Appendix Table 4.2 Data cleaning steps for air pollutants

	Multipollutant			NO2			SO2			CO		
	Hazard of death 1 year after diagnosis			Hazard of death 1 year after diagnosis			Hazard of death 1 year after diagnosis			Hazard of death 1 year after diagnosis		
	Duration of exposure from diagnosis			Duration of exposure from diagnosis			Duration of exposure from diagnosis			Duration of exposure from diagnosis		
	1 yr bf	3 yrs bf	5 yrs bf	1 yr bf	3 yrs bf	5 yrs bf	1 yr bf	3 yrs bf	5 yrs bf	1 yr bf	3 yrs bf	5 yrs bf
Air pollutants and weather elements												
NO2	1.04	1.06*	1.09**	1.08***	1.11***	1.14***						
	(.98 , 1.1)	(1 , 1.14)	(1.01 , 1.17)	(1.03 , 0.00)	(1.05 , 0.00)	(1.08 , 0.00)						
SO2	1.32***	1.43***	1.45***				1.29***	1.37***	1.36***			
	(1.18 , 1.48)	(1.26 , 1.61)	(1.28 , 1.63)				(1.15 , 1.44)	(1.22 , 1.55)	(1.21 , 1.52)			
CO	3.02***	3.39***	3.81***							3.26***	4.02***	4.88***
	(1.64 , 5.56)	(1.68 , 6.85)	(1.8 , 8.03)							(1.96 , 5.43)	(2.32 , 6.96)	(2.81 , 8.49)
Precipitation	1.01	1.00	1.10	0.97	0.91	0.99	1.00	0.98	1.06	0.99	0.94	1.02
	(.9 , 1.13)	(.84 , 1.18)	(.9 , 1.35)	(.83 , 1.14)	(.72 , 1.16)	(.8 , 1.22)	(.88 , 1.15)	(.81 , 1.19)	(.85 , 1.31)	(.85 , 1.14)	(.75 , 1.16)	(.83 , 1.25)
Snow	2.02	0.04	0.00**	1.29	0.16	0.05	3.42	0.36	0.19	2.03	0.09	.02*
	(.03 , 133.75)	(00.0 , 3.23)	(0.00 , 0.90)	(.04 , 37.86)	(.01 , 4.42)	(0.00 , 2.76)	(.16 , 75.38)	(.01 , 12.09)	(0.00 , 9.55)	(.06 , 65.81)	(0.00 , 4.15)	(0.00 , 1.5)
Daily temperature minimum	1.02*	1.02	1.04***	1.02*	1.02*	1.04***	1.00	0.99	1.00	1.02*	1.02*	1.04***
	(1 , 1.04)	(1 , 1.04)	(1.01 , 1.06)	(1 , 1.04)	(1. , 1.04)	(1.01 , 1.07)	(.99 , 1.02)	(.97 , 1.01)	(.97 , 1.02)	(1. , 1.04)	(1. , 1.04)	(1.01 , 1.07)
Treatment options (reference :lobectomy)												
Limited resection with	0.41	0.30	0.33	0.82	0.66	0.73	0.79	0.69	0.66	1.08	0.87	0.88
adjuvant radiotherapy	(.08 , 2.09)	(.06 , 1.5)	(.07 , 1.59)	(.21 , 3.2)	(.18 , 2.5)	(.21 , 2.59)	(.17 , 3.66)	(.15 , 3.23)	(.14 , 3.03)	(.28 , 4.16)	(.23 , 3.21)	(.25 , 3.1)
Treatment interaction with air pollutant and weather elements												
NO2 * Treatment	1.05	1.06**	1.05**	1.02**	1.03***	1.03**						
	(1.01 , 1.1)	(1.01 , 1.1)	(1.01 , 1.1)	(1. , 1.04)	(1.01 , 1.05)	(1.01 , 1.05)						
SO2 * Treatment	1.14*	1.15*	1.15**				1.14*	1.14*	1.14*			
	(.98 , 1.32)	(1. , 1.32)	(1.01 , 1.32)				(.98 , 1.32)	(.99 , 1.31)	(.99 , 1.31)			
CO * Treatment	0.5	0.52	0.53							1.25	1.29	1.28
	(.21 , 1.2)	(.21 , 1.28)	(.22 , 1.3)							(.82 , 1.92)	(.85 , 1.96)	(.84 , 1.94)
Precipitation * Treatment	1.15	1.1	1.14	1.15	1.1	1.12	1.11	1.07	1.09	1.13	1.07	1.07
	(.83 , 1.59)	(.82 , 1.49)	(.85 , 1.54)	(.86 , 1.55)	(.84 , 1.46)	(.85 , 1.46)	(.81 , 1.54)	(.8 , 1.43)	(.81 , 1.47)	(.83 , 1.55)	(.79 , 1.44)	(.79 , 1.45)
Snow * Treatment	0.00*	0.36	0.04	0	0.21	0.02	0.00*	0.20	0.03	0.00*	0.15	0.02
	(0.00 , 1.89)	(0.00 , 618.98)	(0.00 , 323.52)	(0.00 , 1.83)	(0.00 , 276.44)	(0.00 , 179.91)	(0.00 , 1.7)	(0.00 , 191.19)	(0.00 , 139.29)	(0.00 , 1.3)	(0.00 , 235.13)	(0.00 , 128.5)
Temperature minimum * Treatment	1	1	1	0.99	1	1	1	1	1	1	1	1
	(.98 , 1.02)	(.98 , 1.02)	(.98 , 1.02)	(.98 , 1.01)	(.98 , 1.01)	(.98 , 1.01)	(.98 , 1.01)	(.98 , 1.02)	(.98 , 1.02)	(.98 , 1.01)	(.98 , 1.01)	(.98 , 1.01)

p values: * <0.1%, ** < 0.05%, *** < 0.01%

Appendix Table 4.3 Hazards of death one year after diagnosis for annual average of monthly median values

