UNIVERSITY OF CALIFORNIA SAN DIEGO

SAN DIEGO STATE UNIVERSITY

Methodological Considerations for Studying Neighborhood Contextual Determinants of Lung Cancer Survival

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy

in

Public Health (Epidemiology)

by

Paige Sheridan

Committee in charge:

University of California San Diego

Professor Tarik Benmarhnia, Chair
Professor Kimberly Brouwer
Professor James Murphy

San Diego State University

Professor Eyal Oren
Professor Humberto Parada
Professor Caroline Thompson

2021

The Dissertation of Paige Sheridan is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California San Diego

San Diego State University

2021

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to acknowledge the support of my advisor and dissertation chair, Tarik. Over the past five years, Tarik has supported, encouraged, and challenged me. I am grateful for his mentorship.

For all of their patience, support, and love throughout this journey, I would like to acknowledge my parents, Libby and Barnaby, and my partner, George. I would not be where I am today without them.

Chapter 2, in part, is currently being prepared for submission for publication of the material. Sheridan, Paige; Caroline, Thompson; Benmarhnia, Tarik. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, has been submitted for publication of the material as it may appear in American Journal of Epidemiology. Sheridan, Paige; Chen, Chen; Thompson, Caroline; Benmarhnia, Tarik. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in full, has been submitted for publication of the material as it may appear in Cancer Epidemiology, Biomarkers and Prevention. Sheridan, Paige; Thompson, Caroline; Benmarhnia, Tarik. The dissertation author was the primary investigator and author of this paper.

Vita

Education

2021 Doctor of Philosophy, University of California San Diego & San Diego State University

2014 Master of Public Health, University of California Los Angeles

2012 Bachelor of Science, University of Oregon

Publications

Sherer MV, Deka R, Salans MA, Nelson TJ, **Sheridan P**, Rose BS. Androgen deprivation therapy and acute kidney injury in patients with prostate cancer undergoing definitive radiotherapy. *Prostate Cancer and Prostatic Diseases*. 2021

Brennan M, Lowenstern A, **Sheridan P**, Boero I, Thourani VH, Vemulapalli S, Wang, TY, Liska O, Gander S, Jager J, Leon M, Peterson ED. Association between Patient Survival and Clinician Variability in Treatment Rates for Aortic Valve Stenosis. *Journal of the American Heart Association.* 2021

Lowenstern A, **Sheridan P**, Wang TY, Boero I, Vemulapalli S, Thourani VH, Leon MB, Peterson ED, Brennan M. Sex Disparities in Patients with Symptomatic Severe Aortic Stenosis. *American Heart Journal.* 2021

Sheridan P, LeBrett W, Triplett D, Roeland E, Bruggeman A, Yeung H, Murphy J. Economic impact of palliative care among patients with advanced stage cancer. *American Journal of Hospice and Palliative Medicine*. 2021.

Kumar A, Guss Z, Courtney P, Nalawade V, **Sheridan P**, Sarkar R., Banegas MP, Rose BS, Xu R, Murphy J. Evaluation of the use of cancer registry data for comparative effectiveness research. *JAMA Network Open*. 2020

Sarkar R, Courtney PT, Bachand K, **Sheridan P**, Riviere P, Guss ZD, Lopez CR, Brandel MG, Banegas MP, Murphy JD. Quality of care at safety-net hospitals and the impact on pay-for-performance reimbursement. *Cancer*. 2020

Sheridan P, Trinidad D, McMenamin S, Pierce J, Benmarhnia T. Evaluating the Impact of the California Workplace Smoking Ban in 1995, a Synthetic Control Approach. *Preventive Medicine Reports*. 2020

Vitzthum L, Heidi E, Park H, Williamson C, **Sheridan P**, Huynh-Le MP, Sirak I, Wei L, Tarnawski R, Mahantshetty UM, Nguyen C, Mayadev J, Yashar CM, Sacco AG, Mell LK. Comparison of hematologic toxicity and bone marrow compensatory response in head and neck versus cervical cancer patients undergoing chemoradiotherapy. *Frontiers in Oncology*. 2020

Brennan MJ, Leon M, **Sheridan P**, Boero I, Chen Q, Lowenstern A, Thourani V, Vemulapalli S, Thomas K, Wang T, Peterson E. Racial differences in the use of aortic valve replacement for treatment of symptomatic, severe aortic valve stenosis in the transcatheter aortic valve replacement era. *Journal of the American Heart Association*. 2020

Aguilera R, Hansen K, Gershunov A, Ilango SD, **Sheridan P**, Benmarhnia T. Respiratory hospitalizations and wildfire smoke: a spatio-temporal analysis of the October 2007 firestorm in San Diego County. *Environmental Epidemiology*. 2020

Schwarz L, Malig B, Guzman-Morales J, Guirguis K, Ilango S, **Sheridan P**, Gershunov A, Basu R, Benmarhnia T. The health burden of fall, winter and spring extreme heat events in Southern California and contribution of Santa Ana Winds. *Environmental Research Letters*. 2020.

Sun Y, **Sheridan P**, Laurent O, Li J, Sacks D, Fischer H, Qiu Y, Jiang Y, Yim IS, Jiang LH, Molitor J, Chen JC, Benmarhnia T, Lawrence J, Wu J. Associations between green space and preterm birth: Windows of susceptibility and interaction with air pollution. *Environment International*. 2020.

Ilango SD, Weaver M, **Sheridan P**, Schawrz L, Clemesha RS, Bruckner T, Basu R, Gershunov A, Benmarhnia T. Extreme heat episodes and risk of preterm birth in California, 2005-2013. *Environment International.* 2020

Vitzthum L, Riviere P, **Sheridan P**, Nalawade V, Deka R, Xu R, Mell L, Wallace M, Murphy J. Predicting Persistent Opioid Use, Abuse and Toxicity Among Cancer Survivors. *JNCI.* 2019

Schwarz L, Bruckner T, Basu R, **Sheridan P**, Ilango S, Benmarhnia T. A quantile regression approach to examine fine particles, term low birth weight, and racial/ethnic disparities. *Environmental Epidemiology*. 2019.

Sheridan P, Mair CA, & Quinones AR. Associations between prevalent multimorbidity combinations and prospective disability and self-rated health among older adults in Europe. *BMC Geriatrics.* 2019.

Sheridan P, Ilango S, Bruckner T, Wang Q, Basu R, Benmarhnia T. Air pollution and preterm birth in California: identification of critical exposure windows. *American Journal of Epidemiology*. 2019.

Wang Q, Benmarhnia T, Zhang H, Knibbs LD, **Sheridan P**, Li C, et al. Identifying windows of susceptibility for maternal exposure to ambient air pollution and preterm birth. *Environment International*. 2018.

Quinones AR, Nagel C, Newsom J, Huguet N, **Sheridan P**, Thielke S. Racial and ethnic differences in smoking changes after chronic disease diagnosis among middle-aged and older adults in the United States. *BMC Geriatrics.* 2017.

Abstract of the Dissertation

Methodological Considerations for Studying Neighborhood Contextual Determinants of Lung Cancer Survival

by

Paige Sheridan

Doctor of Philosophy in Public Health (Epidemiology)

University of California San Diego, 2021

San Diego State University, 2021

Professor Tarik Benmarhnia, Chair

In the United States, lung cancer is the second most common cancer in men and women and is the leading cause of cancer death. Lung cancer survival is lower than other leading cancers, and approximately half of people with lung cancer die within one year of being diagnosed. While small advances in treatment and interventions in tobacco cessation have improved survival in

recent years, persistent poor survival suggests that new approaches are needed to identify contextual risk factors to improve lung cancer survival.

It has become widely accepted that where an individual lives or works is an important determinant of health. Place-based exposures such as neighborhood contextual factors may be particularly valuable for intervention in lung cancer survival as they can be intervened on using population-level interventions such as local or state policies. Recently, studies have observed associations between two potential neighborhood contextual factors and lung cancer survival: ambient air pollution and racial residential segregation. However, neighborhood-level exposures require specific methodological considerations because they have spatial and temporal components that need to be appropriately considered in analyses.

The first chapter of this dissertation reviews the epidemiologic evidence regarding spatial heterogeneity in lung cancer survival and the relationship between both air pollution and racial residential segregation on lung cancer survival. The second chapter evaluates how lung cancer survival varies spatially in California and how these spatial patterns may change over time. The third chapter examines the association between air pollution and lung cancer survival and illustrates how a systematic error known as immortal time bias can be introduced when a time-varying exposure such as air pollution is mishandled in the context of a time-to-event outcome. The fourth chapter of this dissertation describes two considerations in the measurement of segregation, the spatial scale of geographies and the spatial relationships of populations. This chapter illustrates how these considerations impact both the absolute measurement of segregation and subsequent conclusions about the impact on lung cancer survival. The final chapter of this dissertation summarizes key findings and highlights future directions to advance the study of neighborhood contextual factors in lung cancer survival.

1 Introduction

1.1 Overview of lung cancer

Lung cancer is a heterogeneous group of diseases characterized by uncontrolled growth of abnormal cells in the lungs. In the United States, lung cancer is the second most common cancer in men and women and is the leading cause of cancer death. By the end of 2021, lung cancer will account for approximately 12% of new cancer cases and 22% of cancer deaths, more deaths than colon, breast, and prostate cancer combined (1). Primary lung cancer can be classified into two histologic groups: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). NSCLC can be further categorized into adenocarcinoma, squamous cell carcinoma, and large cell carcinoma histologic subtypes. NSCLC accounts for approximately 85% of all lung cancers, while SCLC accounts for only 15%. NSCLC and SCLC (and subtypes) are distinct clinicopathological entities which has implications for etiology, prevention, screening, diagnosis, and treatment.

To characterize the extent of disease in NSCLC, clinicians stage the cancer using a TNM staging system. This system combines features of the tumor into stage groups to describe the extent of the cancer that can be easily communicated to others, assist in treatment decisions, and serve as an indicator of prognosis. **Table 1.1** shows a simplified overview of this staging process. The International Association for the Study of Lung Cancer publishes a complete description of TNM lung cancer staging (2). The staging process for SCLC is slightly different from NSCLC, where the cancer is generally classified as limited or extensive, with limited stage characterized by cancer confined to the primary tumor and regional lymph nodes. Cancer stage at diagnosis has substantial implications for etiology, treatment considerations, survival, and quality of life. Because of this, epidemiologic studies and cancer statistics often present stratified analyses by stage at diagnosis.

Table 1.1 NSCLC stage at diagnosis

Stage	Description
Stage 1 (Localized)	Cancer is only in the lung, has not spread to any lymph nodes
Stage 2 (Regional)	Cancer is in the lung and nearby lymph nodes
Stage 3 (Regional)	Cancer is in the lung and lymph nodes in the center of the chest
Stage 4 (Distant)	Cancer is found in the lungs, fluid in the area around the lungs, and/or another organ in the body

Lung cancer incidence has declined in men since the mid-1980s and in women since the mid-2000s (1). This sex difference is primarily due to differences in historical patterns of smoking uptake and cessation (1). Similarly, lung cancer mortality has declined by about 54% since 1990 in men and 30% since 2002 in women, largely due to reductions in smoking prevalence. This pace has accelerated in recent years; in the last six years, the rate has increased by 4% and 5% in women and men, respectively.

Lung cancer incidence and mortality differ by geographic and demographic characteristics, which may be explained by smoking behaviors, population demographics, environmental exposures, and cancer screening prevalence. Overall, lung cancer is more common in men than women, Black patients than white, rural than urban populations (3), and among those with lower socioeconomic status (4). Age and cigarette smoking are two of the strongest risk factors for developing lung cancer. On average, a current smoker has approximately 20 times the risk of lung cancer than nonsmokers. It is estimated that 80% of all lung cancers can be attributed to cigarette smoking (5). Other proposed risk factors include exposure to secondhand smoke, occupational exposure to asbestos, arsenic, chromium, nickel, radiation exposure, family history of lung cancer, HIV infection, beta carotene supplements in heavy smokers, and air pollution exposure. A recent global analysis ranked air pollution exposure and smoking as leading risk factors for respiratory tract cancers (6).

The primary treatments for lung cancer include surgery, chemotherapy, radiotherapy, immunotherapy, and targeted therapies. Lung cancer treatment depends on stage of disease, tumor size, metastasis, and histology of the cancer. The treatment goal in early-stage lung cancer is curative, while it is to prolong survival and improve quality of life in late-stage disease. Surgical resection is the most consistent and successful option for early-stage lung cancer (7). Early-stage cancer can be additionally treated with chemotherapy with or without radiation for NSCLC and chemoradiation for SCLC. Advanced stage lung cancer is treated primarily with systemic therapies to relieve symptoms, improve quality of life, and increase survival time.

1.2 Overview of lung cancer survival

In the US, lung cancer survival is lower than other leading cancers, and more than half of people with lung cancer die within one year of being diagnosed. While the five-year survival rate for lung cancer is 59% for cases detected at localized stage, only 17% of cases are detected this early (1). For tumors diagnosed at distant stages, the five-year survival rate is only 5%. Over the past several decades, the five-year overall survival rate has increased from around 10% in 1973 to 20% in 2017 (8). Survival is increasing steadily across all racial and ethnic groups, and all age groups have seen an increase in survival over time, particularly among patients 15-44 years old.

Despite this overall improvement in survival, substantial sociodemographic and clinical disparities remain. A higher proportion of female patients are diagnosed with adenocarcinoma and at early-stage of disease than male patients. Female patients are more likely to be younger, never smokers, have better performance status, and have better five-year survival than male patients (9-12) (13). Younger patients have better survival than older patients, and Black patients have consistently lower survival than all other racial/ethnic groups (14-16).

Stage at cancer diagnosis is one of the strongest prognostic factors in lung cancer survival. Other important prognostic factors include lifestyle factors, socioeconomic factors, access to care and receipt of guideline-concordant treatment, immunotherapy, comorbid conditions, tumor cell grade differentiation, smoking cessation, tumor molecular markers, drug metabolism, and DNA repair (cells with reduced DNA repair have a higher sensitivity to treatment) (14, 15). Survival varies between patients with the same stage at diagnosis and similar treatment modalities, demonstrating that other factors may influence survival. The factors that improve survival in long-term survivors (alive at five years after diagnosis) are largely unknown (16, 17). Previous studies have proposed that younger age and treatment duration may be potential prognostic factors (18).

1.3 The emerging role of contextual determinants in lung cancer survival

Advances in lung cancer survival in recent years are primarily due to improvements in treatment (13). Despite the effort to increase early-stage diagnosis, interventions through screening have been largely unsuccessful. Low dose computed tomography screening is recommended for heavy current and former smokers (30 pack-years) between 55 and 80 years. However, only 4-5% of eligible adults are participating in screening (19). Another focused effort has been on smoking cessation based on evidence that continued smoking by lung cancer patients is associated with less effective treatment and a poorer prognosis (20). Current evidence suggests that approximately 50-60% of lung cancer patients are current smokers at the time of diagnosis and anywhere from 13 – 30% continue to smoke after diagnosis (21-23). Despite efforts to improve lung cancer survival through individual behavioral interventions (early detection, smoking cessation) and treatment advances, only small improvements in lung cancer survival have been made (24). Other than these interventions, most predictors of cancer survival are non-modifiable (tumor cell grade of differentiation, tumor molecular markers, drug metabolism, cellular DNA repair), resulting in

small survival gains over the past few decades. For this reason, contextual determinants of health have been targeted to identify novel risk factors to improve lung cancer survival.

It has become widely accepted that the place where an individual lives or works is an important determinant of health (25). For example, factors like distance to medical facilities, access to green space, and availability of fresh food are all associated with adverse health outcomes (26-28). Similarly, residents near traffic corridors are more likely to experience adverse respiratory outcomes (29). Recently, spatial epidemiology has emerged as an important approach to document the role of the geographic environment in health outcomes and disparities. Identifying spatial heterogeneity in health outcomes can inform risk factor identification, local health interventions, and resource allocation decisions. Further, neighborhood contextual factors may be valuable for intervention in lung cancer survival as they do not rely on individual behavior modification or intervention and can be used to create targeted interventions through local health policies. In order to identify these potential place-based risk factors, it is necessary to evaluate how lung cancer survival varies across space and time. Using these patterns, it is then possible to identify risk factors that vary concomitantly.