	Multipollutant			NO2			SO2			CO		
	Hazard of death 3 year after diagnosis			Hazard of death 3 year after diagnosis			Hazard of death 3 year after diagnosis			Hazard of death 3 year after diagnosis		
	Duration of exposure from diagnosis			Duration of exposure from diagnosis			Duration of exposure from diagnosis			Duration of exposure from diagnosis		
	1 yr bf	3 yrs bf	5 yrs bf	1 yr bf	3 yrs bf	5 yrs bf	1 yr bf	3 yrs bf	5 yrs bf	1 yr bf	3 yrs bf	5 yrs bf
Air pollutants and weather elements												
NO2	1.04***	1.06***	1.08***	1.07***	1.09***	1.13***						
	(1.01 , 1.07)	(1.02 , 1.09)	(1.05 , 1.12)	(1.04 , 1.81)	(1.06 , 1.49)	(1.1 , 5.78)						
SO2	1.21***	1.23***	1.23***				1.21***	1.22***	1.21***			
	(1.15 , 1.27)	(1.16 , 1.29)	(1.17 , 1.29)				(1.15 , 1.27)	(1.16 , 1.29)	(1.15 , 1.27)			
CO	1.79***	1.91***	2.11***							2.24***	2.62***	3.23***
	(1.3 , 2.46)	(1.37 , 2.65)	(1.51 , 2.95)							(1.69 , 2.96)	(1.97 , 3.48)	(2.43 , 4.29)
Precipitation	1	0.95	1	0.95	.87**	0.91	1	0.97	1.02	0.96	.88**	0.92
	(.92 , 1.08)	(.85 , 1.07)	(.89 , 1.13)	(.87 , 1.04)	(.77 , .99)	(.81 , 1.03)	(.92 , 1.09)	(.86 , 1.08)	(.9 , 1.15)	(.87 , 1.05)	(.78 , .99)	(.82 , 1.04)
Snow	1.1	0.34	.04**	0.49	0.49	.11**	1.58	1.18	0.47	1.3	0.44	.06**
	(.13 , 9.38)	(.04 , 2.68)	(0.00 , .48)	(.07 , 3.26)	(.08 , 2.85)	(.01 , .94)	(.25 , 10.02)	(.18 , 7.55)	(.05 , 4.05)	(.19 , 8.94)	(.06 , 3.25)	(.01 , .66)
Daily temperature minimum	1.01	1.01*	1.03***	1.01	1.01*	1.03***	0.99	.99**	1	1.01	1.01*	1.03***
	(1. , 1.02)	(1. , 1.02)	(1.02 , 1.05)	(1. , 1.02)	(1. , 1.02)	(1.02 , 1.05)	(.98 , 1.)	(.98 , 1.)	(.98 , 1.01)	(1. , 1.02)	(1. , 1.02)	(1.01 , 1.04)
Treatment options (reference :lobectomy)												
Limited resection with	0.85	0.81	0.84	0.66	0.63	0.65	1.18	1.12	1.05	0.79	0.75	0.74
adjuvant radiotherapy	(.46 , 1.59)	(.42 , 1.57)	(.42 , 1.67)	(.37 , 1.16)	(.35 , 1.12)	(.36 , 1.17)	(.68 , 2.07)	(.61 , 2.04)	(.56 , 1.98)	(.46 , 1.35)	(.43 , 1.31)	(.42 , 1.29)
Treatment interaction with air pollutant and weather elements												
NO2 * Treatment	1.02	1.02	1.01	1.02***	1.02***	1.01**						
	(1. , 1.04)	(.99 , 1.04)	(.99 , 1.04)	(1.01 , 1.03)	(1.01 , 1.03)	(1. , 1.02)						
SO2 * Treatment	0.97	0.98	0.98				0.99	1	1			
	(.89 , 1.04)	(.91 , 1.05)	(.91 , 1.05)				(.92 , 1.06)	(.93 , 1.07)	(.93 , 1.07)			
CO * Treatment	1.01	1.03	1.03							1.37***	1.33***	1.28**
	(.65 , 1.57)	(.67 , 1.6)	(.66 , 1.59)							(1.1 , 1.69)	(1.08 , 1.64)	(1.05 , 1.56)
Precipitation * Treatment	0.99	0.93	0.92	0.98	0.93	0.92	0.98	0.94	0.94	0.97	0.93	0.91
	(.81 , 1.2)	(.77 , 1.14)	(.75 , 1.13)	(.8 , 1.19)	(.77 , 1.13)	(.76 , 1.12)	(.81 , 1.2)	(.77 , 1.14)	(.77 , 1.15)	(.79 , 1.18)	(.76 , 1.13)	(.75 , 1.12)
Snow * Treatment	0.8	1.31	1.27	1.18	1.58	1.54	0.63	0.78	0.77	0.99	1.53	1.65
	(.07 , 9.34)	(.13 , 13.34)	(.08 , 20.23)	(.12 , 11.83)	(.16 , 15.18)	(.11 , 22.3)	(.06 , 6.8)	(.09 , 7.03)	(.06 , 10.18)	(.1 , 9.97)	(.16 , 14.58)	(.11 , 24.11)
Temperature minimum* Treatment	1	1	1	1	1	1	1	1	1	1	1	1
	(.99 , 1.)	(.99 , 1.01)	(.99 , 1.01)	(.99 , 1.01)	(1. , 1.01)	(1. , 1.01)	(.99 , 1.)	(.99 , 1.)	(.99 , 1.01)	(.99 , 1.01)	(.99 , 1.01)	(.99 , 1.01)

p values: * <0.1%, ** < 0.05%, *** < 0.01%

Appendix Table 4.4 Hazards of death three years after diagnosis for annual average of monthly median values

	Multipollutant			NO2			SO2			CO		
	Hazard of death 5 year after diagnosis			Hazard of death 5 year after diagnosis			Hazard of death 5 year after diagnosis			Hazard of death 5 year after diagnosis		
	Duration of exposure from diagnosis			Duration of exposure from diagnosis			Duration of exposure from diagnosis			Duration of exposure from diagnosis		
	1 yr bf	3 yrs bf	5 yrs bf	1 yr bf	3 yrs bf	5 yrs bf	1 yr bf	3 yrs bf	5 yrs bf	1 yr bf	3 yrs bf	5 yrs bf
Air pollutants and weather elements												
NO2	1.04***	1.05***	1.08***	1.05***	1.07***	1.1***						
	(1.02 , 1.06)	(1.03 , 1.08)	(1.06 , 1.11)	(1.03 , 1.62)	(1.05 , 1.81)	(1.08 , 8.09)						
SO2	1.18***	1.19***	1.19***				1.17***	1.18***	1.17***			
	(1.12 , 1.23)	(1.14 , 1.24)	(1.14 , 1.24)				(1.12 , 1.23)	(1.13 , 1.23)	(1.12 , 1.22)			
CO	1.39***	1.41***	1.52***							1.73***	1.89***	2.27***
	(1.09 , 1.78)	(1.1 , 1.81)	(1.17 , 1.96)							(1.39 , 2.14)	(1.51 , 2.36)	(1.81 , 2.85)
Precipitation	1.01	1	1.08	0.96	0.93	0.98	1.02	1.02	1.09*	0.97	0.93	0.98
	(.94 , 1.09)	(.91 , 1.1)	(.97 , 1.2)	(.89 , 1.04)	(.84 , 1.03)	(.89 , 1.09)	(.95 , 1.1)	(.93 , 1.12)	(.99 , 1.21)	(.9 , 1.04)	(.84 , 1.02)	(.89 , 1.09)
Snow	0.76	0.28	0.05***	0.45	0.29	.09***	1.24	0.8	0.44	1.02	0.38	.09***
	(.16 , 3.64)	(.06 , 1.34)	(.01 , .33)	(.1 , 1.94)	(.06 , 1.3)	(.02 , .5)	(.3 , 5.07)	(.18 , 3.6)	(.08 , 2.31)	(.24 , 4.31)	(.08 , 1.8)	(.02 , .54)
Daily temperature minimum	1.01	1.01**	1.03***	1.01	1.01**	1.03***	0.99**	0.99***	0.99*	1	1.01	1.02***
	(1. , 1.02)	(1. , 1.02)	(1.02 , 1.04)	(1. , 1.02)	(1. , 1.02)	(1.02 , 1.04)	(.98 , 1.)	(.98 , .99)	(.98 , 1.)	(1. , 1.01)	(1. , 1.02)	(1.01 , 1.03)
Treatment options (reference :lobectomy)												
Limited resection with	1.01	0.99	1.01	0.74	0.71	0.75	1.36	1.32	1.25	0.9	0.85	0.84
adjuvant radiotherapy	(.64 , 1.59)	(.62 , 1.59)	(.62 , 1.64)	(.49 , 1.13)	(.47 , 1.09)	(.49 , 1.13)	(.89 , 2.07)	(.85 , 2.06)	(.79 , 1.98)	(.6 , 1.36)	(.57 , 1.28)	(.56 , 1.26)
Treatment interaction with air pollutant and weather elements												
NO2 * Treatment	1.02	1.02**	1.02*	1.02***	1.02***	1.01***						
	(1. , 1.04)	(1. , 1.04)	(1. , 1.03)	(1.01 , 1.03)	(1.01 , 1.03)	(1.01 , 1.02)						
SO2 * Treatment	0.95	0.95	0.96				0.98	0.98	0.98			
	(.88 , 1.02)	(.89 , 1.02)	(.9 , 1.03)				(.91 , 1.05)	(.91 , 1.05)	(.92 , 1.05)			
CO * Treatment	0.99	1.01	1.01							1.32***	1.31***	1.29***
	(.7 , 1.41)	(.71 , 1.44)	(.71 , 1.43)							(1.07 , 1.63)	(1.07 , 1.61)	(1.07 , 1.56)
Precipitation * Treatmer	0.93	0.94	0.94	0.92	0.93	0.93	0.94	0.94	0.95	0.91	0.93	0.92
	(.81 , 1.07)	(.84 , 1.05)	(.84 , 1.05)	(.8 , 1.06)	(.83 , 1.04)	(.83 , 1.04)	(.82 , 1.07)	(.84 , 1.05)	(.85 , 1.06)	(.79 , 1.05)	(.83 , 1.03)	(.82 , 1.03)
Snow * Treatment	0.84	1.04	1.03	1.12	1.32	1.27	0.64	0.65	0.67	1.11	1.34	1.51
	(.15 , 4.76)	(.18 , 5.89)	(.15 , 6.96)	(.2 , 6.26)	(.22 , 8.02)	(.18 , 9.03)	(.12 , 3.5)	(.12 , 3.55)	(.1 , 4.23)	(.21 , 5.84)	(.24 , 7.6)	(.23 , 9.93)
Temperature minimum * Treatment	1	1	1	1	1	1	1	1	1	1	1	1
	(.99 , 1.)	(.99 , 1.)	(.99 , 1.)	(1. , 1.)	(1. , 1.)	(1. , 1.01)	(.99 , 1.)	(.99 , 1.)	(.99 , 1.)	(.99 , 1.)	(1. , 1.)	(1. , 1.01)

p values: * <0.1%, ** < 0.05%, *** < 0.01%

Appendix Table 4.5 Hazards of death five years after diagnosis for annual average of monthly median values