Previous studies have demonstrated that there is large-scale geographic variation in lung cancer survival. Richards et al. analyzed net survival by state and found sizeable geographic variation in lung cancer survival from 2001-2009. In the 37 states with comprehensive cancer registries, five-year survival ranged from 15% to 25% among whites and 7% to 22.7% among Black patients (30). These established geographic disparities in lung cancer survival indicate that there may be similar small-scale heterogeneity in survival that can be explained by neighborhood-level factors that vary across space and time. However, few studies have focused on small-scale variation to identify neighborhood variation driven by spatial or temporal patterns in risk factors.

Therefore, this dissertation will describe spatial heterogeneity in lung cancer survival in California and identify clusters of exceptionally high or low survival. This will be the focus of the first aim.

The presence of spatio-temporal variability in lung cancer survival indicates that there may be important modifiable contextual determinants that may contribute to this variability. Recently, two neighborhood-level factors have been targeted as potential modifiable factors: ambient fine particulate matter air pollution and racial residential segregation.

1.4 Air pollution and lung cancer survival

Ambient air pollution is a complex mixture of particulate matter (PM), gaseous pollutants (e.g., nitrogen dioxide, ozone), organic compounds, and heavy metals. Exposure to air pollution is the fifth leading cause of death globally and is responsible for approximately 3.3 million premature deaths worldwide, primarily attributed to $PM_{2.5}$ (31, 32). It is hypothesized that PM is the component of air pollution primarily responsible for adverse health effects due to its broad range of toxic substances. Major components of PM include transition metals, ions, organic compounds, minerals, reactive gases, and biologic materials. PM is classified by aerodynamic diameter: coarse (< 10 μm, PM_{10}), fine (<2.5 μm, $PM_{2.5}$), and ultrafine (< 0.1 μm, $PM_{0.1}$). The health effects vary by particle size; PM_{10} is primarily deposited along the upper airways and can be cleared through mucociliary clearance, while $PM_{2.5}$ and $PM_{0.1}$ are small enough to penetrate the lungs gas-exchange region where they can then be absorbed by lung endothelial cells and transported into systemic circulation, causing systemic inflammation and cell damage (33-36).

Air pollution is ubiquitous in the US, with important inequalities in exposure across the US and within cities. This means that all populations are exposed, and vulnerable populations are disproportionality exposed (37). Because of its established relationship with adverse health, $PM_{2.5}$

is systematically monitored and regulated in the US. While air pollution was decreasing in the US over the last two decades, there has been a sudden increase in the last two years, partially attributable to climate change, economic activity, increased number of wildfires and extreme weather events, and decreases in clean air enforcement actions (38). As air pollution exposure increases, adverse health impacts attributable to air pollution are expected to increase concomitantly. Therefore, understanding the adverse health impacts of air pollution exposure, particularly among vulnerable populations, is increasingly important.

Epidemiologic studies and animal models have established that $PM_{2.5}$ exposure is responsible for adverse health effects, mainly through the cardiovascular and respiratory systems. PM comprises a complex mixture of carcinogenic and mutagenic substances that play a role in chronic systemic inflammation, oxidative stress, and DNA damage in human tissues, resulting in adverse health outcomes (39-42). These adverse health effects can be seen after short and long-term exposure to $PM_{2.5}$, even at very low exposure levels (43-45). Studies have shown an exposure-response relationship between $PM_{2.5}$ and mortality risk, where adverse health effects are dependent on both exposure concentration and length of exposure (46).

The association between air pollution and lung cancer incidence is well established (47-50). It is hypothesized that the mixture of carcinogenic and mutagenic substances present in $PM_{2.5}$, including polycyclic aromatic hydrocarbons (PAHs) and other transition metals, can be metabolized in the body and cause DNA damage, genomic instability, and promote malignant neoplasms (51-53). Based on substantial evidence for a causal association between particulate matter air pollution and increased risk of lung cancer, air pollution exposure was classified as a carcinogen by the World Health Organization International Agency for Research on Cancer (IARC) in 2013 (52).

While the carcinogenic effects of air pollution are well established, less is known about the impact of air pollution on lung cancer survival. There are several plausible pathways through which $PM_{2.5}$ exposure may influence lung cancer survival. Exposure to $PM_{2.5}$ results in exposure to free radicals, organic chemicals, and transition metals. This exposure generates reactive oxygen species (ROS), resulting in genotoxicity, oxidative stress, and inflammation (54, 55). Further, $PM_{2.5}$ can act directly on lymphocytes, monocytes, and macrophages, affecting immune and inflammatory responses (56). Together, these responses can then result in systemic inflammation (57, 58), impaired immune response (59, 60), a decline in lung function (61), vascular impairment (62), cardiovascular disease (39, 63), respiratory infections (60, 64) and impaired cardiometabolic response. Exposure to $PM_{2.5}$ may impact lung cancer survival more directly by reducing treatment efficacy, resulting in adverse surgical outcomes, altering immune response, or promoting proinflammatory states (65-67). It is plausible that the established carcinogenicity of $PM_{2.5}$ contributes to lung cancer survival through the acceleration of carcinogenic processes, the development of secondary metastases, and secondary primary tumors (68, 69).

The following section reviews recent epidemiologic studies that examine the relationship between air pollution and cancer and lung cancer survival.

1.5 Review of epidemiologic studies on air pollution and cancer survival

Five studies worldwide have published research on ambient air pollution after diagnosis and cancer survival, beginning in 2013 (70-74) (**Table 1.2**). Of these, only two examined lung cancer survival (70, 73). These studies examined this relationship using data from either California Cancer Registry or Surveillance, Epidemiology and End Results Program (SEER). These studies used time from date of cancer diagnosis to date of death or last follow-up as the primary outcome of interest. Cox proportional hazards models were used to estimate hazard ratios for the impact of

average air pollution over follow-up on survival time. Notably, all of the identified studies assigned air pollution exposure using an average exposure from diagnosis to date of last follow-up. All five studies found positive associations between increased average air pollution exposure and decreased survival after a cancer diagnosis. In four of the five studies, effect modification was identified by stage at diagnosis, where hazards increased for patients with early-stage disease (70-72, 74).

Table 1.2 Air pollution and cancer survival studies

Reference	Study Population	Period	Exposure	Statistical method	Main Findings
Hu, 2013 (72)	Breast cancer cases from California SEER	1999-2009	County monthly mean $PM_{2.5}$	Cox proportional hazards models, breast cancer-specific hazard	$PM_{2.5}$ associated with decreased breast cancer survival, HR 1.86, 95% CI 1.12-3.10
Xu, 2013 (73)	Respiratory cancer cases from Honolulu and Los Angeles SEER	1992-2008	County mean $PM_{2.5}$ over follow-up	Cox proportional hazards models, all-cause and respiratory cancer-specific hazard	$PM_{2.5}$ associated with decreased respiratory cancer survival, HR 1.49, 95% CI 1.45 – 1.53
Eckel, 2016 (70)	Lung cancer cases from California Cancer Registry	1998-2009	Zip code monthly mean $PM_{2.5}$	Cox proportional hazards models, all-cause and lung cancer-specific hazards, stratified by stage and histology	$PM_{2.5}$ associated with decreased lung cancer survival particularly among early-stage tumors, HR 1.16, 95% CI 1.16-1.17
Deng, 2017 (71)	Liver cancer cases from California Cancer Registry	2000-2009	Zip code monthly mean $PM_{2.5}$	Cox proportional hazards models, all-cause and liver cancer-specific hazards, overall and stratified by stage	$PM_{2.5}$ was associated with decreased liver cancer survival with a nonlinear association, HR 1.18, 95% CI 1.16-1.20
Villanueva, 2021 (74)	Ovarian cancer cases from California Cancer Registry	1996-2014	Address mean $PM_{2.5}$ over follow-up	Cox proportional hazards models, ovarian cancer-specific survival, overall and stratified by stage	$PM_{2.5}$ was associated with decreased ovarian cancer survival, HR 1.45, 95% CI 1.41-1.49

1.5.1 Methodological challenges and proposed solutions

It is well known that in a time-to-event setting with time-varying exposures, it is necessary to account for both the time-varying nature of the exposure and differences in length of follow-up between censored and non-censored individuals. However, the existing literature overlooks this necessity and instead treats a time-varying exposure (air pollution) as time fixed by averaging exposure over follow-up for all individuals. This averaged exposure creates a study population where it is no longer possible to account for differences in follow-up lengths or ensure comparable exposure periods between censored and non-censored individuals. This is because the process of averaging an exposure over follow-up encodes follow-up duration into the exposure value itself. Using attained duration of follow-up to assign exposure may lead to one form of a common systematic error in epidemiology, immortal time bias (ITB). To date, this error has not been formally described in this context. Therefore, this dissertation will demonstrate how this error can be introduced and avoided using established methods in a case study of air pollution and lung cancer survival. This will be the focus of the second aim.

1.6 Racial residential segregation and lung cancer survival

Racial/ethnic residential segregation refers to the extent to which two or more groups live separately from one another in a geographic area (75). Racial segregation is not a choice made by residents but a consequence of generations of systemic racism and discriminatory policies (76, 77). Policies in the 1930s were designed to increase housing for white Americans, driving Black and other populations of color into urban housing projects. This segregation was strengthened with the introduction of policies that systematically denied these neighborhoods opportunities for mortgages. While these policies have since been made illegal, lasting effects have resulted in unequal opportunity for individuals and communities of color (78). Cities today remain deeply

segregated despite the introduction of policies in the 1960s which were introduced to promote equity and combat housing discrimination.

Segregation is well established as a fundamental cause of racial disparities and poor health. Segregation concentrates poverty, perpetuates racial bias, and shapes neighborhoods' social and physical features (79). On average, segregated neighborhoods have lower access to quality affordable housing, education, quality jobs, healthy food, green space, physical activity resources, and quality health care (79-81). Segregated neighborhoods have fewer primary and specialty care physicians, resulting in decreased utilization of preventive care and higher use of emergency departments for routine care (82-84). This neighborhood-built environment results in concentrated stress and poor health outcomes.

Beyond factors that impact communities' daily health and safety, one of the fundamental ways racial residential segregation can impact cancer is through access to care. In general, segregated neighborhoods have increased difficulty in accessing and using health care resources (85). Recent studies have documented that hospital-level segregation and hospital racial composition are associated with disparities in treatment and utilization (86-88). Safety net hospitals, with higher proportions of uninsured and underserved patients, are more often located near segregated areas and have higher rates of adverse patient safety events and lower quality of care (89, 90). Studies have demonstrated that patients living in segregated neighborhoods have decreased receipt of appropriate cancer care (91, 92).

Capturing an inherently multidimensional and spatial process through a single index is a challenge that has been discussed for decades (93). To date, there is no unified conceptual or methodological approach that is considered “best practice” in measuring racial residential segregation. It was first proposed that segregation can be conceptualized in five distinct

dimensions: evenness, exposure, clustering, concentration, and centralization (75). Recent work has demonstrated that after accounting for spatial relationships, these dimensions can be reduced to only: evenness and exposure (94). Several indices can be used to capture these dimensions, such as the commonly used isolation and dissimilarity indices. Other less commonly used indices have been proposed, such as hyper segregation (95) and index of concentration at the extremes (ICE) (96). Some studies use the percent of a racial/ethnic group in an area to measure segregation. However, this crude measure has been widely criticized because this does not describe how two or more population groups are distributed across space (97).

The following section reviews recent population studies examining the relationships between air pollution on cancer and lung cancer survival.

1.7 Review of epidemiologic studies on segregation and cancer survival

There is growing evidence that cancers are unequally distributed across geographic areas that differ by socioeconomic status, racial/ethnic composition, and rurality (98, 99). There are substantial racial and ethnic disparities in lung cancer, despite overall improvements in survival over the last several decades (100). In lung cancer, many factors may contribute to adverse health outcomes. Studies have demonstrated that patients with cancer in segregated communities are more likely to die from their cancer (101), be diagnosed at later stages of disease (102), and receive lower rates of appropriate care (103). To review the current literature on segregation and cancer survival, we searched Google Scholar using the terms: racial segregation, residential segregation, and cancer survival, cancer prognosis. Theoretical articles, commentaries, or those without explicit measures of segregation or cancer survival were excluded. Using this criteria, seven articles were identified and outlined in **Table 1.3**.

Table 1.3 Racial residential segregation and cancer survival studies

Reference	Study Population	Period	Segregation measure	Main Findings
Warner, 2010 (104)	Breast cancer cases from the California Cancer Registry	1996-2004	Hyper-segregation, high segregation on two or more indices	Metropolitan segregation did not explain differences in survival between Black and white women
Johnson, 2016 (103)	Early-stage NSCLC patients from the Georgia Cancer Registry	2000-2009	Isolation index	Black patients living in areas with high economic deprivation and high segregation were more likely to die, even after controlling for surgery 1.31, (1.04-1.66)
Zhou, 2017 (105)	Colorectal cancer cases from Wisconsin Cancer Reporting System	2004-2011	Racial bias, redlining, location quotient (local area segregation measure)	Racial bias in lending was associated with greater hazard (1.37, 1.06-1.17) No associations between redlining or location quotient
Poulson, 2020 (106)	Colorectal Cancer cases from SEER	2005-2015	Index of dissimilarity	Black patients had decreased cancer-specific survival (HR 1.43, 1.17-1.74)
Poulson, 2020 (107)	Breast Cancer cases from SEER	2005-2015	Index of dissimilarity	Black patients had an increased hazard of death with increasing segregation (RR 1.29, 1.04-1.60)
Westrick, 2020 (108)	Epithelial ovarian cancer cases from Florida Cancer Data System	2001-2015	5 ICE variables: economic, race/ethnicity, racialized economic segregation	Economic and racialized economic residential segregation influenced EOC survival more than race or ethnic segregation alone
Harris, 2021 (109)	Oral squamous cell carcinoma from SEER	2005 - 2015	Index of dissimilarity	Highly segregated counties, advanced age, advanced stage at time of diagnosis, and lack of surgery were poor prognostic factors among Black patients

These studies used a variety of different segregation measures to capture their exposure of interest. Three of the ten papers used one of the most common measures, the index of dissimilarity, to capture racial residential segregation (106, 107, 109). This index measures evenness across space and captures the proportion of a group's population that would need to change residence for

each neighborhood to have the same proportion of that group as the general geographic area. Of the seven identified papers, only one examined lung cancer survival (103). All four papers that used common segregation measures (described above) identified an association between racial residential segregation and decreased survival after a cancer diagnosis (103, 106, 107, 109). Of the three that used other measures to capture segregation impacts, only one identified that economic segregation played a larger role in epithelial ovarian cancer survival than racial or ethnic segregation (108). One study that measured hyper-segregation found that segregation did not explain any survival disparities between Black and white women with breast cancer (104).

1.7.1 Methodological challenges and proposed solutions

The current literature in segregation and cancer survival has overlooked specific methodological challenges in the measurement of segregation. There are two primary factors in measuring segregation that need to be considered: the spatial scale of geographies and spatial relationships of populations. First, it is necessary to consider the spatial scale of both the subareas and the larger geographic area used in analysis. Most common segregation measures combine information on two geographic scales: subareas (e.g., neighborhoods, census tracts) within larger geographic areas (e.g., cities, counties). These two geographic scales are required because the racial makeup of one neighborhood is meaningless without the context of the surrounding area. It is well established that the size of these areal units can change the value of the segregation index (110), which can then impact the effect estimate of interest. For example, studies have demonstrated that segregation impacts on health are usually more substantial when segregation is measured in small areas (111). For this reason, it is necessary to make explicit the choice of spatial scale or to conduct sensitivity analyses using different spatial sizes.

Second, it is essential to consider the spatial relationships between populations in measuring segregation (94). Common segregation measures treat neighborhoods as discrete units with no interaction, with exogenously assigned boundaries treated as impenetrable walls (112). Because of this feature, these measures cannot differentiate between population distribution patterns, instead only evaluating population mix within a spatial unit (113). To overcome this limitation, spatial segregation measures have been proposed that explicitly consider this spatial element (94, 112, 114, 115). These measures incorporate an adjacency component to account for the fact that there is more interaction between closer areas, which decreases as distance increases. Previous studies have demonstrated that when segregation indices that account for both scale and spatial orientation of neighborhoods are used, the absolute estimation of segregation changes (116).

All seven studies that examine segregation and lung cancer survival do not make explicit the choice of spatial scale for analysis, nor do they use segregation measures that account for spatial relationships of populations (**Table 1.4**). Two of the seven do not provide information for how the segregation measure was calculated (104, 109). A systematic comparison of different ways to calculate segregation that may impact an exposure's amplitude and effect size on lung cancer survival among different racial groups is lacking. Therefore, this dissertation will demonstrate how these two considerations in measuring racial residential segregation can impact effect estimates using a case study of segregation and lung cancer survival. This will be the focus of the third and final aim.

Table 1.4 Racial residential segregation and cancer survival studies: segregation measures

Reference	Segregation Measure	Spatial scale (subgroups)	Spatial scale (larger area)	Discussion of spatial scale	Measure that accounts for spatial relationships
Warner, 2010	Hyper-segregation, high segregation on two or more indices	Unknown	Metropolitan Area	No	No
Johnson, 2016 (103)	Isolation index	Census block	Census tract	No	No
Zhou, 2017 (105)	Racial bias, redlining, location quotient (local area segregation measure)	Zip code	Metropolitan area	No	No
Poulson, 2020 (106)	Index of dissimilarity	Census block	County	No	No
Poulson, 2020 (107)	Index of dissimilarity	Census block	County	No	No
Westrick, 2020 (108)	5 ICE variables: economic, race/ethnicity, racialized economic segregation	Neighborhood	Unknown	No	No
Harris, 2021 (109)	Index of dissimilarity	Unknown	County	No	No

1.8 Summary

This dissertation expands on the current literature on neighborhood contextual determinants in lung cancer survival by addressing the outlined methodological challenges. First, we will describe how lung cancer varies spatially in California and examine potential clusters of exceptionally high or low survival. Next, we will assess the association between air pollution and lung cancer survival using a method that treats air pollution as time-varying and avoids systematic

errors. Finally, we will estimate the association between racial residential segregation and lung cancer survival using spatial and aspatial segregation measures across different geographic scales. We illustrate how both of these considerations can impact effect estimates and subsequent conclusions.

1.9 Specific aims

This dissertation examined the relationship between neighborhood contextual factors (air pollution and racial residential segregation) on lung cancer survival. The following aims are addressed:

Aim 1: Describe spatial patterning in age-adjusted lung cancer survival in the three largest California counties

1a: Describe the changes in spatial heterogeneity over time

1b: Identify potential high and low clusters in age-adjusted survival over California

Aim 2: Identify and correct the systematic bias in studies that use average air pollution as a time-fixed exposure in a time-to-event analysis under a target trial framework

2a: Estimate the association between air pollution after a lung cancer diagnosis on five-year lung cancer survival using a discrete-time hazards approach to account for time-varying air pollution

2b: Compare the effect estimate between the naïve analysis to the proposed solution using discrete-time models

Aim 3: Estimate the association between residential racial segregation and lung cancer survival using segregation indices calculated using different approaches

3a: Calculate segregation using different geographic areal unit sizes (census tract vs. census block, five- vs. ten-mile neighborhood), and using spatial and aspatial indices

3b: Compare conclusions between spatial and aspatial indices and by the size of geographical unit concerning racial/ethnic disparities in lung cancer survival

This research studied populations of lung cancer cases in California, United States. All three aims in this dissertation used lung cancer cases from the California Cancer Registry from 2000 – 2010. California is the most populous state in the nation, with more than 39 million residents in 2020. California is exceptionally diverse, with no single racial/ethnic group making up most of California's population. Further, Hispanic populations of any race are the largest single ethnic group in the state. In 2015, 16,000 individuals were diagnosed with lung cancer, and 12,000 died of the disease in California. These aims provide valuable evidence about potential prognostic factors in lung cancer that vary over space and time and address methodological challenges prevalent in the literature.

2 Spatial variation in lung cancer survival across time in Southern California

2.1 Abstract

Background: Studies frequently examine trends in lung cancer survival over time but rarely examine heterogeneity over space. Identifying variability in lung cancer survival over space and time can be used to generate hypotheses about what place-based factors may drive variability and allow for targeted interventions.

Methods: Lung cancer cases from the California Cancer Registry from 2000 – 2010 in the three largest counties in California (Los Angeles, Orange, San Diego) were used to calculate age-adjusted zip code specific median survival. The ratio of observed to expected median survival was mapped across all zip codes. We assessed spatial autocorrelation of the age-standardized median survival times using the *Getis-Ord Gi** statistic to identify local clusters of above and below expected survival in Los Angeles County.

Results: This study included 44,995 lung cancer cases. Over the five-year follow-up period, 84% (n=38,058) of lung cancer cases died. The five-year median survival for lung cancers diagnosed during the study period was 4.95, 1.62, and 0.38 years for localized, regional, and distant stage at disease, respectively. There was wide geographic variation in the observed to expected ratio in all three California counties. The cluster analysis revealed significant hot and cold spots across Los Angeles County.

Conclusions: There is important spatial and temporal heterogeneity in lung cancer survival in California, suggesting that there are likely prognostic factors in survival that vary in similar patterns. It is necessary to evaluate disparities in lung cancer survival over space and time to reveal

potential risk factors and design targeted interventions. Evaluating global trends alone will miss important small-area heterogeneity over space and time.

2.2 Introduction

By the end of 2021, lung cancer will account for approximately 8% of new cancer cases and 20% of cancer deaths, more deaths than colon, breast, and prostate cancer combined (117). Lung cancer survival is lower than other leading cancers, and more than half of people with lung cancer die within one year of being diagnosed. Small advances in lung cancer survival to date are primarily attributable to individual-level interventions in tobacco use, early diagnosis, and improved treatment (13, 24). This continued poor survival suggests that new approaches are needed to identify modifiable and structural risk factors. Recently, there has been growing interest in understanding small-area spatial variation in cancer survival as a mechanism to identify novel place-based risk factors and inform local policies to create targeted interventions (118).

It has become widely accepted that the place where an individual lives or works is an important determinant of health. Studies have demonstrated that neighborhood social and built environments are independently associated with mortality (119). In the US, the place an individual lives can determine their access to education, income, green space, healthy food, clean air and water, and quality health care (120-122). In lung cancer, identifying place-based risk factors may be particularly valuable as they do not rely on individual interventions. Examining spatial variation in lung cancer survival can help identify areas with disproportionately high or low survival rates. Recently, studies have illustrated that identifying changes in spatial patterns of lung cancer survival over time can aid in the identification of risk factors that vary both spatially and temporally (123-126).

Previous work has identified geographic variation in cancer survival between counties (118) and within the United States (127). Richards et al. analyzed net survival by state in the US and found large geographic variation in lung cancer survival, with notable racial disparities. In the 37 states with comprehensive cancer registries, the five-year survival ranged from 15% to 25% among white patients and 7% to 23% among Black patients (30). At the same time, few studies have examined spatial heterogeneity in lung cancer survival over time. Cramb et al. examined changes in lung cancer survival in Queensland AU and found large spatial variation in survival, with variation in changes over time (124). This evidence of large-scale spatial heterogeneity in lung cancer survival highlights the possibility that spatially and temporally varying risk factors are driving these geographic patterns.

To date, no studies have described how lung cancer survival varies spatially in California and how these spatial patterns change over time. A recent study examined cancer survival in California over the past two decades and found that five-year survival improved from 18 to 29%. However, study authors noted that they did not assess how these trends varied spatially across the state (128). California has an exceptionally diverse racial and ethnic composition, with no single racial or ethnic group constituting most of the population. California has the largest population in the US (making up 12% of the US population), and 95% of its residents live in urban areas. Unfortunately, California is only ranked 16th in lung cancer survival rates in the nation (127). Understanding how lung cancer survival varies spatially across urban areas in the state is necessary to identify potential modifiable risk factors and influence resource allocation to improve health equity in the state.

This study describes how lung cancer survival varies spatially across the three largest counties in California: Los Angeles, San Diego, and Orange County. To do this, we apply indirect

standardization to obtain age-adjusted median survival for all zip codes in these counties. To illustrate how spatial heterogeneity in survival may change over time, we stratify the study period between 2000-2005 and 2006-2010. Finally, we evaluate the presence of spatial autocorrelation in Los Angeles to identify potential clusters (hot or cold spots) where lung cancer survival may be better or worse than expected after adjusting for potential confounding by age distribution. All analyses are stratified by stage at diagnosis.

2.3 Methods

All lung cancer cases in the three largest counties in California (Los Angeles, San Diego, Orange) from 2000 – 2010 were obtained using the California Cancer Registry (CCR). CCR is a statewide population-based cancer surveillance system that collects information on all cancers (except for non-melanoma skin cancer) diagnosed in California. Lung cancer cases were identified with the International Classification of Disease of Oncology (ICD-0-3) site codes (C34.0 – C34.3, C34.8, C34.9). Cases were included in the study if they had a primary, histologically confirmed cancer of the lung or bronchus and had an address in the study area at the time of cancer diagnosis on record with CCR (N=49,202). Patients with in situ cancer, diagnosed at autopsy, or with an invalid date of diagnosis or follow-up (N=3,661) were excluded from analysis. Patients who were lost to follow-up before five years were excluded from analysis (N=546). The final study population included 44,995 patients. This study was reviewed and approved by Institutional Review Boards (IRBs) at San Diego State University, the University of California San Diego, and the California Department of Public Health Committee for the Protection of Human Subjects.

Standardization

Measures of cancer prognosis most often report five or ten-year survival rates. When using absolute survival rates, it is necessary to account for age distribution in the target population (129). This is because survival varies strongly with age. In order to make comparisons of cancer survival between populations in different areas or across time, it is necessary to remove the influence of age (when there are differences in the age distribution across comparison areas). Because of this, comparisons between cancer survival rates are age standardized. Age-standardized survival rates use a weighted average of age-specific survival rates within age subgroups, equal to the proportion of patients in these age groups in some standard population (130).

The outcome of interest for this study was median survival, censored at five years. We used an indirect standardization procedure to age-adjust zip code specific median survival for each county separately. Here, our standard population was the county in which the zip code resides. This standardization procedure used quintiles of the age distribution of all cancer cases for each county. We then calculated the median survival times for each age group quintile at the county level. We multiplied these age group specific median survival times by the number of lung cancer cases in each age group at the zip code level. We then summed these median survival times and divided them by the number of people in each zip code. This results in median survival times for each zip code had the zip code taken on the age distribution of the county. A ratio of observed over expected was used to present median survival relative to what would be expected in the county. This means that for values greater than one, the observed survival is greater than expected. For values less than one, observed survival is worse than expected, given the population's age distribution.

Spatial Autocorrelation

We assessed spatial autocorrelation of the age-standardized survival rates using local *Getis-Ord Gi** statistic to identify local patterns and clusters of above and below expected survival. *Getis-Ord Gi** is a type of Local Indicator of Spatial Associations, known as a LISA statistic. Neighbors were defined using contiguity, where common sides and points of polygons define the neighbor relation. The *Getis-Ord Gi** statistic was then used to categorize and visualize cold and hotspots (blue and red) and is represented with a Z-score.

To evaluate how these spatial patterns may change over time, the standardized median survival times and spatial autocorrelation analysis were stratified by year at diagnosis to create two time periods (2000-2005 and 2006-2010). All visualizations and the spatial autocorrelation analysis were stratified by stage at diagnosis, as stage is one of the strongest prognostic factors for survival. All data analysis was conducted using R software V 4.0.2.

2.4 Results

A total of 44,995 lung cancer cases were included in this analysis. The average age at diagnosis was 69 (SD 12), and over half of all cases were diagnosed at distant stage at disease (63%). The majority of the population was non-Hispanic white (65%), followed by Hispanic (12%) and Asian/Pacific Islander (12%). Over the five-year follow-up period, 84% (n=38,058) of lung cancer patients died. The five-year median survival for lung cancers diagnosed during the study period was 4.95, 1.62, and 0.38 years for localized, regional, and distant stage at disease, respectively. During the study period, overall median survival improved for all stages of disease and age groups (Supplementary Material **Table 2.2**). Study population characteristics stratified by county are presented in **Table 2.1**.

Table 2.1 Study population characteristics, stratified by county of residence, California Cancer Registry, 2000 - 2010

	Los Angeles	Orange	San Diego
	n=25,072	n=9,567	n=10,356
Age Category			
[18,59]	5280 (21.1)	1871 (19.6)	1957 (18.9)
(59,67]	5341 (21.3)	2036 (21.3)	2081 (20.1)
(67,73]	4854 (19.4)	1889 (19.7)	2021 (19.5)
(73,79]	4825 (19.2)	1882 (19.7)	2165 (20.9)
(79,101]	4772 (19.0)	1889 (19.7)	2132 (20.6)
Male	13233 (52.8)	4834 (50.5)	5174 (50.0)
Partnered	12675 (51.8)	5320 (57.2)	5267 (52.0)
Race			
American Indian / Alaskan Native	35 (0.1)	14 (0.1)	37 (0.4)
Asian/Pacific Islander	3591 (14.3)	1223 (12.8)	849 (8.2)
Hispanic	3657 (14.6)	771 (8.1)	1055 (10.2)
NH Black	3741 (14.9)	115 (1.2)	496 (4.8)
NH White	14022 (56.0)	7435 (77.8)	7912 (76.5)
Stage at Diagnosis			
Distant	16322 (65.1)	5792 (60.5)	6412 (61.9)
Localized	3612 (14.4)	1665 (17.4)	1699 (16.4)
Regional	5138 (20.5)	2110 (22.1)	2245 (21.7)
Time of follow-up (Mean)	1.5 (1.7)	1.7 (1.8)	1.5 (1.8)

There was wide geographic variation in the observed to expected ratio in all three California counties (**Figures 2.1, 2.2 and 2.3**). In Los Angeles, distinct spatial patterning was consistent across both time periods and all stages of diagnosis. In general, areas in the west experienced better than expected survival (observed to expected ratio greater than 1), while areas in the east experienced worse than expected survival (observed to expected ratio less than 1). Over the study period, this pattern remained broadly consistent, except for some areas where observed

survival increased or decreased relative to the expected. A similar pattern was observed in San Diego and Orange County, where areas on the coast tended to have better than expected survival, while areas inland tended to have worse than expected survival. This pattern remained consistent across both time periods and all stages of diagnosis.