	Multipollutant			NO2			SO2			CO		
	Hazard of death 1 year after diagnosis			Hazard of death 1 year after diagnosis			Hazard of death 1 year after diagnosis			Hazard of death 1 year after diagnosis		
	Duration of exposure from diagnosis			Duration of exposure from diagnosis			Duration of exposure from diagnosis			Duration of exposure from diagnosis		
	1 yr bf	3 yrs bf	5 yrs bf	1 yr bf	3 yrs bf	5 yrs bf	1 yr bf	3 yrs bf	5 yrs bf	1 yr bf	3 yrs bf	5 yrs bf
Air pollutants and weather elements												
NO2	1.01	1.01	0.99	1.06***	1.07***	1.06***						
	(.98 , 1.05)	(.98 , 1.05)	(.95 , 1.03)	(1.03 , 0.00)	(1.04 , 0.00)	(1.03 , 0.00)						
SO2	1.03**	1.05***	1.07***				1.04**	1.06***	1.08***			
	(1. , 1.06)	(1.02 , 1.08)	(1.03 , 1.1)				(1.01 , 1.07)	(1.03 , 1.09)	(1.05 , 1.11)			
CO	2.66***	3.44***	4.55***							2.82***	3.85***	4.26***
	(1.81 , 3.91)	(2.24 , 5.3)	(2.71 , 7.62)							(2.03 , 3.93)	(2.62 , 5.67)	(2.75 , 6.59)
Precipitation	1**	0.99***	0.99***	1.**	.99***	.99**	1.00	1.00	1.00	1**	0.99***	0.99***
	(0.99 , 1)	(0.99 , 1)	(0.99 , 1)	(.99 , 1.)	(.99 , 1.)	(.99 , 1.)	(1. , 1.)	(.99 , 1.)	(.99 , 1.)	(.99 , 1.)	(.99 , 1.)	(.99 , 1.)
Snow	0.99	0.98***	0.97***	0.99	.98**	.97***	1.00	0.99	.98***	0.99	0.98***	0.97***
	(0.98 , 1)	(0.97 , 1)	(0.95 , 0.98)	(.98 , 1.)	(.97 , 1.)	(.95 , .99)	(.98 , 1.01)	(.98 , 1.)	(.96 , .99)	(.98 , 1.01)	(.97 , 1.)	(.95 , .98)
Daily temperature minimum	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.01	1.00	1.00
	(1. , 1.02)	(.99 , 1.02)	(.98 , 1.02)	(.99 , 1.02)	(.99 , 1.02)	(.98 , 1.02)	(.99 , 1.01)	(.98 , 1.01)	(.97 , 1.01)	(1. , 1.02)	(.99 , 1.02)	(.98 , 1.02)
Treatment options (reference :lobectomy)												
Limited resection with	0.52	0.21	0.21	1.32	0.77	0.8	1.55	0.38	0.3	1.88	0.9	0.74
adjuvant radiotherapy	(.03 , 10.12)	(.01 , 8.01)	(. , 8.84)	(.09 , 19.47)	(.04 , 15.43)	(.03 , 19.37)	(.07 , 32.14)	(.01 , 12.08)	(.01 , 9.36)	(.11 , 32.45)	(.03 , 25.63)	(.02 , 25.03)
Treatment interaction with air pollutant and weather elements												
NO2 * Treatment	1.04	1.04**	1.04**	1.02***	1.02***	1.02***						
	(1.01 , 1.08)	(1.01 , 1.08)	(1. , 1.07)	(1.01 , 1.03)	(1.01 , 1.03)	(1. , 1.03)						
SO2 * Treatment	1.03*	1.04**	1.04**				1.04**	1.05**	1.05**			
	(.99 , 1.08)	(1 , 1.08)	(1 , 1.08)				(1. , 1.08)	(1.01 , 1.09)	(1.01 , 1.09)			
CO * Treatment	0.64	0.64	0.65							1.23*	1.2	1.17
	(.35 , 1.16)	(.35 , 1.19)	(.35 , 1.21)							(.98 , 1.54)	(.93 , 1.54)	(.91 , 1.51)
Precipitation * Treatment	1	1	1	1	1	1	1	1	1	1	1	1
	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)
Snow * Treatment	1	1	0.99	1	1	1	1	1	1	1	1	1
	(.98 , 1.01)	(.98 , 1.02)	(.97 , 1.02)	(.98 , 1.02)	(.98 , 1.02)	(.98 , 1.02)	(.98 , 1.01)	(.98 , 1.02)	(.98 , 1.02)	(.98 , 1.01)	(.98 , 1.02)	(.98 , 1.02)
Temperature minimum * Treatment	1	1	1	0.99	1	1	0.99	1	1.01	0.99	1	1
	(.97 , 1.02)	(.97 , 1.03)	(.97 , 1.03)	(.97 , 1.01)	(.97 , 1.02)	(.97 , 1.02)	(.97 , 1.02)	(.98 , 1.03)	(.98 , 1.03)	(.97 , 1.02)	(.97 , 1.03)	(.97 , 1.03)

p values: * <0.1%, ** < 0.05%, *** < 0.01%

Appendix Table 4.6 Hazards of death one year after diagnosis for annual average of monthly maximum values