Figure 2.1 Map of Los Angeles County: Ratio of observed to expected median survival, age-adjusted to Los Angeles County age distribution, California Cancer Registry, 2000 - 2010

Figure 2.2 Map of Orange County: Ratio of observed to expected median survival, age-adjusted to Orange County age distribution, California Cancer Registry, 2000 - 2010

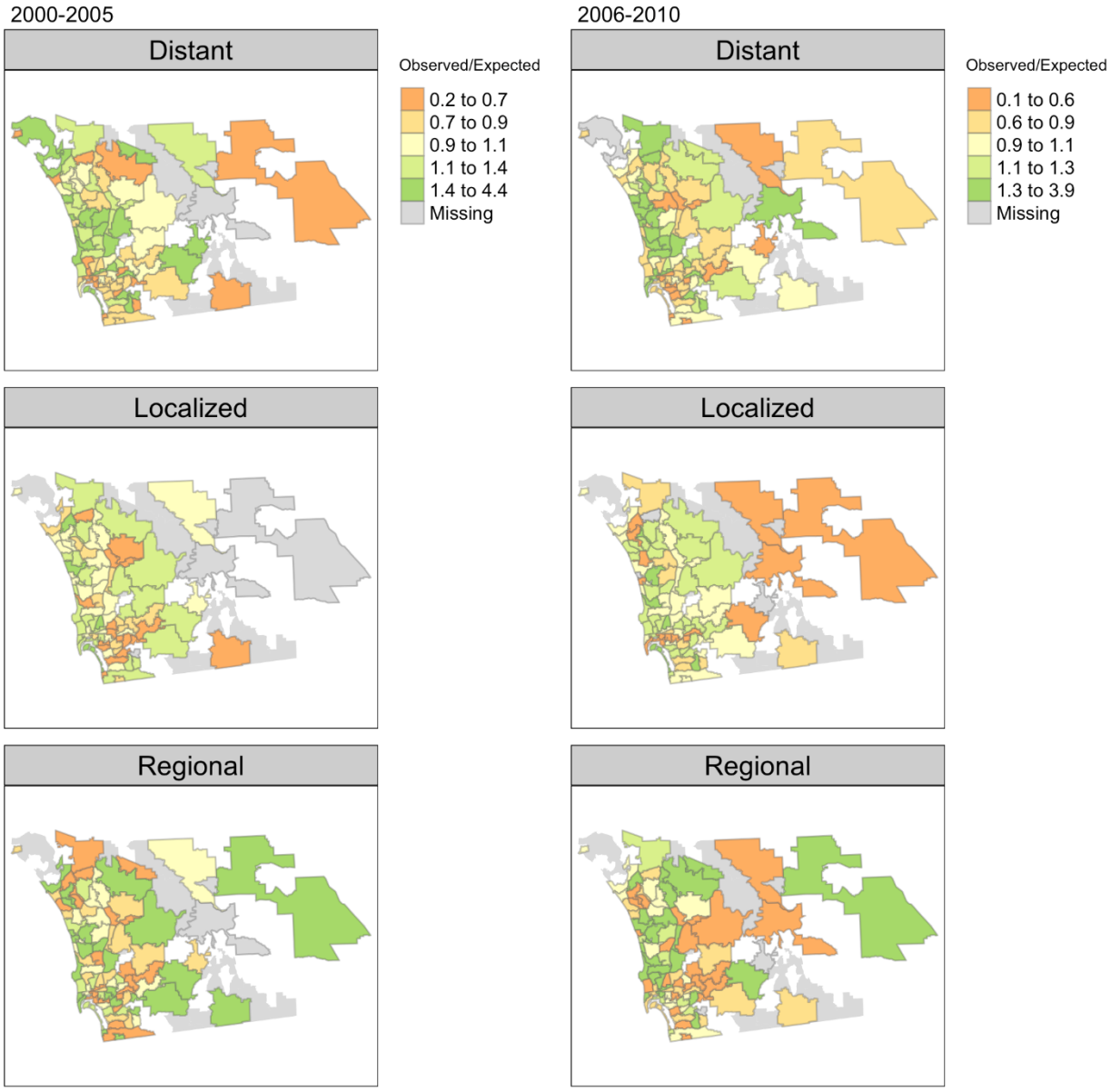

Figure 2.3 Map of San Diego County: Ratio of observed to expected median survival, age-adjusted to San Diego County age distribution, California Cancer Registry, 2000 - 2010

In the spatial autocorrelation analysis, hot and cold spots were identified in Los Angeles (**Figures 2.4 and 2.5**). These hot and cold spots corresponded to areas where better and worse than expected survival tended to be clustered in **Figure 2.1**. There was a consistent pattern across all

stages and both time points, where areas in the west were consistently identified as ‘cold spots’, indicating better than expected survival. Similarly, in the east, there were ‘hot spots’ where survival was worse than expected. Over the study period, these hot and cold spots changed slightly, with some areas becoming stronger clusters and others becoming weaker. However, in general, the patterns remained largely the same.

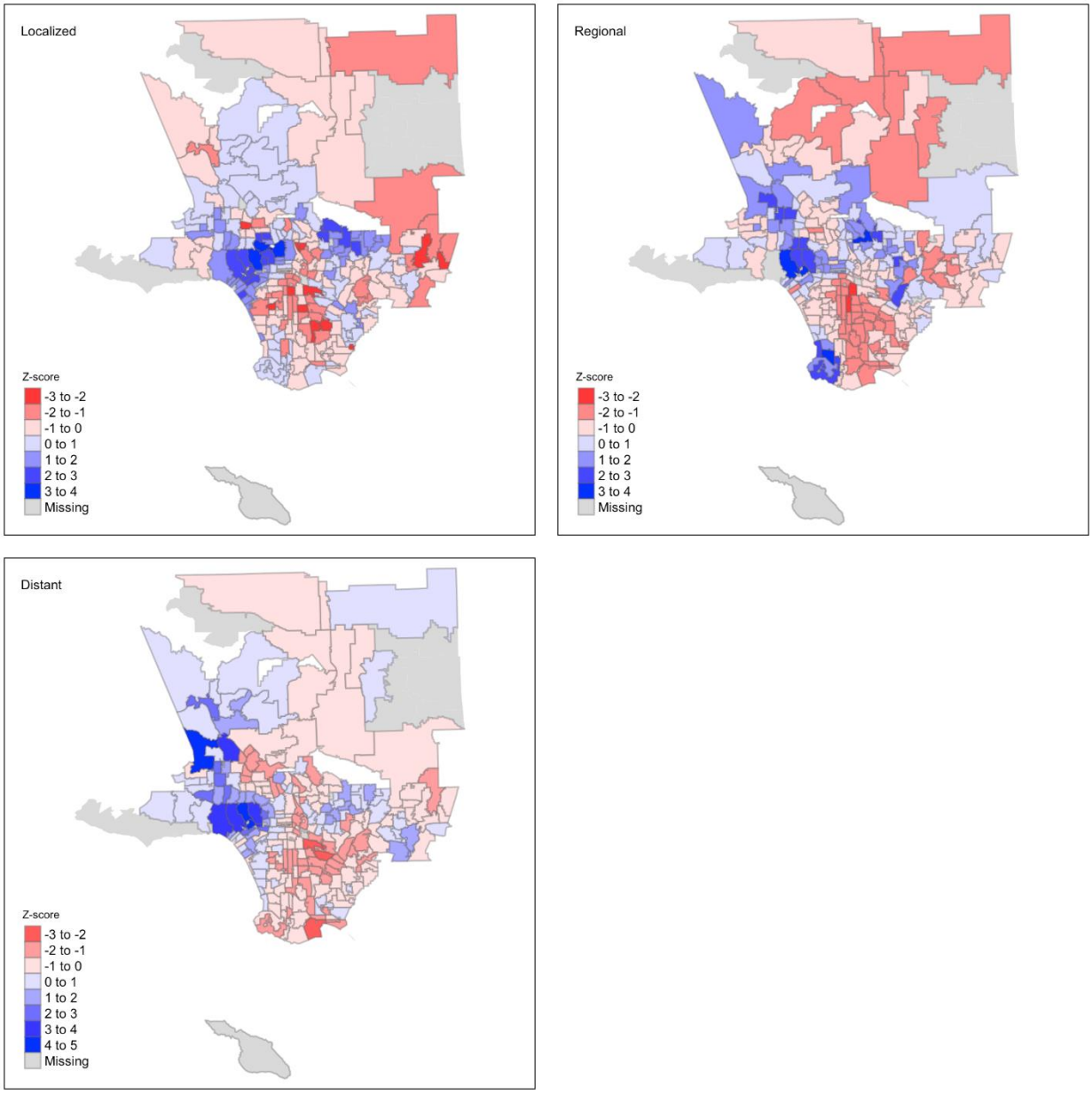

Figure 2.4 Getis-Ord Gi* cluster analysis, Los Angeles, California Cancer Registry, 2000 - 2005

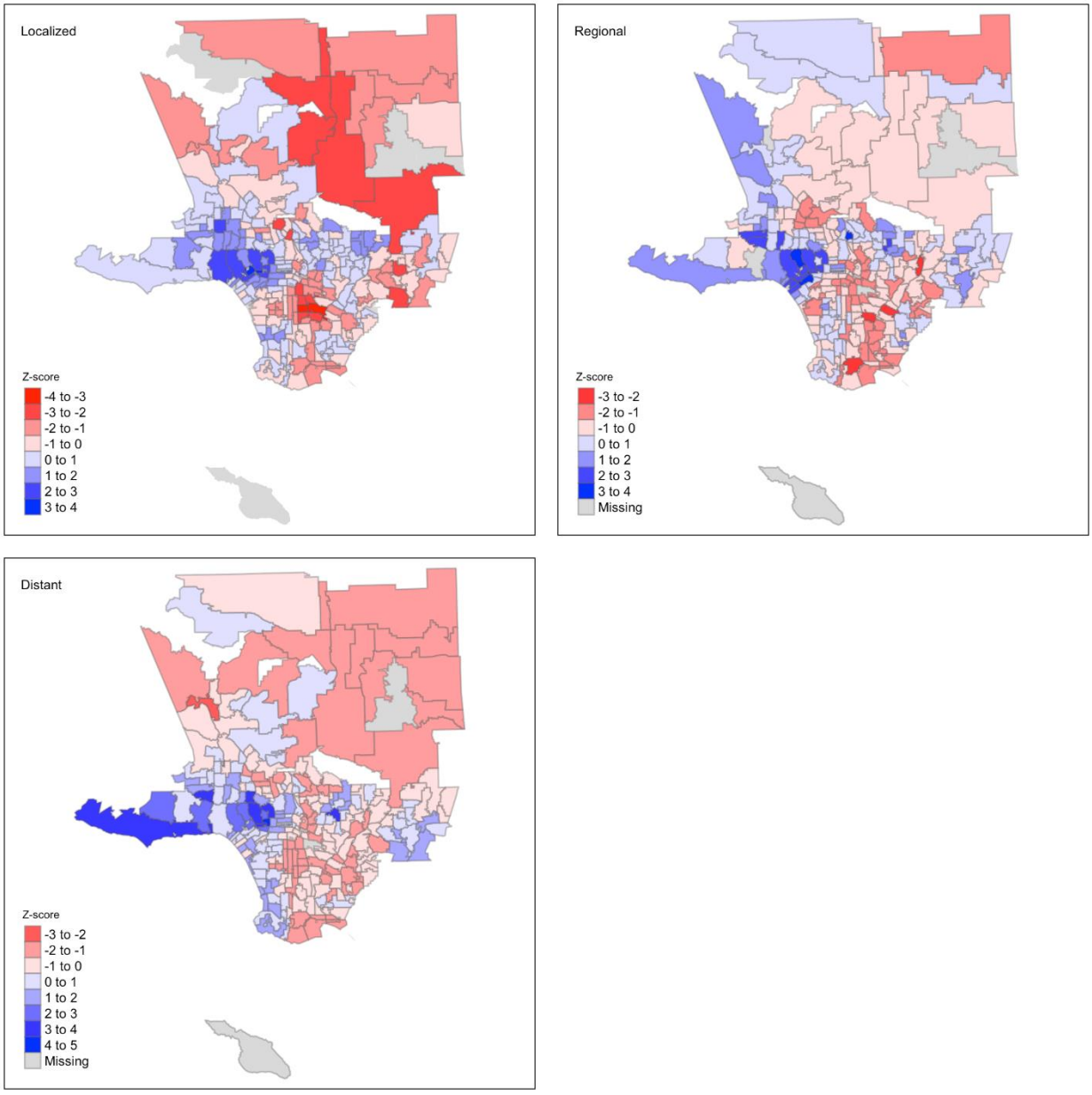

Figure 2.5 Getis-Ord Gi* cluster analysis, Los Angeles, California Cancer Registry, 2006 – 2010

2.5 Discussion

In the three largest counties in California, lung cancer survival varied between zip codes and across time. There were areas where the observed to expected ratio remained consistent across

time, while other areas increased or decreased. These results suggest that it is not sufficient to describe lung cancer survival globally (state or county-wide) without considering small-scale spatial variability as this may miss important areas of exceptionally low or high vulnerability. Our spatial autocorrelation analysis identified patterns of hot and cold spots that are consistent across diagnosis stages and time. Our findings highlight that there may be important drivers of lung cancer survival that vary geographically to create this spatial heterogeneity and clusters of exceptional and poor survival.

Several factors vary across space that may contribute to spatial heterogeneity in lung cancer survival. Stage at diagnosis, receipt of guideline-concordant care, and treatment at high volume centers are important prognostic factors in lung cancer. Because of this, spatial factors that affect access to care are likely also to influence survival (131-133). Studies have demonstrated that distance to cancer centers and access to other supportive care are associated with later stage at diagnosis, decreased receipt of appropriate treatment, and subsequent survival in cancer (134, 135). People living in rural areas worldwide are more often diagnosed at later stages of disease and have higher mortality, particularly for Black patients (3, 125, 136). One study found that the uptake of novel treatments for lung cancer varied across counties and regions, which may contribute to some spatial survival disparity (118). Lastly, studies have examined lung cancer survival over time and found that treatment center significantly influences survival (137).

Other factors unrelated to access to care may also be drivers of spatial heterogeneity in cancer survival. Social and environmental factors have been shown to impact a patient's cancer prognosis differentially (138). Socio-economic deprivation is associated with lower rates of treatment in lung cancer patients, which may partially explain spatial variation in survival. Similarly, some studies have demonstrated that racial residential segregation may adversely impact

cancer survival (103, 105, 107, 109). Finally, there is evidence that exposure to air pollutants and toxins may impact survival after a cancer diagnosis (70, 73, 74, 139-141). These contextual determinants are promising avenues for future research as they relate to interventions at the population level and may have important implications for improving health inequities.

In California, lung cancer survival improved over the last two decades. However, there are disparities in survival and rates of change in survival across population subgroups. Women have consistently better survival than men (partially due to increased receipt of surgery) and have had faster increases in survival than their male counterparts (128). Black patients have had lower survival rates, lower surgical rates, and are more likely to be diagnosed with advanced stage disease (100). At the same time, Asian American and Pacific Islander patients have the highest survival rates compared with white patients. While many spatial factors may contribute to these disparities, studies have demonstrated that factors associated with access to care and stage at diagnosis, such as neighborhood socioeconomic status, play the largest role in racial disparities for Black and Hispanic patients (100).

There are several limitations in the present work. First, we estimated absolute survival. While absolute survival can be beneficial to give a complete picture of health in a population across zip codes, it may reflect underlying factors that result in poor health across a range of outcomes. In this work, we age-adjusted within each county. This means that we cannot make any conclusions comparing survival between counties. Finally, mapping small-area variation in lung cancer is always constrained by small cases numbers. In this study, we have some zip codes with small population numbers that may cause some instability in our estimates. The main objective of this paper was descriptive, and we did not attempt to quantify nor identify drivers behind geographic

variation and changes in survival over time. Future work should consider how neighborhood-level factors, such as social or environmental factors, may impact this geographic heterogeneity.

In this study, we identified important spatial and temporal heterogeneity in lung cancer survival in California, suggesting that there are likely prognostic factors in survival that vary in similar patterns. It is necessary to evaluate disparities in lung cancer survival over space and time to reveal potential risk factors and design targeted interventions. Evaluating global trends alone will miss important small-area heterogeneity.

Supplementary Material

Table 2.2 Median survival by age group from 2000 - 2005 & 2006 - 2010, California Cancer Registry

	Age Category	2000 - 2005	2006 - 2010
Distant	[18,59]	0.58	0.66
	(59,67]	0.45	0.54
	(67,73]	0.37	0.38
	(73,79]	0.27	0.3
	(79,101]	0.2	0.22
Regional	[18,59]	2.09	2.88
	(59,67]	1.88	2.38
	(67,73]	1.44	1.72
	(73,79]	1.3	1.47
	(79,101]	0.87	0.94
Localized	[18,59]	4.99	4.99
	(59,67]	4.99	4.99
	(67,73]	4.91	4.99
	(73,79]	3.48	4.39
	(79,101]	1.58	2.12

Chapter 2, in part, is currently being prepared for submission for publication of the material. Sheridan, Paige; Caroline, Thompson; Benmarhnia, Tarik. The dissertation author was the primary investigator and author of this paper.

3 Immortal time bias with time-varying exposures in environmental epidemiology: a case study in lung cancer survival

3.1 Abstract

Immortal time bias is a well-recognized bias in clinical and pharmacoepidemiology but is rarely discussed in environmental epidemiology settings. Under the target trial framework, this bias is formally conceptualized as a misalignment between start of study follow-up (time zero) and treatment assignment. This misalignment can occur when attained duration of follow-up is encoded into treatment status using minimums, maximums, or averages. The bias can be further exacerbated in the presence of important time trends commonly found in environmental exposures. We replicated previous studies that average air pollution exposure over follow-up in a time-to-event model using lung cancer cases from the California Cancer Registry (2000-2010) linked with $PM_{2.5}$ estimates. We compared this approach to a discrete-time approach that ensures alignment between time zero and treatment assignment. In the former approach, the estimated hazard ratios for a 5 $\mu g/m^3$ increase in average $PM_{2.5}$ were 1.35 (95% CI: 1.30-1.41), 1.29 (95% CI: 1.25-1.33), and 1.17 (95% CI: 1.15-1.18) for localized, regional, and distant stage of disease, respectively. By contrast, under the discrete-time approach, the estimated pooled ORs were 1.01 (95% CI: 0.99-1.01), 1.00 (95% CI: 0.98-1.01), and 1.00 (95% CI: 0.99-1.01). We identify that immortal time bias likely drives the differences in these effect estimates as a result of averaging exposure over differential follow-up periods. Our findings highlight the importance of appropriately conceptualizing a time-varying exposure in environmental epidemiology under the target trial framework to avoid introducing preventable systematic errors.