	Multipollutant			NO2			SO2			CO		
	Hazard of death 3 year after diagnosis			Hazard of death 3 year after diagnosis			Hazard of death 3 year after diagnosis			Hazard of death 3 year after diagnosis		
	Duration of exposure from diagnosis			Duration of exposure from diagnosis			Duration of exposure from diagnosis			Duration of exposure from diagnosis		
	1 yr bf	3 yrs bf	5 yrs bf	1 yr bf	3 yrs bf	5 yrs bf	1 yr bf	3 yrs bf	5 yrs bf	1 yr bf	3 yrs bf	5 yrs bf
Air pollutants and weather elements												
NO2	1.03***	1.03***	1.02*	1.05***	1.06***	1.06***						
	(1.01 , 1.04)	(1.01 , 1.05)	(1. , 1.04)	(1.04 , 1.7)	(1.04 , 2.01)	(1.05 , 2.83)						
SO2	1.04***	1.04***	1.04***				1.04***	1.05***	1.05***			
	(1.02 , 1.05)	(1.02 , 1.06)	(1.03 , 1.06)				(1.02 , 1.05)	(1.03 , 1.06)	(1.03 , 1.07)			
CO	1.95***	2.13***	2.27***							2.28***	2.63***	2.77***
	(1.61 , 2.36)	(1.71 , 2.65)	(1.78 , 2.9)							(1.92 , 2.73)	(2.18 , 3.19)	(2.28 , 3.36)
Precipitation	1***	1***	1***	1***	1***	1**	1	1	1	1.***	1.***	1.***
	(1. , 1.)	(.99 , 1.)	(.99 , 1.)	(1. , 1.)	(1. , 1.)	(1. , 1.)	(1. , 1.)	(1. , 1.)	(1. , 1.)	(1. , 1.)	(.99 , 1.)	(.99 , 1.)
Snow	0.99*	0.99***	0.98***	.99*	.99***	.98***	1	.99*	.99***	1	.99**	.98***
	(.99 , 1.)	(.98 , 1.)	(.97 , .99)	(.99 , 1.)	(.98 , 1.)	(.97 , .99)	(.99 , 1.)	(.99 , 1.)	(.98 , 1.)	(.99 , 1.)	(.98 , 1.)	(.97 , .99)
Daily temperature minimum	1.01***	1.01**	1.01**	1.01**	1.01**	1.02**	1	1	1.01	1.01***	1.01***	1.02***
	(1. , 1.02)	(1. , 1.02)	(1. , 1.03)	(1. , 1.02)	(1. , 1.02)	(1. , 1.03)	(1. , 1.01)	(1. , 1.01)	(.99 , 1.02)	(1. , 1.02)	(1. , 1.02)	(1. , 1.03)
Treatment options (reference :lobectomy)												
Limited resection with	0.66	0.47	0.40	0.6	0.45	0.39	0.9	0.52	0.38	0.67	0.49	0.39
adjuvant radiotherapy	(.17 , 2.52)	(.1 , 2.23)	(.08 , 1.98)	(.17 , 2.11)	(.11 , 1.85)	(.09 , 1.63)	(.22 , 3.68)	(.11 , 2.4)	(.08 , 1.73)	(.18 , 2.48)	(.11 , 2.15)	(.09 , 1.71)
Treatment interaction with air pollutant and weather elements												
NO2 * Treatment	1.02	1.02**	1.02**	1.01***	1.01***	1.01***						
	(1. , 1.04)	(1. , 1.03)	(1. , 1.03)	(1.01 , 1.02)	(1. , 1.02)	(1. , 1.02)						
SO2 * Treatment	0.99	1	1				1	1.01	1.01			
	(.97 , 1.02)	(.98 , 1.02)	(.98 , 1.02)				(.98 , 1.02)	(.99 , 1.03)	(.99 , 1.03)			
CO * Treatment	0.89	0.88	0.87							1.18***	1.15**	1.13**
	(.68 , 1.16)	(.67 , 1.15)	(.66 , 1.14)							(1.05 , 1.34)	(1.02 , 1.3)	(1. , 1.27)
Precipitation * Treatment	1	1	1	1	1	1	1	1	1	1	1	1
	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)
Snow * Treatment	1	1	1	1	1	1	1	1	1	1	1	1
	(.99 , 1.01)	(.99 , 1.01)	(.99 , 1.01)	(.99 , 1.01)	(.99 , 1.01)	(.99 , 1.01)	(.99 , 1.01)	(.99 , 1.01)	(.99 , 1.01)	(.99 , 1.01)	(.99 , 1.01)	(.99 , 1.01)
Temperature minimum* Treatment	1	1	1	1	1	1	1	1	1.01	1	1	1
	(.99 , 1.01)	(.99 , 1.01)	(.99 , 1.02)	(.99 , 1.01)	(.99 , 1.01)	(.99 , 1.02)	(.99 , 1.01)	(.99 , 1.01)	(.99 , 1.02)	(.99 , 1.01)	(.99 , 1.01)	(.99 , 1.02)

p values: * <0.1%, ** < 0.05%, *** < 0.01%

Appendix Table 4.7 Hazards of death three years after diagnosis for annual average of monthly maximum values

	Multipollutant			NO2			SO2			CO		
	Hazard of death 5 year after diagnosis			Hazard of death 5 year after diagnosis			Hazard of death 5 year after diagnosis			Hazard of death 5 year after diagnosis		
	Duration of exposure from diagnosis			Duration of exposure from diagnosis			Duration of exposure from diagnosis			Duration of exposure from diagnosis		
	1 yr bf	3 yrs bf	5 yrs bf	1 yr bf	3 yrs bf	5 yrs bf	1 yr bf	3 yrs bf	5 yrs bf	1 yr bf	3 yrs bf	5 yrs bf
Air pollutants and weather elements												
NO2	1.02***	1.02***	1.02***	1.04***	1.05***	1.05***						
	(1.01 , 1.04)	(1.01 , 1.04)	(1.01 , 1.04)	(1.03 , 1.5)	(1.04 , 1.72)	(1.04 , 2.35)						
SO2	1.04***	1.04***	1.04***				1.04***	1.04***	1.04***			
	(1.02 , 1.05)	(1.02 , 1.05)	(1.03 , 1.05)				(1.02 , 1.05)	(1.03 , 1.05)	(1.03 , 1.05)			
CO	1.55***	1.61***	1.64***							1.79***	1.97***	2.05***
	(1.34 , 1.79)	(1.36 , 1.9)	(1.36 , 1.98)							(1.58 , 2.04)	(1.72 , 2.24)	(1.8 , 2.33)
Precipitation	1***	1***	1***	1***	1***	1*	1.00	1.00	1.00	1.***	1.***	1.***
	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)
Snow	1.00	.99**	.99***	1.00	.99**	.99***	1.00	1.00	0.99	1.00	0.99	.99***
	(.99 , 1.)	(.99 , 1.)	(.98 , 1.)	(.99 , 1.)	(.99 , 1.)	(.98 , 1.)	(.99 , 1.01)	(.99 , 1.01)	(.99 , 1.)	(.99 , 1.)	(.99 , 1.)	(.98 , 1.)
Daily temperature minimum	1.01***	1.01***	1.02***	1.01***	1.01***	1.02***	1.00	1.00	1.01	1.01***	1.01***	1.02***
	(1. , 1.02)	(1.01 , 1.02)	(1.01 , 1.03)	(1. , 1.01)	(1. , 1.02)	(1.01 , 1.03)	(.99 , 1.)	(.99 , 1.01)	(1. , 1.02)	(1. , 1.02)	(1.01 , 1.02)	(1.01 , 1.03)
Treatment options (reference :lobectomy)												
Limited resection with	1.40	1.07	1.11	0.94	0.73	0.76	1.69	1.08	0.89	1.14	0.85	0.77
adjuvant radiotherapy	(.5 , 3.88)	(.35 , 3.26)	(.35 , 3.51)	(.36 , 2.43)	(.27 , 2.)	(.28 , 2.1)	(.57 , 5.04)	(.35 , 3.31)	(.29 , 2.75)	(.42 , 3.1)	(.3 , 2.38)	(.28 , 2.16)
Treatment interaction with air pollutant and weather elements												
NO2 * Treatment	1.02	1.02***	1.02***	1.01***	1.01***	1.01***						
	(1.01 , 1.04)	(1.01 , 1.04)	(1.01 , 1.03)	(1.01 , 1.02)	(1.01 , 1.02)	(1.01 , 1.02)						
SO2 * Treatment	0.99	0.99	0.99				1	1	1			
	(.97 , 1.)	(.97 , 1.01)	(.97 , 1.01)				(.98 , 1.01)	(.98 , 1.02)	(.99 , 1.02)			
CO * Treatment	0.85	0.85	0.86							1.16**	1.16**	1.15***
	(.68 , 1.07)	(.69 , 1.06)	(.69 , 1.06)							(1.03 , 1.32)	(1.03 , 1.3)	(1.03 , 1.28)
Precipitation * Treatmer	1	1	1	1	1	1	1	1	1	1	1	1
	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)	(1 , 1)
Snow * Treatment	1	1	1	1	1	1	1	1	1	1	1	1
	(.99 , 1.)	(.99 , 1.)	(.99 , 1.)	(.99 , 1.)	(.99 , 1.01)	(.99 , 1.01)	(.99 , 1.)	(.99 , 1.)	(.99 , 1.)	(.99 , 1.01)	(.99 , 1.01)	(.99 , 1.01)
Temperature minimum * Treatment	.99*	0.99	0.99	1	1	1	0.99	1	1	1	1	1
	(.98 , 1.)	(.99 , 1.)	(.98 , 1.)	(.99 , 1.)	(.99 , 1.01)	(.99 , 1.01)	(.99 , 1.)	(.99 , 1.01)	(.99 , 1.01)	(.99 , 1.)	(.99 , 1.01)	(.99 , 1.01)