3.2 Introduction

Epidemiologic studies often seek to measure the impact of time-varying environmental exposures on time-to-event outcomes. In this setting, it is necessary to account for the time-varying nature of the exposure and differences in length of follow-up between censored and non-censored individuals. However, studies often overlook these requirements and attempt to simplify the analysis by treating the exposure as time-fixed to avoid the added complexity of accounting for a time-varying exposure. For instance, some studies choose to average the exposure over follow-up for each individual (70-74). While this may initially appear to be an appropriate simplification, this averaged exposure unintentionally creates a study population where it is impossible to account for differences in length of follow-up between individuals or the time-varying nature of the exposure. When an averaged environmental exposure is used in this context, a well-known systematic error can be introduced into the study, called immortal time bias (ITB). This paper uses the target trial framework to identify how averaging a time-varying environmental exposure in a time-to-event analysis can introduce ITB into an observational study. We then illustrate an approachable analytic solution that prevents ITB and appropriately considers the time-varying nature of these exposures. The terms “exposure” and “treatment” will be used synonymously throughout this paper.

The target trial framework views an observational study through the lens of a hypothetical randomized trial (target trial) to avoid fundamental flaws in the analysis that can result in biases, including ITB (142). This framework defines that the fundamental principles that guide the design and analysis of randomized trials must be applied in observational studies to avoid systematic errors. One of these central principles is the alignment of start of follow-up (time zero), eligibility criteria, and treatment assignment. This alignment at time zero occurs by design in a randomized

trial (at randomization), so systematic errors can be introduced when this principle is ignored in an observational study. While several errors can occur if this alignment principle is violated, we focus on ITB in this paper. Here, ITB occurs when the start of follow-up and eligibility criteria are aligned, but post-baseline information is used to assign treatment status (143) (creating a misalignment between treatment and time zero). For this discussion, we will refer to the "principle of alignment at time zero" to describe alignment between time zero and treatment assignment (assuming alignment of eligibility criteria).

ITB was first described in the context of the heart transplantation studies by Gail in 1972 (144). In this instance, heart transplantation was treated as a time-fixed exposure (transplanted or not) at baseline, and survival distributions were compared between those who never received a transplant and those who did. The principle of alignment at time zero was violated here because researchers retrospectively assigned individuals to the transplant group if they had received a heart transplant at any time over follow-up, therefore using post-baseline information to assign baseline treatment status. This approach ensured that patients in the transplant group had a guaranteed survival ("immortal time") during the time they waited to receive the transplant. Because the transplanted group was guaranteed this additional follow-up time, it produced an artificial increase in survival time among the treated, suggesting a survival benefit from heart transplant. After excluding this "immortal time", reanalysis of the same data revealed no survival benefit in the transplanted subjects (145).

While ITB is most often identified when there is a well-defined period of "immortal time" as described above, there are other subtle ways treatment status is assigned using post-baseline information. One example of this occurs when treatment assignment is dependent on an individual's attained duration of follow-up (146). This appears in studies that use maximums,

minimums, or averages of exposure during follow-up as the exposure of interest (143). This has been described in studies of "long-term" and "short-term" drug users, where a participant must have had a longer duration of follow-up to be classified as a long-term user (147, 148). For example, in an observational study that demonstrated a protective effect of statin use on lung cancer when comparing long-term statin users with non-users (149), individuals were classified using their observed duration of statin therapy *during* follow-up. This results in a study sample where being defined as a long-term statin user depends on having a longer length of follow-up. While there is no apparent period of "immortal time" here, we can identify that the principle of alignment at time zero is violated due to assigning treatment using attained duration of follow-up, resulting in ITB.

Because ITB is most commonly discussed in clinical and pharmacoepidemiology (145, 150, 151), its occurrence in environmental epidemiology is rarely discussed. However, we can identify that ITB can be introduced when environmental exposures are averaged over follow-up (in a time-to-event setting) using the target trial framework. Averaging an exposure over follow-up defines the exposure value based on an individual's observed length of follow-up, creating misalignment of treatment and time zero (as described above) (143). This results in a study sample where individuals with longer follow-up times have more exposure observations included in their average than individuals with shorter follow-up times. These averaged exposures (between individuals with different follow-up times) are no longer inherently comparable as a result of the different number of values included in the calculation. While this discrepancy in calculating averages is problematic in and of itself, the induced systematic differences between individuals with and without the outcome of interest can be even greater when the exposure has a strong temporal trend, as with air pollution (152, 153). When this is the case, the averaged exposure level

is associated with the length of follow-up and the temporal trend. For example, if an exposure decreases over calendar time, those with longer follow-up times will have lower averages than those with shorter follow-up times due to the temporal trend in the exposure. This results in an even greater discrepancy in assigned exposure values between those with longer and shorter follow-up times.

To date, no studies have illustrated how averaging time-varying environmental exposures over follow-up time in a time-to-event analysis can introduce ITB. This study presents a case study examining the impact of air pollution after a lung cancer diagnosis on five-year survival. We use two approaches to estimate this association. First, we define air pollution exposure using an average over follow-up (we refer to this as the naïve approach). We then use an alternate method that upholds the principle of alignment at time zero, a discrete-time approach. We compare effect estimates between the naïve and discrete-time approach to illustrate how ITB can impact effect estimates.

3.3 Methods

Several studies have examined the impact of air pollution exposure after a cancer diagnosis on cancer survival in a time-to-event framework (70-74). These studies treated air pollution exposure as time-fixed by averaging air pollution exposure from date of diagnosis to date of last follow-up and use Cox proportional hazards models to estimate hazard ratios for the association between average air pollution and survival. These studies found strong associations between higher average air pollution during follow-up and decreased survival after a cancer diagnosis. In this case study, we will first use the approach from this previous work that creates a misalignment between assigned exposure at time zero, where air pollution is averaged across follow-up and treated as time-fixed in a Cox model (naïve approach). We then estimate this same association using the

proposed discrete-time hazards approach (discrete-time approach). This is an established method for estimating the causal effect of a time-varying exposure in a time-to-event framework that upholds the principle of alignment at time zero outlined by the target trial framework (154).

Study population

A retrospective, population-based cohort of lung cancer patients in California was created by linking data from the California Cancer Registry (CCR) with $PM_{2.5}$ (fine particulate matter air pollution) estimates using patient zip code at diagnosis. CCR is a statewide population-based cancer surveillance system that collects information on all cancers (except for non-melanoma skin cancer) diagnosed in California. Lung cancer cases diagnosed between 2000 and 2010 and registered by CCR were identified with International Classification of Disease of Oncology (ICD-0-3) site codes (C34.0 – C34.3, C34.8, C34.9). Cases were included in the study if they had a primary, histologically confirmed cancer of the lung or bronchus and had an address in California at the time of cancer diagnosis on record with CCR (N=123,706). Patients with in situ cancer (N=48), diagnosed at autopsy (N=355), or with an invalid date of diagnosis or follow-up (N=2,035) were excluded from analysis. For this analysis, all cancer cases that survived less than 30 days from date of diagnosis were excluded (N=15,625). The initial study population included 105,643 patients. This study was reviewed and approved by Institutional Review Boards (IRBs) at San Diego State University, the University of California San Diego, and the California Department of Public Health Committee for the Protection of Human Subjects.

Exposure

Daily zip code specific concentrations of fine air particulate matter ($PM_{2.5}$) were estimated from 24-hour daily means sampled and reported by the US Environmental Protection Agency Air

Quality System (AQS) in California. Measurements from the three nearest stations within 25 km of each population-weighted zip code centroid were assigned to each zip code week using inverse distance weighting as applied previously (155). The measured concentration was weighted by the squared inverse distance to each zip code centroid, giving more weight to values at stations closer to the centroid. In this analysis, these estimated daily values were averaged as monthly zip code $PM_{2.5}$ values for each lung cancer patient. Patients with less than two daily values in any month over a five-year follow-up were excluded from the analysis (N=16,605).

Outcome

Survival time was calculated from the date of lung cancer diagnosis to the date of death from any cause. All patients were censored at five years after diagnosis to avoid heterogeneity of effect associated with long-term survivorship (139).

Covariates

The following covariates were considered as potential confounders of the association between air pollution and lung cancer survival either directly or through unmeasured variables: age at diagnosis (years), sex (male/female), marital status (single, partnered), race/ethnicity (non-Hispanic white, non-Hispanic black, Hispanic, Asian American, Native Hawaiian or Pacific Islander, Other/Unknown), treatment received during the first six months after diagnosis (surgery, radiation, chemotherapy), histology, month of diagnosis, year of diagnosis, and socioeconomic status quintile. Socioeconomic status was measured using a composite residential neighborhood-level index that combined census measures of education, income, occupation, and cost of living at the census block group treated as quintiles (156). Standard histology groupings were created using

ICD-0-3 morphology codes for carcinoma, squamous cell, adenocarcinoma, small cell, large cell, and other carcinomas.

Statistical Analysis

Descriptive statistics were calculated for survival, $PM_{2.5}$ exposure, and other covariates. Two approaches were used to estimate the impact of air pollution on survival after a lung cancer diagnosis. The first approach (naïve approach) emulates previous literature where $PM_{2.5}$ exposure is averaged over follow-up to create a single value representing all exposure during the follow-up period. This average exposure is then treated as time-fixed in a Cox proportional hazards model to estimate a single hazard ratio for the effect of $PM_{2.5}$ on survival over the entire follow-up period

The second approach (discrete-time approach) treats $PM_{2.5}$ as time-varying by using monthly averages. These monthly averages are then included in a discrete-time hazards model (i.e., a pooled generalized linear model with relatively short periods), where outcomes are assessed at each month of follow-up (157). This model estimates at each month for each person, the conditional probability of remaining free of the outcome given exposure, baseline covariates and time of follow-up. In this approach, the discrete-time hazard during each month *t* is defined as the risk of the outcome during month *t* among those who reached month *t* free of the outcome. This ensures that exposure is assigned at each outcome-time (month), which guarantees comparable lengths of exposure for those with and without the outcome at each month, ensuring alignment between treatment assignment and time-zero (for each case month distinct time zero in the context of this time varying discrete-time approach). The coefficient estimated from a generalized linear model with a complementary log-log link function will approximate those of a proportional hazards model (158). Time-varying hazards were accounted for by modeling follow-up time using cubic splines. Bootstrap confidence intervals were used to account for correlated observations. All

models for both approaches included potential confounding variables that were associated with both $PM_{2.5}$ exposure and lung cancer survival described above and include a random intercept on zip code at diagnosis to account for heterogeneity in survival across space. Models were stratified by stage at diagnosis (local, regional, distant) due to established differences in etiology, treatment, and survival. Overall models included patients with unknown stage at diagnosis.

All models for both approaches included potential confounding variables associated with $PM_{2.5}$ exposure and lung cancer survival described above and included a random intercept on zip code at diagnosis to account for heterogeneity in survival across space. Models were stratified by stage at diagnosis (local, regional, distant) due to established differences in etiology, treatment, and survival. Overall models included patients with unknown stage of disease.

To illustrate the difference in effect estimates over follow-up between both methods, we estimated adjusted survival curves using terciles of $PM_{2.5}$ exposure. For the naïve approach, we used the R package "*survminer*". In this package, a survival curve is plotted for each level of the grouping variable ($PM_{2.5}$ tercile), and the distribution is adjusted to the reference population using inverse probability weighting as described in the package documentation. For the discrete-time approach, we used the method described by Hernán (154). This procedure calculates conditional survival by multiplying the model's predicted values through time *t* to estimate conditional survival at *t* for all subjects. Survival is then predicted at time *t* for each subject, and conditional survival is averaged for all subjects under each exposure. This results in survival curves for each exposure level that are standardized to the distribution of the covariates in the study. All analyses were performed using R Version 4.0.2.

3.4 Results

The final study population included 89,038 lung cancer cases diagnosed in California between 2000 and 2010. Baseline characteristics of the study population are presented in **Table 3.1**. On average, lung cancer cases were 70.4 (SD 11.4) years old at diagnosis and predominantly non-Hispanic white (68%). More than half of the lung cancers were diagnosed at a late disease stage (55.6%). During the five-year study period (82%) of the population died, with the majority dying in the first year (53%). Median survival years for localized, regional, and distant stages at diagnosis were 4.92, 1.69, and 0.53 years, respectively. Average $PM_{2.5}$ exposure was 12.4 $\mu g/m^3$ across all lung cancer cases.

Table 3.1 Baseline characteristics of lung cancer patients, California Cancer Registry, 2000 - 2010

	Localized	Regional	Distant	Unknown
	n=14,304	n=19,294	n=49,525	n=5,915
PM (mean (SD))	12.6 (3.8)	12.9 (4.2)	13.3 (4.7)	14.1 (4.8)
Age (mean (SD))	69.9 (11.3)	68.7 (11.0)	67.9 (11.7)	75.1 (11.7)
Male (%)	6436 (45.0)	9740 (50.5)	25795 (52.1)	2809 (47.5)
Years Follow Up (mean (SD))	3.4 (1.8)	2.4 (1.9)	1.0 (1.2)	1.2 (1.4)
Histology (%)				
Adenocarcinoma	6904 (49.3)	7376 (38.8)	19252 (40.0)	961 (25.7)
Squamous Cell	3040 (21.7)	4679 (24.6)	6486 (13.5)	579 (15.5)
Other	2438 (17.4)	3356 (17.7)	9569 (19.9)	631 (16.9)
Large Cell	962 (6.9)	1401 (7.4)	5259 (10.9)	1140 (30.4)
Small Cell	670 (4.8)	2183 (11.5)	7615 (15.8)	433 (11.6)
Race/Ethnicity (%)				
NH White	10504 (73.6)	13495 (70.0)	32209 (65.1)	4178 (70.9)
Asian/PI	1489 (10.4)	2205 (11.4)	6874 (13.9)	631 (10.7)
Hispanic	1288 (9.0)	1849 (9.6)	5756 (11.6)	626 (10.6)
NH Black	952 (6.7)	1641 (8.5)	4473 (9.0)	436 (7.4)
American Indian	46 (0.3)	78 (0.4)	147 (0.3)	24 (0.4)
Partnered (%)	7696 (54.8)	10455 (55.3)	26013 (53.9)	2306 (40.7)
Radiation (%)	2222 (15.5)	8079 (41.9)	20950 (42.3)	738 (13.2)
Surgery (%)	9649 (67.5)	8160 (42.3)	2407 (4.9)	180 (3.2)
Chemotherapy (%)	1833 (12.9)	9405 (49.7)	26298 (54.2)	954 (18.1)
SES Quintile (%)				
Low	1910 (13.4)	2891 (15.0)	8113 (16.4)	1003 (17.0)
Low Med	2699 (18.9)	3751 (19.4)	9741 (19.7)	1373 (23.2)
Med	3091 (21.6)	4205 (21.8)	10770 (21.7)	1346 (22.8)
Med High	3199 (22.4)	4242 (22.0)	10878 (22.0)	1226 (20.7)
High	3405 (23.8)	4205 (21.8)	10023 (20.2)	967 (16.3)

Over the study period, there were long-term downward trends in $PM_{2.5}$ (Supplementary Material **Figure 3.1)**. **Table 3.2** shows the estimated hazard ratios for five-year survival for both the naïve approach and discrete-time approaches.