p values: * <0.1%, ** < 0.05%, *** < 0.01%

Appendix Table 4.8 Hazards of death five years after diagnosis for annual average of monthly maximum values

CURRICULUM VITAE
Naiya Patel, PhD, MPH, DDS

PROFESSIONAL EXPERIENCE

Department of Health Management and Systems Sciences, SPHIS

Teaching Assistant

July 2023 *July 2022-*

Two Ph.D. courses assisted are:

- Health Services Research Methods 2
- Economic Evaluation in Health Care

- Prepare drafts of each class slides and class material under the guidance of the primary course instructor
- Assist in developing syllabus, and class rubrics
- Assist in effective and efficient course assignments
- Grade assignments and help solve student queries
- Play active role in teaching and advising process for the courses

The Commonwealth Institute of Kentucky

Research Scholar *August 2020- July 2022*

- Perform data analysis and data investigation for real-world data
- Assist in key project decision making i.e. ICD coding, appropriate data analysis technique, relevant study cohort grouping, assigning work to team members
- Crosswalk ICD 9 to 10 for Medicaid Claims data
- Develop data frames and structure raw claims data utilizing Knime, SQL, Pycharm and STATA
- Investigate big data problems and propose solutions to overcome it
- Work with raw environmental data for the U.S. and clean it to merge Air Pollutants, Weather data, and County demographics
- Perform literature search and data mining relevant to healthcare cost data and Medicaid Claims data
- Draft reports and presentations
- **Project completed** : The Effect of Market Changes on Kentucky Medicaid: Utilization and Cost Savings for Three Chronic Diseases—Cancer, Diabetes, and CVD

Louisville Metro Government, Kentucky United States

Project Intern

January 2022 *October 2021-*

- Prepare and draft literature review surrounding COVID-19 variants for Louisville Metro Department of Public Health
- Draft COVID19 state/county projection reports for publication and release
- Inform team manager about epidemiologic model parameters value from the literature reviews to support model simulation
- Train undergraduates and graduate interns in the department on systematic literature reviewing process, teach novel reviewing techniques that are time efficient and review their work

Center for Health Organization Transformation - NSF funded

Research Assistant *August 2019-2020*

- Engage in literature review and literature search using a database of interest
- Prepare annotated bibliographies and assist a faculty member with a research project
- Assist in all stages of survey data collection, analysis, project decision making, and reporting
- Worked with different survey instruments and large data-sets using a variety of analytical techniques to prepare research reports
- Administrative work includes facilitating weekly meetings, prepare a timeline for a research project, design and customize gift cards for incentives, preparing email drafts for marketing strategy
- Prepare manuscripts of an ongoing research study using citation tools
- Develop a research poster when required, participate in conferences
- Strategic planning and stakeholder engagement

Jersey Smile Care, Jersey City, New Jersey *April 2018- January 2019*

Oral Health Educator and Dental Assisting,

- Assist the dentist during and after the dental procedure for educating the patient
- Follow up with the patient regarding recall visits, making new appointments and ad-hoc appointments
- Contact insurance companies to verify the eligibility of the patient, enter the details in Dentrix and maintain a patient record in Dentrix as well as patient chart

- Act as a liaison between dentist, patient, and insurance company and navigate them as required. Well versed with medical and dental coding, terminologies for filing insurance claims and eligibilities
- Maintain patient chart and assist the dentist in documenting the updated information at every visit
- Perform instrument sterilization, setup material and instrument trays for the patient appointment/treatment procedures
- Keep the Dentrix application updated including entering and rescheduling patient appointments, patient information, and patient insurance information, treatment procedures, treatment costs and payments and answering patient queries over the phone

Bristol-Myers Squibb, Princeton, New Jersey *February 2017- May 2017*

Feasibility Specialist Intern, Clinical Trials, and Business Analytics.

- Data-mined completed oncology studies to retrospectively assess screen-failure reasons and avoidable amendments in clinical trial studies
- Analyzed the study metrics of interest for specific clinical trial phase study
- Utilized pressure testing as well as building and running queries in CE works
- By comparing the data with real-world data outcomes, could predict values and generate insights
- Performed study protocol reviewing and developed framework
- Well versed with real-world database like CE works, TA scan, Globocan as well as Health landscape and GIS system database
- Hands-on experience in Clinical Trial Management (CTM) system. Knowledge about MEDIDATA analysis tool required for analysis of new clinical trial site inflow
- Developed protocol inclusion-exclusion criteria after screening real-world patient scenario
- Experience interpreting Electronic Medical Record

Long Island University, Brooklyn, New York

Graduate Research Assistant, Department of Public Health, October 2016-January 2019.