Table 3.2 Estimated five-year cancer-free survival hazard ratios and odds ratios for a 5 μg/m^3 increase in $PM_{2.5}$ exposure over the follow-up period, California Cancer Registry, 2000–2010

	Unadjusted		**Adjusted** [a]	
Naïve	**HR**	**95% CI**	**HR**	**95% CI**
Local	1.43	1.38 - 1.48	1.35	1.30 - 1.41
Regional	1.35	1.32 - 1.38	1.29	1.25 - 1.33
Distant	1.35	1.32 - 1.38	1.17	1.15 - 1.18
Overall	1.28	1.27 - 1.3	1.35	1.33 - 1.37
Discrete Time[b]	**OR**	**95% CI**	**OR**	**95% CI**
Local	1.03	1.01 - 1.05	1.01	0.99 - 1.03
Regional	1.03	1.01 - 1.04	1.00	0.98 - 1.01
Distant	1.02	1.01 - 1.03	1.00	0.99 - 1.01
Overall	1.02	1.02 - 1.03	1.00	0.99 - 1.01

[a]Adjusted for age at diagnosis (years), sex (male/female), marital status (single, partnered), race/ethnicity (non-Hispanic white, non-Hispanic black, Hispanic, Asian American, Native Hawaiian or Pacific Islander, Other/Unknown), treatment received during the first six months after diagnosis (surgery, radiation, chemotherapy), histology, month of diagnosis, year of diagnosis, and socioeconomic status quintile. Estimates for overall (all stages) additionally adjusted for stage at diagnosis (localized, regional, distant, unknown).

[b]Odds Ratios are for a five-unit increase in monthly $PM_{2.5}$ over follow-up. Confidence intervals are obtained using bootstrap estimation with 1000 replications using the percentile method.

In the naïve approach, the HRs for all-cause mortality associated with a 5 μg/m^3 increase in average $PM_{2.5}$ were 1.35 (95% CI: 1.30-1.41), 1.29 (95% CI: 1.25-1.33), and 1.17 (95% CI: 1.15-1.18) for localized, regional, and distant stage of disease, respectively. By contrast, under the

discrete-time approach, the estimated pooled ORs were 1.01 (95% CI: 0.99-1.01), 1.00 (95% CI: 0.98-1.01), and 1.00 (95% CI: 0.99-1.01). for localized, regional, and distant stage of disease, respectively. Adjusted survival curves for both the naïve and discrete-time approaches are shown in **Figures 3.1 and 3.2.**

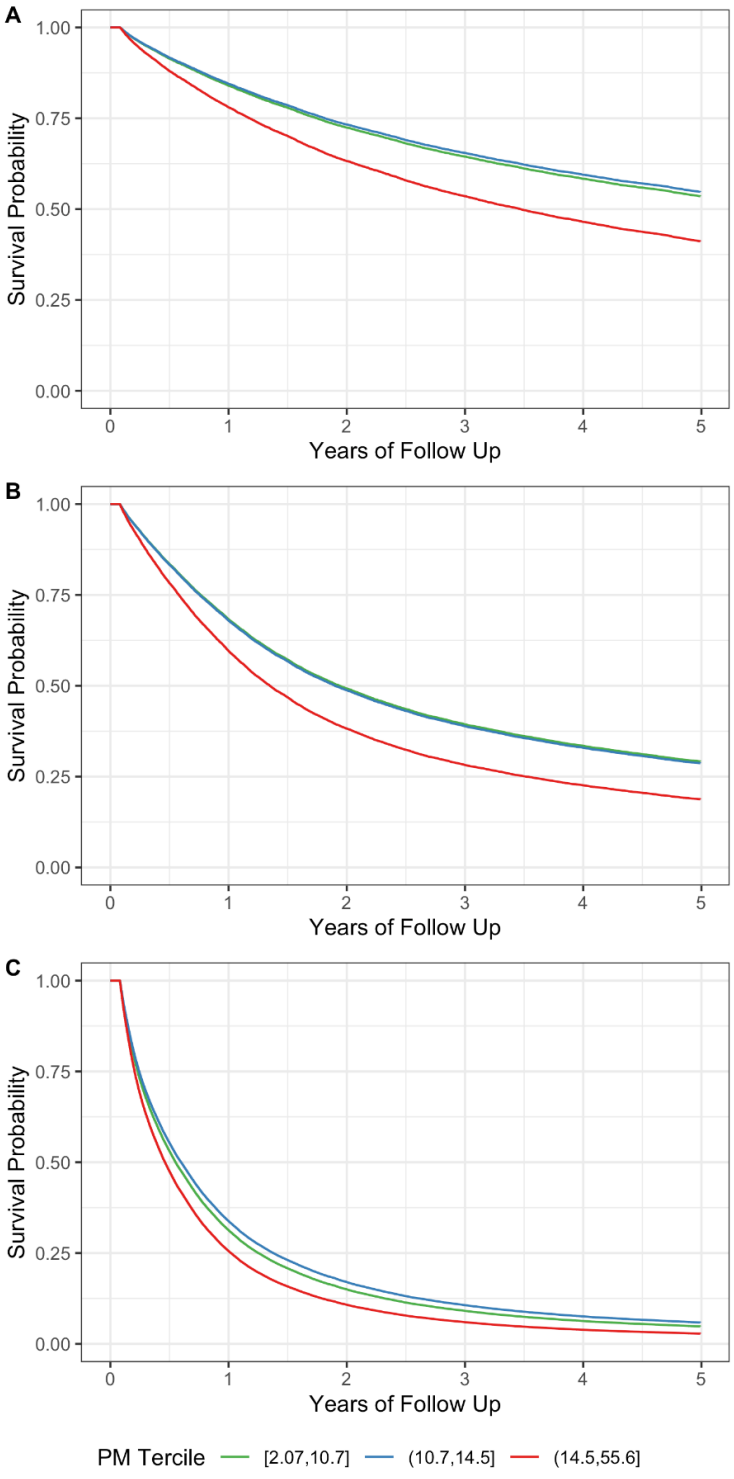

Figure 3.1 Adjusted survival curves from the naïve analysis using average $PM_{2.5}$ exposure over follow-up for a) Localized, b) Regional, and C) Distant stage at disease.

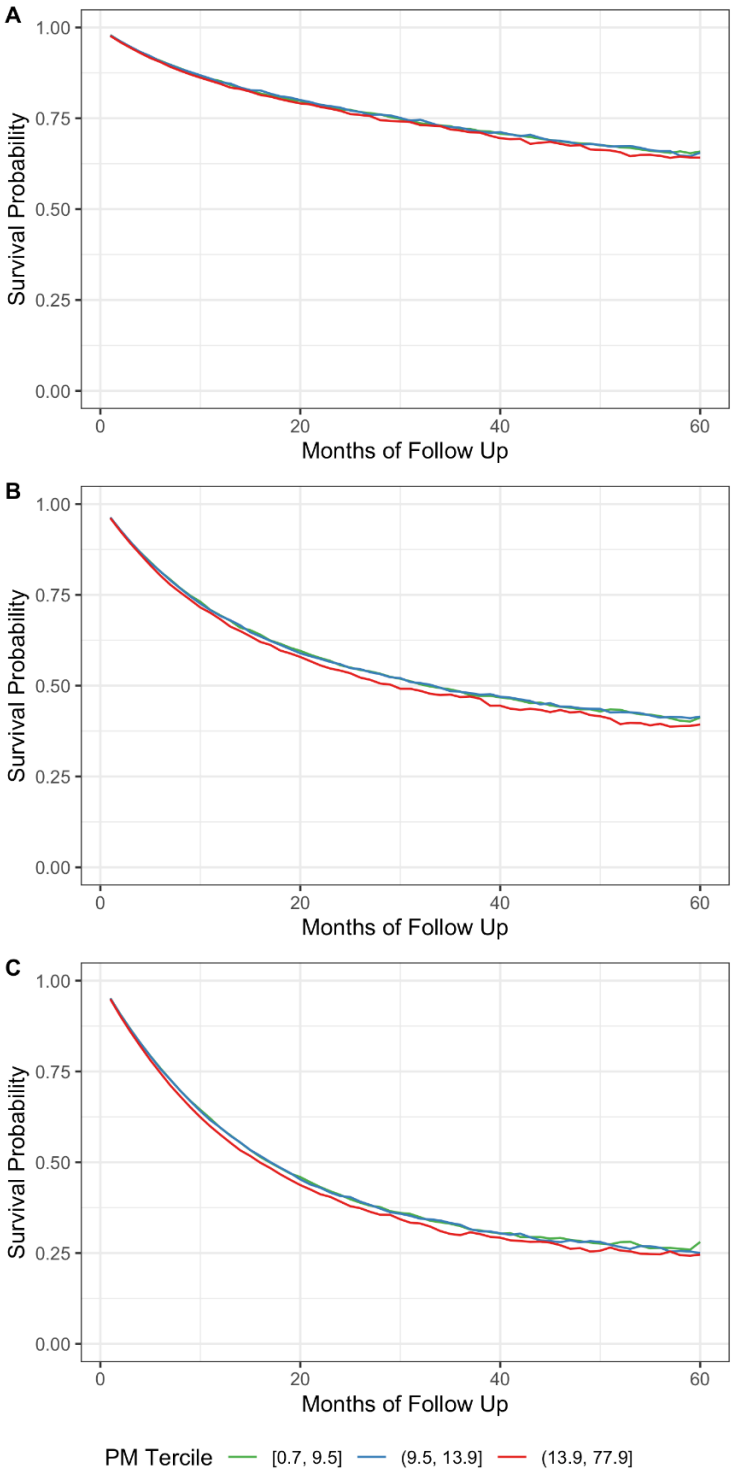

Figure 3.2 Adjusted survival curves from the discrete-time analysis using monthly $PM_{2.5}$ exposure over follow-up for a) Localized, b) Regional, and C) Distant stage at disease.

3.5 Discussion

While ITB is most often discussed in the context of clinical and pharmacoepidemiology, the errors that introduce systematic bias as a result of a failure to emulate a target trial correctly can occur in environmental epidemiology as well. In this case study, we used the target trial framework to identify that averaging an environmental exposure over follow-up in a time-to-event setting can introduce ITB into the study population and lead to substantial differences in the estimated effects and conclusions in the inference of interest. When an exposure is averaged over follow-up, treatment assignment is made using post-baseline information, which unhitches treatment assignment from time zero (143, 146). We illustrate that when ITB is introduced into a study in this way, the use of a time-to-event model does not adequately control for differences in follow-up lengths (a well-known function of these models). Instead, it is necessary to use alternative approaches that ensure alignment at time zero.

In this paper, we proposed an analytic solution using a discrete-time hazards approach. This method ensures alignment between time-varying treatment and time zero at each outcome time because the hazard is re-assessed at each follow-up period, including updating information on any time-varying covariates or exposures. This guarantees that only comparable exposure periods (concerning time) are being used to assess the hazard at each outcome time. In a recent paper that outlines the use of causal diagrams for ITB, Mansournia et al. highlight how ITB can be considered as either misclassification or selection bias and then propose strategies to avoid ITB, including the discrete-time approach we use here. They make the helpful comparison of the discrete-time approach to a series of mini-randomized trials at monthly intervals (159). In this comparison, individuals have their exposure value re-assigned in each mini-randomized trial (at each month), preventing ITB.

Here we chose to illustrate the discrete-time hazard approach because it allows for time-varying hazards and hazard ratios, avoiding common pitfalls in reporting causal effects using an average HR to represent risk over entire follow-up periods. Using an average HR can be uninformative because of time-varying HRs, which can occur due to built-in selection bias. The average HR ignores the distribution of events over follow-up, which guarantees that the magnitude of the HR will depend on the length of follow-up when it is time-varying. Hernán recommends using discrete-time hazards models to obtain adjusted survival curves to present HRs over time to overcome the misleading representation of a single HR that depends on length of follow-up (154). Furthermore, the discrete-time approach can be easily integrated into more advanced methods, including g-computation or inverse probability weighting methods (160, 161). These approaches could be beneficial when dealing with complex time-varying exposure and confounding settings that may cause exposure-confounding feedback (i.e., collider stratification bias) as described elsewhere (162, 163). Additionally, there are other potential analytic approaches that can be used to assess the impact of time-varying exposure in a time-to-event framework while maintaining alignment at time zero. Previous studies have appropriately treated such exposures in time-varying Cox proportional hazards models or by including only comparable exposure lengths (i.e. one year before baseline) that are consistent across all individuals, regardless of follow-up length (164-166).

After applying this discrete-time approach to our case study with lung cancer cases in California, we find little evidence that $PM_{2.5}$ exposure after a cancer diagnosis impacts survival. By contrast, in the naïve approach that uses an average exposure over follow-up, we find estimates in magnitude to results from the emulated studies (70, 73). In a recent study that examined the impacts of average air pollution exposure over follow up on survival in California, study authors found the HRs for all-cause mortality associated with a 5 $\mu g/m^3$ increase in average $PM_{2.5}$ were

1.38 (1.35-1.41), 1.26 (1.24 – 1.28), and 1.10 (1.09 to 1.11), for localized, regional, and distant stage at disease, respectively. We propose that ITB likely drives these effect estimates. Of note, the apparent effect modification by stage at diagnosis seen here and in other previous studies that use the naïve approach (74) is no longer apparent after using the discrete-time approach. We hypothesize that this effect heterogeneity is likely driven by the larger differences in survival time among those with localized stage of disease relative to those with distant stage of disease. These comparatively larger differences in follow-up time then exacerbate the impact of ITB.

This case study demonstrates that ITB in this particular setting is substantial and partially attributable to the strong downward trend in $PM_{2.5}$ over the study period (152, 153). Because of this temporal trend, the averaged exposure level was informed not only by the number of observations available (length of follow-up) but also by the decreasing trend. This means that individuals with longer follow-up times had systematically lower averages than those with shorter follow-up times, creating a noticeable difference in average exposure. Researchers in environmental epidemiology are aware of the strong temporal trends in most environmental exposures and often control for season and year in the analysis to adjust for this trend (70). Unfortunately, controlling for year of diagnosis using any strategies (stratification, indicator variable) will not control for temporal trends in the exposure or prevent ITB when done in the presence of an averaged exposure over follow-up.

While we focused on a specific example in $PM_{2.5}$ exposure and lung cancer survival, ITB can occur in any observational setting. For example, a recent paper that examined the impact of air pollution on mortality in a Medicare population used yearly average air pollution exposure from date of Medicare enrollment to date of death or last follow-up in Cox proportional hazards models (167). Another recent study examined the impact of air pollution and temperature on COVID-19

case fatality (168). They used an average exposure from date of diagnosis through date of death or recovery. In both of these examples, air pollution exposure was assigned using attained length of follow-up. Because of this, these analyses may be vulnerable to ITB through the same mechanism as described in this case study. Because treatment assignment is not defined at time zero inherently in observational studies as in RCTs, epidemiologists need to explicitly ensure this alignment to avoid these systematic errors.

There are several limitations in this work worth considering. First, we did not evaluate the conditions under which ITB would be introduced in environmental epidemiology studies. ITB in environmental epidemiology may be subject to different considerations than its "classical" counterpart. Notably, ITB is most often identified in studies with a discrete exposure value, where subjects are considered either unexposed or exposed (targeting a traditional randomized trial with two arms and one single randomization at baseline). In air pollution and other environmental exposures, exposure is often measured continuously and is ubiquitous, which means no one is truly unexposed. In these settings, the target trial corresponds to multiple randomizations at different times, as in crossover or stepped wedge designs. In this case study, the bias was considerable because of the strong temporal trend in $PM_{2.5}$ over the study period and the significant differences between follow-up in lung cancer cases. While this bias may be minimized in other settings where the temporal trend is not as strong, it is still necessary to use approaches that ensure alignment at (multiple) time zero under the target trial framework to avoid systematic errors.

There are additional limitations in the case study itself. Air pollution exposure was assigned using interpolated estimates from ground-level fixed-site monitors, which are not evenly spaced across California. This may have resulted in differential precision of exposure measurement by region, specifically in rural areas with fewer monitors. Because air pollution was assigned using

zip code at diagnosis, we do not have information on individual variability in air pollution exposure from place of work or other factors that influence daily exposure. CCR does not include information on potentially important confounding variables, including smoking, alcohol use, access to care (although insurance status can be an appropriate proxy), or treatment received more than six months after diagnosis.

In this study, we identified that averaging a time-varying environmental exposure over follow-up in a time-to-event context creates a misalignment between treatment assignment and start of study follow-up, introducing ITB. The bias resulting from this misalignment can be further exacerbated in the presence of important time trends commonly found in environmental exposures such as air pollution. In this context, we recommend treating air pollution as time-varying in a model that ensures alignment between the start of study follow-up and treatment assignment. Future studies should evaluate under what conditions ITB in this setting is minimized or exacerbated.