- Conducted research studies under the guidance of Head of the Department; in the process of data analysis using SPSS and manuscript review. The research studies are IRB approved
- Engage in literature review and literature search using a database of interest

- Prepare annotated bibliographies and assist a faculty member with a research project
- Participate in all stages of survey data collection, analysis, and reporting
- Worked with different survey instruments and large data-sets using a variety of analytical techniques to prepare research reports
- Administrative work includes facilitating weekly meetings, prepare a timeline for a research project, design and customize gift cards for incentives, preparing email drafts for marketing strategy
- Prepare manuscripts of an ongoing research study using citation tools

New York University Tobacco (SocioEconomic Evaluation of Dietary Decisions) SEED Lab, **New York, New York** *September 2015 - September 2016*

Research Assistant,

- Assisted in roles related to research, administration, and communications, which also included preparation of materials for conferences. Currently, the research study manuscript titled "Tobacco and Dental Education: Dental Student Use, Knowledge, and Beliefs about Tobacco Products" has been accepted by " Journal of Dental Education" and is in the process of publication
- Attend a weekly meeting to discuss the status of the ongoing research study
- Engage in literature review using a database of interest and prepare a draft for manuscript
- Rectify as well as test the survey questionnaire (Pre-test) and make necessary changes before distributing it to the study sample
- Engage in analysis of the results and generate possible insights

Aastha Dental Clinic, **Vadodara, India**

General Dentist, *November 2014 - June 2015*

- Performed various dental procedures, such as dental fillings, tooth extractions, denture placement, and patient counseling
- Assist in major surgeries like tooth implant and multiple extractions under general anesthesia as well as bone grafting
- Pre and Post counseling of relevant patients as well as filling their case report forms
- Verify the eligibility, measure vital signs and keep their medical records in files for future record
- Follow up with treated patients

Indian Red Cross Society, **Vadodara, India.**

Volunteer/Health Educator and General dentist, *May 2014 - May 2015*

- Prepared informative brochures for upcoming campaigns
- Prepare powerpoint presentations in local layman dialect and show the presentations in every campaign to facilitate the education campaigns for health
- Volunteered blood donation camps
- Demonstrated correct brushing techniques with the help of models to facilitate public understanding in a health education campaign
- Empower and educate people regarding wrong notions of oral hygiene, correct nutrition, and blood donation myths

EDUCATION

The University of Louisville, School of Public Health and Information Science,
Doctor of Philosophy- Ph.D. in Health Policy June 2023
Thesis topic : Clinical and Environmental Factors Affecting the Survival Outcomes among Stage IA T1-2N0M0 First Primary Non-Small Cell Lung Carcinoma Patients in the United States
Utilized database: Geo code U.S. EPA air pollution, NOAA's Global Historical Climatology Network (GHCN)-daily data for U.S., Historical Area Resource Files, and SEER-18 plus Cancer Registry data. The thesis determines impact of weather and air pollution on treated lung cancer patients as well as factors determining treatment receipt for such patients. Study period is 1988-2016.

Long Island University, Brooklyn, New York
Masters of Public Health, December 2017

Global Institute of Public Health, NEW YORK UNIVERSITY, New York, New York
Masters of Public Health (Epidemiology), September 2015 – December 2015

Karnavati School of Dentistry (KSD), Ahmedabad, India
Bachelor of Dental Surgery, October 2014
GPA: 3.5, Stood among the top 20 students out of 1000 at the state level.

CERTIFICATES

- **NIH Training** course "Protecting Human Research Participants," September 2015.
- **CITI Training (GCP)** for Good Clinical Practice, the NYU Medical School, November 2015.
- **Indian Red Cross Society**, Vadodara, India, obtained FIRST-AID certificate for all types of first aid training for emergency lifesaving situations, May 2014.

- **FEMA(Federal Emergency Management Agency)** required for emergency preparedness, January 2016.
- **Health Insurance Portability and Accountability Act (HIPAA)** patient protection, January 2025

AWARDS:

- Graduate Scholars Award at Long Island University, awarded to those who have achieved an outstanding GPA
- Academic Excellence Award for MPH at Long Island University
- Multicultural Association of Graduate Students Research Award (MAGS) for being selected as top 5 student research abstract worldwide to present research poster presentation at ISPOR Europe 2020, at The University of Louisville
- Women in Public Health travel award for being selected at AcademyHealth 2021 research conference to present the selected research abstract

PUBLICATIONS

Research Papers

- Patel, N., & Patel, N. (2017). Modern technology and its use as storytelling communication strategy in public health. MOJ Public Health, 6(3), 00171.
- Patel N. Asthma and Its Environmental Triggers—A Systematic Comparative Review Study. Global Scientific Research Journal of Public Health, 1(1), 2018, pp. 1-5
- Shearston, J. A., Shah, K., Cheng, E., Moosvi, R., Park, S. H., Patel, N., ... & Weitzman, M. Tobacco and Dental Education: Dental Student Use, Knowledge, and Beliefs about Tobacco Products.
- Patel, N. (2018). Bridging the gap of translation research in public health-from research to real world. MOJ Public Health, 7(6), 347-349.
- Patel N (2018) Cellphone Radiations and its Effects in Public Health- Comparative Review Study. MOJ Public Health 7(1): 00197. DOI: 10.15406/mojph.2018.07.00197
- Berić-Stojšić B, Patel N, Blake J, Johnson D. Flipped Classroom Teaching and Learning Pedagogy in the Program Planning, Implementation, and Evaluation Graduate Course: Students' Experiences. Pedagogy in Health Promotion. 2019 Apr 9:2373379919839073.
- Patel N, Patel N. What Does a College Attending Men's Community Perceive about Transgender Women?-A Non-Experimental Qualitative Study.