Supplementary Material

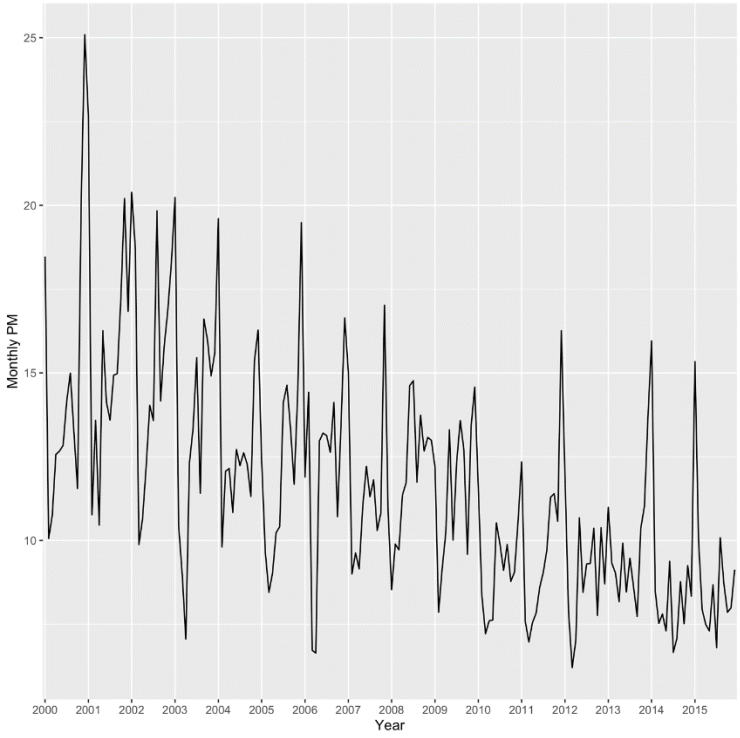

Figure 3.3 Fine particulate matter trends in California, 2000 – 2015

Chapter 3, in full, has been submitted for publication of the material as it may appear in Epidemiology. Sheridan, Paige; Chen, Chen; Thompson, Caroline; Benmarhnia, Tarik. The dissertation author was the primary investigator and author of this paper.

4 Methodological considerations for measuring racial residential segregation as a determinant of lung cancer survival

4.1 Abstract

Background: Several studies have attempted to quantify the relationship between racial residential segregation and cancer outcomes but are limited by methodological challenges in measuring segregation. There are two primary considerations for measuring segregation using an index: the spatial scale and spatial relationships. This study assesses how these considerations impact the absolute estimation of segregation and subsequent conclusions in lung cancer survival.

Methods: Lung cancer cases from the California Cancer Registry from 2000-2010 were linked with segregation measures calculated using American Community Survey data. Segregation was captured using the dissimilarity index. The index was calculated in eight ways to examine the relative contribution of considering the spatial scale and spatial relationships: using census tract vs. census block, five vs. ten-mile regions, and using a spatial vs. aspatial measure of segregation.

Results: Segregation among white lung cancer patients was generally associated with protective or null effects on survival, while segregation among Black cancer patients was generally harmful. Both considerations for calculating dissimilarity influenced the absolute value of segregation and qualitatively impacted the effect estimates for the association between segregation and five-year survival.

Conclusion: Explicit consideration should be made for the impact of spatial scale and spatial relationships when measuring segregation. Reliable and accurate estimation of segregation is necessary to understand how racial residential segregation influences cancer outcomes and disparities.

4.2 Introduction

Racial residential segregation is a well-established determinant of racial health disparities. Generally, racial residential segregation refers to the uneven distribution between racial and ethnic groups across urban space, a fundamental characteristic of cities (169, 170). Importantly, racially segregated neighborhoods do not occur organically or by resident choice but instead are the physical manifestation of individual and structural racism that has systematically (over multiple decades) denied equal opportunities to communities of color (78, 171). Segregation is associated with restricted access to quality education, employment, equal income, safe streets, clean air, and quality health care (100, 172-180). Racially isolated neighborhoods are more likely to have primary care shortages, offer fewer ambulatory facilities, and have a lower supply of surgeons (82, 181). These factors are well-known social determinants of health that result in disparities across the life course (182). While policies were enacted in the 1960s to increase integration and promote equity, many cities in the US remain deeply segregated.

Over the past decade, there has been growing interest in the influence of neighborhood environment on racial disparities in lung cancer outcomes, including racial residential segregation (183-185). Despite improved lung cancer survival over the last several decades, racial disparities in survival have remained persistent (100, 176, 186, 187). Non-Hispanic (NH) Black patients have the highest death rate and shortest survival of any racial/ethnic group, with only 16% surviving to five years, compared to 19% of white patients (1). Disparities in lung cancer outcomes may be explained by disparities in access to quality health care, socioeconomic factors, and physician-patient relationships (188). For example, NH Black patients are less likely to receive recommended care for their cancer than their white counterparts (132, 189). Studies have demonstrated that while insurance status partially contributes to these disparities, it is not the only driver (176, 190).

Several studies have attempted to quantify the relationship between racial residential segregation and lung cancer outcomes (101-103, 185, 191). However, the literature has not yet addressed methodological inconsistencies in the measurement of segregation. Segregation is notoriously challenging to measure, partially because it is a complex, multidimensional and spatial process (93). It is generally agreed upon that segregation involves multiple groups and that segregation measures must capture how these different population groups are distributed across space (192). Because of this, segregation indices capture information on two geographic scales: subareas (e.g., neighborhoods, census tracts) within larger geographic areas (e.g., cities, counties). Similarly, segregation indices generally capture one or more dimensions of segregation: most often evenness (differential distribution of groups across space) and isolation (potential interaction between groups) (75, 193). To ensure interpretability and consistency in the literature, it is necessary to consider two factors in calculating a segregation index: the spatial scale of geographies and spatial relationships of populations.

In order to calculate a segregation index, it is first necessary to define the spatial scale of both the subareas and the larger geographic area. While these geographic scales are often chosen arbitrarily or for convenience, the size of both spatial scales can impact the absolute measurement of segregation. It is well established that when administrative boundaries are used to delineate neighborhoods to calculate segregation, the value of the index can change depending on the size of the areal units chosen (110). For example, segregation levels vary between counties and metropolitan areas partially because of their different sizes (194). This size-dependency is known as the modifiable areal unit problem (MAUP) (114), which occurs when artificially delineated areal units (like zip codes or census tracts) are used to analyze geographically continuous phenomena.

After defining the spatial scale for analysis, it is critical to use meaningful segregation indices that consider spatial relationships of populations (94). Common segregation measures treat neighborhoods as discrete units with exogenously assigned boundaries as impenetrable walls to interaction (112). Because of this feature, these measures cannot differentiate between population distribution patterns, instead only evaluating population mix within a spatial unit (113). This is also known as the checkerboard landscape problem (114), where a gridded area with alternating exclusively white and black areas will not be evaluated as any less segregated than when the same areas are rearranged into larger areas. To overcome this limitation, spatial segregation measures that explicitly consider this spatial element have been proposed (94, 112, 114, 115). These measures incorporate an adjacency component to account for the fact that there is more interaction between areas in closer proximity, which then decreases as distance increases. Previous studies have demonstrated that when segregation indices that account for the spatial orientation of neighborhoods are used, the absolute estimation of segregation changes. (116).

To date, there has been no comprehensive assessment of how these two factors impact the absolute measurement of segregation or subsequent conclusions in cancer epidemiology. In this study, we extend the existing literature by evaluating the impact of racial residential segregation on lung cancer survival in California while explicitly considering the impact of both spatial scale and spatial relationships. We use the dissimilarity index, which captures the evenness of population groups across space (75). This index is one of the most widely used, partially due to its relative ease of calculation and interpretation. We will compare three considerations under the two factors outlined above. To evaluate the impact of spatial scale, we use two different areas for subareas of the analysis (census tract vs. census block) and two different areas for the larger geographic area over which segregation is defined (five vs. ten-mile neighborhoods). To evaluate the impact of

using indices that consider spatial relationships, we use both an aspatial and spatial dissimilarity index.

4.3 Methods

Study Population

A retrospective, population-based cohort of cancer patients in California was created by linking the California Cancer Registry (CCR) with residential segregation data from the American Community Survey using the zip code of the patient's residence at diagnosis. CCR is a statewide population-based cancer surveillance system that collects information on all cancers (except for non-melanoma skin cancer) diagnosed in California. Lung cancer cases were identified with the International Classification of Disease of Oncology (ICD-0-3) site codes (C34.0 – C34.3, C34.8, C34.9) diagnosed between 2000 and 2010 and registered by CCR. Cases were included if they have a primary, histologically confirmed cancer of the lung or bronchus and had an address in California at the time of cancer diagnosis on record with CCR (n=123,303). In-situ cancers, non-carcinoma histology, and diagnoses at autopsy were not included in the study sample. Patients with incomplete diagnosis dates, date last follow-up, or zero days of follow-up time were excluded (n=2,035). Because this study aimed to examine racial disparities in lung cancer outcomes between Black and white populations, only NH white and NH Black lung cancer patients were included in the final analysis (n=95,159). While there is substantial racial segregation in Hispanic and Asian communities in California, these populations have a lower cancer burden and do not have significant survival disparities (195). This study was reviewed and approved by Institutional Review Boards (IRBs) at San Diego State University, the University of California San Diego, and the California Department of Public Health Committee for the Protection of Human Subjects.

Exposure

The exposure of interest in this study was racial residential segregation measured using the dissimilarity index. Conceptually, the dissimilarity index measures the percentage of a group's population that would have to change residence for each neighborhood to have the same distribution as the metropolitan area overall. As described above, this contrast is achieved by using information on two geographic scales: subareas within larger geographic areas. The index ranges from zero (complete integration) to one (complete segregation). To illustrate the impact of considerations for measuring segregation on lung cancer survival, we calculated a dissimilarity index using eight combinations of the three considerations of interest: census tract vs. census block, five vs. ten-mile neighborhood, and aspatial vs. spatial. The patient's census tract at residence was used to assign each lung cancer case their segregation values.

Race and ethnicity population data at the census block and census tract level was obtained using the 2010 American Community Survey (ACS) five-year detailed tables (196). The R package 'seg' was used to calculate the eight dissimilarity indices. This package takes an input of the larger geographic area of interest (five or ten-mile neighborhood), divided into smaller subareas (census tract or block). In order to create this input, we first calculated the geographic centroid for all census tracts in the study area. We then calculated a five- and ten-mile diameter buffer from this centroid to create neighborhoods surrounding each census tract. We then used all subareas (tract or block) within this neighborhood. This created a single dissimilarity index for each census tract, where every tract has its own five- or ten-mile diameter neighborhood. For each census tract and census block, both an aspatial and spatial dissimilarity index was calculated, using both a five and ten-mile neighborhood, for a total of eight dissimilarity measures. A flow chart

describing how the dissimilarity indices were calculated is shown in Supplementary Material **Figure 4.4.**

Outcome

Survival time was calculated from the date of lung cancer diagnosis to the date of death from any cause. All non-deceased patients were censored after five years of follow-up to avoid heterogeneity of effect in long-term survivorship (139).

Statistical Analysis

Aspatial dissimilarity was calculated using the method proposed by Massey and Denton (75):

$$D = \frac{1}{2}\sum_{i=1}^{n} \left| \frac{w_i}{W_T} - \frac{b_i}{B_T} \right|$$

Where n represents the total number of subgroups (tracts, blocks) in a region, w_i represents the white population size in subgroup i, W_T represents the white population size in the region, b_i represents the Black population size in subgroup i, and B_T represents the Black population size in the region.

The spatial dissimilarity index was calculated using the method proposed by Reardon and O'Sullivan (94). This method modifies the composite population counts used in the aspatial dissimilarity index by applying a spatial kernel centering at the reference unit. Here, the kernel center has a larger weight than farther locations (197). This approach accounts for spatial autocorrelation between adjacent spatial units by creating modified composite population counts that are weighted population counts of neighboring units (198).

Cox proportional hazards models were used to estimate the effect (summarized using Hazard Ratios, HR) of all eight dissimilarity measures on five-year lung cancer survival. Each dissimilarity index was included in the model as a quartile indicator, where the contrast of interest was the highest quartile of dissimilarity compared with the lowest. Models were stratified by stage at diagnosis (local, regional, distant) and race (NH Black and NH white) due to hypothesized effect heterogeneity. All models were adjusted for potential confounders, including age at diagnosis (years), sex (male/female), and year of diagnosis. Socioeconomic status (SES) was measured using a composite residential neighborhood-level index that combined census measures of education, income, occupation, and cost of living at the census block group treated as quintiles (156). We explored whether SES was a potential effect modifier of the relationship between residential segregation and lung cancer survival. We included interaction terms to evaluate the presence of multiplicative interactions, and none were identified. All analyses were performed using R Version 4.0.2.

4.4 Results

Our final study population included 95,159 lung cancer patients. Of these, less than 10% identified as NH Black (n=9,197), and 90% identified as NH white (n=85,962). The average age at diagnosis was 70, and over half of all cases were diagnosed at distant stage of disease (n=54,912). The median survival time among cases with localized, regional, and distant stage at disease was 4.3, 1.4, and 0.3 years, respectively. **Table 4.1** summarizes baseline patient characteristics for the study population.

Table 4.1 Baseline characteristics of NH white and NH Black lung cancer patients, California Cancer Registry, 2000- 2010

	NH White	NH Black
	n=85,962	n=9,197
Age (mean (SD))	70.2 (11.2)	65.8 (11.5)
Male (%)	42180 (49.1)	4791 (52.1)
Years Follow Up (mean (SD))	1.5 (1.7)	1.4 (1.6)
Stage at Diagnosis (%)		
Distant	49174 (57.2)	5738 (62.4)
Localized	13057 (15.2)	1082 (11.8)
Regional	17566 (20.4)	1851 (20.1)
Unknown	6165 (7.2)	526 (5.7)
SES Quintile (%)		
Low	10471 (12.2)	3325 (36.2)
Low Med	17486 (20.3)	2454 (26.7)
Med	20013 (23.3)	1751 (19.0)
Med High	19729 (23.0)	1088 (11.8)
High	18263 (21.2)	579 (6.3)

Figures 4.1 and 4.2 present the HRs for the association between the eight dissimilarity measures and five-year lung cancer survival. Models are stratified by stage at diagnosis (local, regional, distant) and by race. First, we found that, while "white segregation" (**Figure 4.1**) was generally associated with protective (HR<1) or null effects, the opposite pattern was observed for "Black segregation" (**Figure 4.2**). When comparing the different stages at diagnosis, most of the variability across effect estimates was identified for localized or regional cancers, especially among NH white cancer patients. Overall, we did not identify an effect of segregation on survival when considering distant cancer stages.