- Naiya Patel (2019) Why New Drugs, Treatments, and Medical Devices Still Needs to be Tested Clinically Before Making it Available in the Market?. Journal of Neurological Research And Therapy - 3(1):1-5.
- BHAGAT, Pramod Kumar; PATEL, Naiya; PATEL, Neel. Effects of denture cleansers on denture material properties —An observational Case—Control Study.. Journal of Oral Research, [S.l.], v. 9, n. 1, p. 72-80, feb. 2020. ISSN 0719-2479. Available at: <http://www.joralres.com/index.php/JOR/article/view/joralres.2020.011>.
- Patel N. Impact on Dental Economics and Dental Healthcare Utilization in COVID-19: An Exploratory Study. Journal of Advanced Oral Research. August 2020. doi:10.1177/2320206820941365
- Karimi, Seyed; DuPre, Natalie; McKinney, W. Paul; Little, Bert B.; Patel, Naiya; and Moyer, Sarah (2020) "Projecting the COVID-19 Weekly Deaths and Hospitalizations for Jefferson County, Kentucky," The University of Louisville Journal of Respiratory Infections: Vol. 4 : Iss. 1 , Article 44.
- Patel,N; Patel,N (2021) ": Expansion of Preventive Dental Service Coverage for Certain Medicaid Beneficiaries in Texas- Call for Dental Policy Effectiveness Action", Journal of American Dental Association (under review).
- Patel, N. (2020) Qualitative Research Methodology and its Scope in Health Services Research. Journal of Neurological Research And Therapy - 3(2):18-21.
- Patel, N. Building a culture of health in Kentucky to address racism a public health crisis. MOJ Public Health. 2022;11(3):75–81.DOI: 10.15406/mojph.2022.11.00379
- Patel, N. (2021) COVID-19 Vaccine Production and Potential Market Characteristics for Pharmaceutical Companies to Enter the Market-Perfect Competition, Profit Maximization, and Externalities. DOI: 10.46998/IJCMCR.2021.14.000350
- Patel, N. (2021). Understanding Barriers and Facilitators for Telehealth Implementation in Healthcare Delivery System During COVID-19-Call for Action. Biomedical Journal of Scientific & Technical Research, 38(2), 30125-30132.

Technical Reports

- The published reports on COVID-19 informing Kentucky state governor and health policy decision makers can be found at https://louisville.edu/sphis/research/covid-19-projections
- The published health policy brief titled "The Impact of Medicaid Managed Care on Non-Office-Based Service Utilization Among Kentucky Medicaid Beneficiaries

with Chronic Disease" informing both private and public stakeholders can be found at https://louisville.edu/sphis/departments/cik/policy-briefs

- The published data brief titled "Area Health Resource Files (AHRF) Data" on Area Health Resource Files can be found at https://louisville.edu/sphis/departments/cik/data-briefs
- Health Policy report titled "How Useful Are Digital Health Terms for Outcomes Research? An ISPOR Special Interest Group Report" published in Value in Health journal can be found at https://doi.org/10.1016/j.jval.2022.04.1730

Book

- B Suresh Lal and Naiya Patel (2020). Economics of Covid-19 Digital Health Education and Psychology, Adhyayan Publishers New Delhi, India ISBN-10 : 9388804775. *Foreword provided by highest Indian civilian award winner Prof. Sukhadeo Thorat (Padma Shri).*

Working Papers

- Factors affecting treatment receipt and survival outcomes for stage 1A Non Small Cell Lung Cancer utilizing evidence synthesis for constructing directed acyclic graphs
- Disparity factors that affect treatment receipt for stage 1A Non Small Cell Lung Cancer patients in the United States
- Air pollution and weather affects survival outcomes among stage 1A Non Small Cell Lung Cancer patients in the United States

Media Highlights

- Telangana Today press article on our COVID-19 book released by Padma Bhushan (highest Indian civilian award) achiever Dr.Krishna Ella in Hyderabad and foreword provided by Padma Shri (highest Indian civilian award) achiever Dr.Sukhadeo Thorat
- Long Island University Headlines on contribution towards international book on COVID-19
- COVID-19 projection reports for Kentucky State- Louisville Business first
- COVID19 death projection report- 89.3 WFPL
- AcademyHealth Blogpost on dentalcare cost reduction and better outcomes
- University of Louisville interview on my COVID-19 published book titled "Economics of COVID-19: Digital Health, Education and Psychology"

Invited Talks

- Student Research Spotlight ISPOR Europe 2020 on "Projecting the COVID-19 Weekly Deaths and Hospitalizations for Jefferson County Kentucky" under conference theme Advancing Evidence to action
- Fireside Chat ISPOR student chapter at University of Washington on my research interests and about ISPOR 2020 student research spotlight https://soundcloud.com/isporfiresidechat/episode-3-naiya-patel-samuel-crawfords-research-spotlight
- ISPOR 2021 Europe conference podium presentation on "Projecting COVID-19 Hospitalizations and Deaths Under Scenarios of Vaccination in Jefferson County, Kentucky"

SERVICE

Reviewer of Journals- Reviewer Record

- MedCrave Journal of Public Health
- Reviewer board member of SOJ Nursing and Healthcare journal
- Pedagogy in Health Promotion by SAGE publications
- Reviewer BMJ open Journal
- Journal of Advanced Oral Research SAGE Publications
- Journal of Health Research and Reviews Wolters Kluwer
- Reviewer Board Panel SAGE

Editor

- Indonesian Journal of Health Administration
- American Journal of Epidemiology and Public Health
- Journal of International Oral Health
- Journal of Environment and Life Science

Faculty Search Committee University of Louisville 2022-2023

- Assist department in the selection and decision-making process for potential faculty candidates

Judge

- 2022 Undergraduate Arts & Research Showcase University of Louisville

POSTER PRESENTATIONS

- **Bristol-Myers Squibb Symposium 2017-** Use of Real-world data in improving clinical trial design
- **ADEA conference California 2017 –** Tobacco and Dental Education: Dental Student Use, Knowledge, and Beliefs about Tobacco Products
- **Long Island University 2018 –** Discovery Day "Student's perceptions about the effectiveness of flipped classroom in public health graduate course"
- **Society for Public Health Education 2018 annual conference abstract submission-**

Student's perceptions about the effectiveness of flipped classroom in public health graduate course

- **CHOT Seattle 2019 -** Diabetes medication adherence among Medicaid enrollee by observing copayment and no copayment- report
- **ISPOR Europe 2020-** Projecting COVID 19 Weekly Deaths and Hospitalizations for Jefferson County Kentucky- research paper abstract
- **KPHA 2021-** Projecting COVID-19 Hospitalizations and Deaths Under Scenarios of Vaccination in Jefferson County, Kentucky
- **AcademyHealth 2021**- Projecting COVID-19 Hospitalizations and Deaths Under Scenarios of Vaccination in Jefferson County, Kentucky