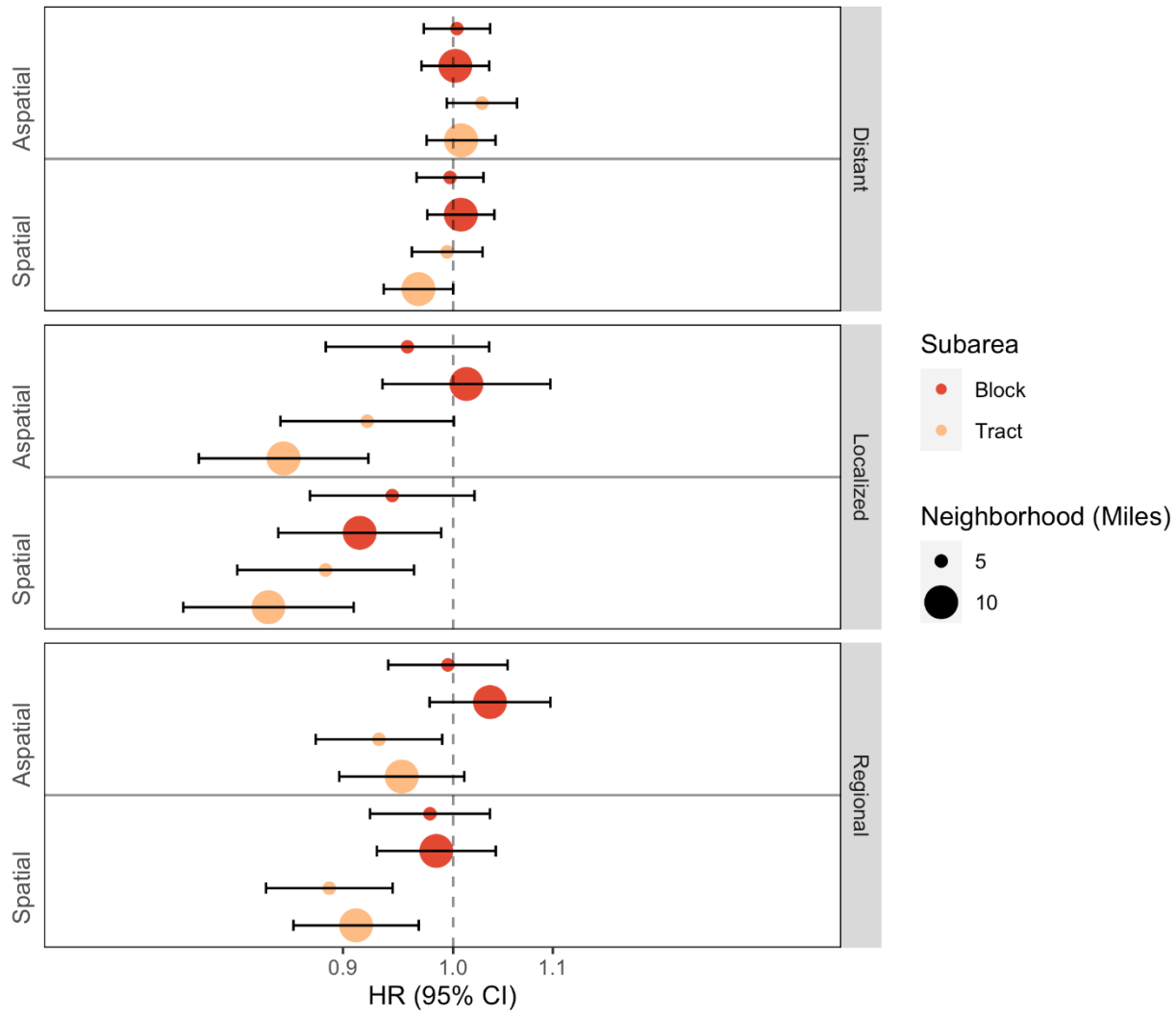

Figure 4.1 NH White Lung Cancer Cases: Hazard Ratios for highest quartile compared to the lowest quartile of segregation measured using dissimilarity index (8 ways), California Cancer Registry, 2000 – 2010.

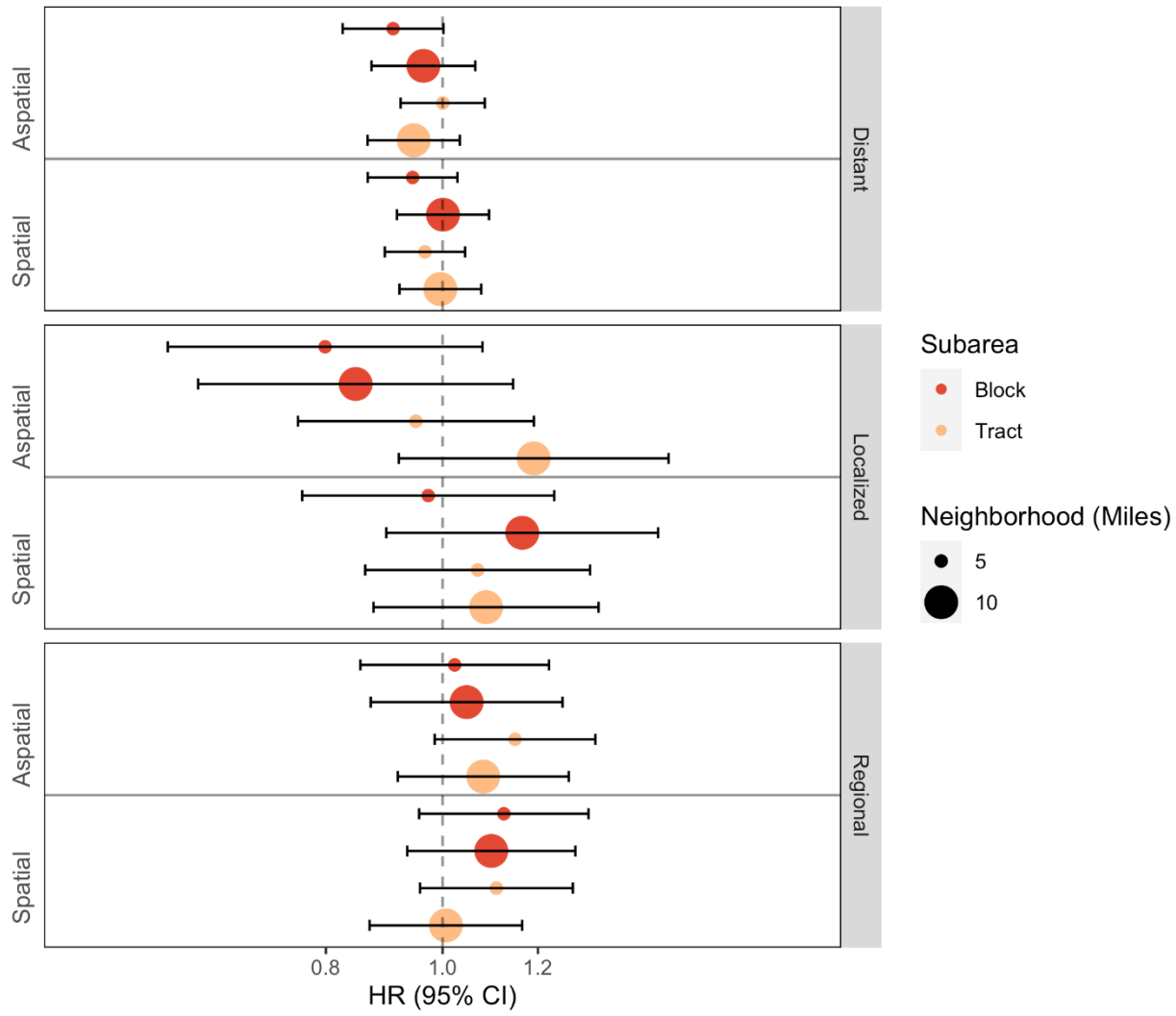

Figure 4.2 NH Black Lung Cancer Cases: Hazard Ratios for highest quartile compared to the lowest quartile of segregation measured using dissimilarity index, eight ways, California Cancer Registry, 2000 – 2010.

In general, each method of dissimilarity index calculation resulted in a different HR. However, some of the analysis decisions resulted in greater differences than others. When comparing dissimilarity indices calculated using census tract vs. census block as subareas of interest, we found that the HRs were varied enough that overall conclusions about the effect of segregation of lung cancer survival were impacted for NH white patients with localized and

regional stages of diagnosis. Among NH Black patients, the HRs varied most substantially between census tract and census block among those with localized stage at disease.

When comparing estimates with five vs. ten-mile neighborhoods, we found some heterogeneity across effect sizes but did not observe a systematic pattern. In some cases, a larger buffer was associated with a higher HR (aspatial measures at the census block with localized stages for NH Black in **Figure 4.2**), and in other cases, the opposite pattern was observed (aspatial measures at the census tract with regional stages for NH Black in **Figure 4.2**).

We found that aspatial vs. spatial segregation measures influenced both the amplitude and direction of the HRs. This spatial consideration was associated with substantial heterogeneity for regional and localized disease stages across NH white and NH Black patients. In some settings, the direction of HRs was entirely reversed when using spatial segregation measures (localized stages at the census block with a ten-mile buffer for NH Black patients in **Figure 4.2**).

Finally, to visualize the differences in the eight dissimilarity measures across space, we used an example in San Diego County (**Figure 4.3**). In San Diego, the segregation value was qualitatively lower when using the spatial dissimilarity measures than the aspatial ones. In a similar pattern, dissimilarity estimates were lower when using census tracts as the subunit of analysis compared to census block. This map also identifies that more heterogeneity is captured across space when using a five-mile buffer to create a neighborhood than with the ten-mile buffer neighborhood. Similar patterns were seen in the entire study area (California), with the most significant differences in average segregation value between aspatial (0.51) vs. spatial estimates (0.42) and census tract (0.46) vs. census block (0.56).

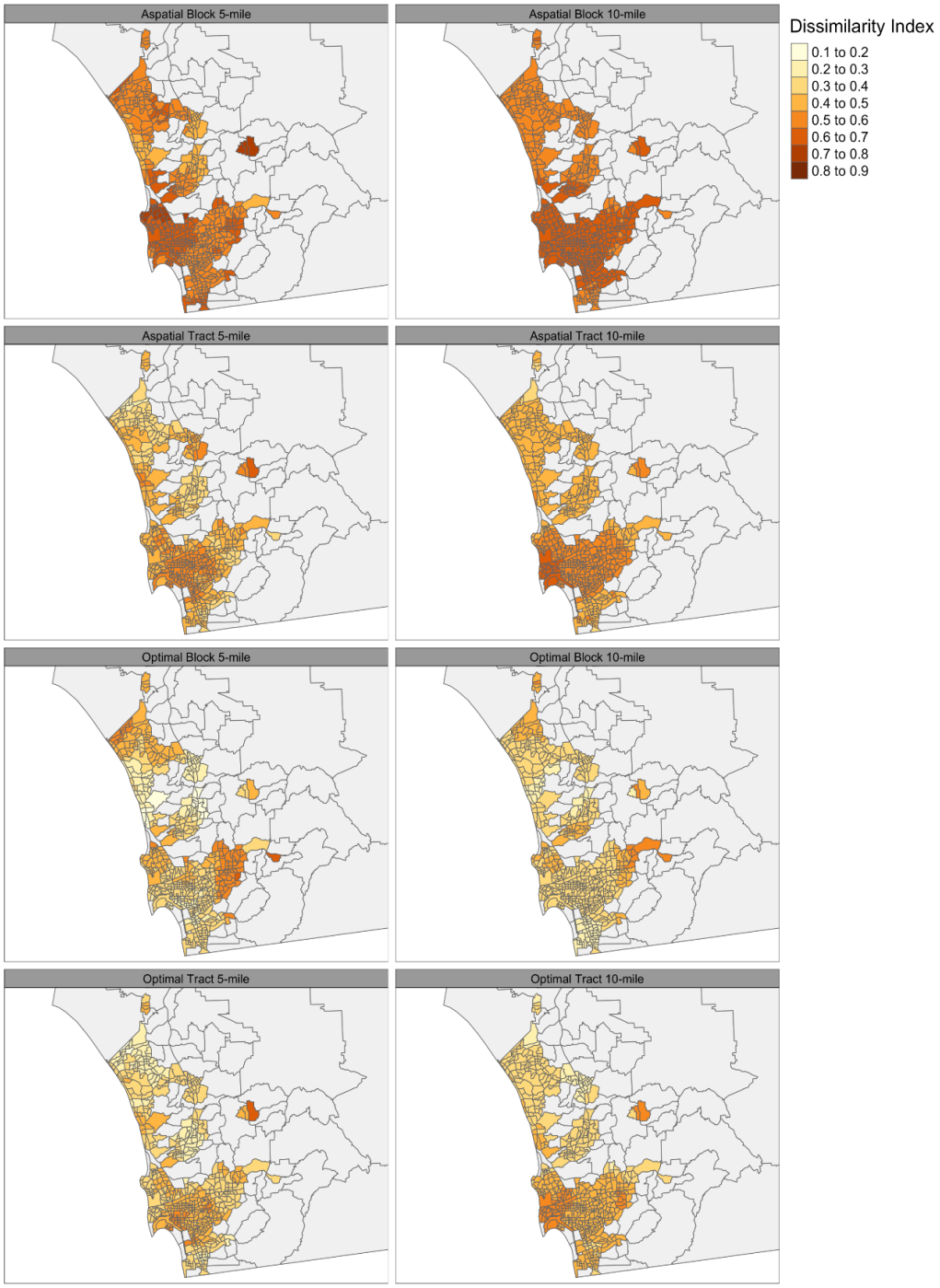

Figure 4.3 Segregation in San Diego, California, measured using the dissimilarity index, eight ways, California Cancer Registry, 2000 – 2010.

4.5 Discussion

The role of residential racial segregation as a determinant of cancer outcomes disparities has received specific attention in the last several years. However, this evidence has lacked discussion in the diversity of choices involved in measuring segregation and implications for epidemiological findings. This paper highlighted two essential considerations for calculating segregation indexes: the spatial scale of units for analysis and segregation measures that account for spatial relationships. We demonstrated that these factors impact the absolute value of segregation and qualitatively impact effect estimates for the association between segregation and five-year survival.

We found that segregation measures using the spatial dissimilarity index resulted in different values (as compared to aspatial measures). This difference highlights that ignoring spatial autocorrelation when estimating segregation measures may lead to considerable biases. For this reason, we recommend that researchers consider spatial dissimilarity indexes when measuring segregation to prevent spatial confounding, as recommended by others (199, 200). We also found differences in effect estimates when the spatial scale of analysis (both subareas and buffer size). Because the choice of the spatial scale may depend on the specific question of interest and what level of segregation one may want to capture, we strongly suggest to 1) explicitly motivate the choice of the spatial scale of analysis; 2) conduct sensitivity analyses with varying choices for the spatial scale (for both subareas and larger geographic area).

Previous studies, focusing on different outcomes, have similarly demonstrated that the use of spatial indices and the choice of spatial scale impacts effect estimates (201). Kramer et al. compared spatial segregation measures using population density surfaces and census tract using aspatial segregation measures and found that while they were highly correlated, the aspatial

measures may hide important differences by region and metropolitan size (202). One study examined income segregation to predict voter turnout and found that using a spatial measure of segregation (Moran's I) results in more income segregation (197). In one study that examined spatial and aspatial segregation measures and the relative impact of the choice of spatial scale for each, they found that aspatial measures yielded higher levels of segregation when smaller areal units were used. This scale effect was also present among spatial measures (113). Finally, studies have demonstrated that when segregation is evaluated based on smaller spatial units, the level of segregation tends to be higher than with larger units (203).

To date, only one study has examined the association between residential segregation and lung cancer treatment and survival among early-stage cancer patients (103). This study uses the aspatial isolation index to measure residential segregation, one of the most commonly used segregation measures. They found that among neighborhoods with high segregation and high economic deprivation, NH Black lung cancer patients with early-stage disease were more likely to die prematurely, even after adjusting for receipt of surgery. Among NH white patients, segregation had no impact on survival. In a related study, authors examined the association between segregation measured using the aspatial dissimilarity index and lung cancer mortality. They found that increasing segregation was associated with lung cancer mortality for Black patients and decreased mortality for white populations (101).

There are limitations in this work worth considering. The dissimilarity index itself is subject to several limitations. First, this measure can only measure the evenness of two groups at a time. This means that when there are multiple populations of interest, their estimated segregation must be calculated independently. A second consideration is that this index measures deviations from evenness and reflects randomness across space. This means that when unit sizes or

populations are small, this can lead to large values for the segregation index. When a population group is small, it is more likely to be unevenly distributed across space by change than a larger population size (204). This further emphasizes the importance of conducting various sensitivity analyses with varying spatial scales. Importantly, measuring residential segregation using the dissimilarity index only captures one dimension of segregation. It is likely that segregation effects on health result from multiple dimensions of segregation (192). In this work, we used aggregated administrative data to estimate segregation. This means that while our spatial segregation measures decrease the influence of these administrative boundaries, there is still constraint that would not be present using individual-level data. Finally, it is necessary to acknowledge that segregation is context-specific, and effect estimates from a given study may not be transportable to other contexts.

The purpose of this work was not to recommend one single way to capture segregation comprehensively. Instead, we highlight that the choices made for defining and measuring segregation can carry implications for the qualitative assessment of segregation and inference in epidemiologic studies. In line with previous work, we recommend that segregation studies be conducted with measures that consider spatial relationships and multiple scales of data to explore the scale robustness of the results (205). Performing an analysis at multiple scale levels can document the role of scale in modifying segregation. Further, we recommend that if a strong prior exists for the choice in the segregation measure and scale of analysis, this choice is outlined explicitly.

Reliable and accurate estimation of segregation is necessary to understand how racial residential segregation influences cancer outcomes and disparities. However, the current literature is limited by a lack of consideration for these factors that can influence the estimated value of

segregation. Because these choices influence both the absolute estimate of segregation and inference about the relationship with survival, it is necessary to consider these factors explicitly.

Supplementary Material

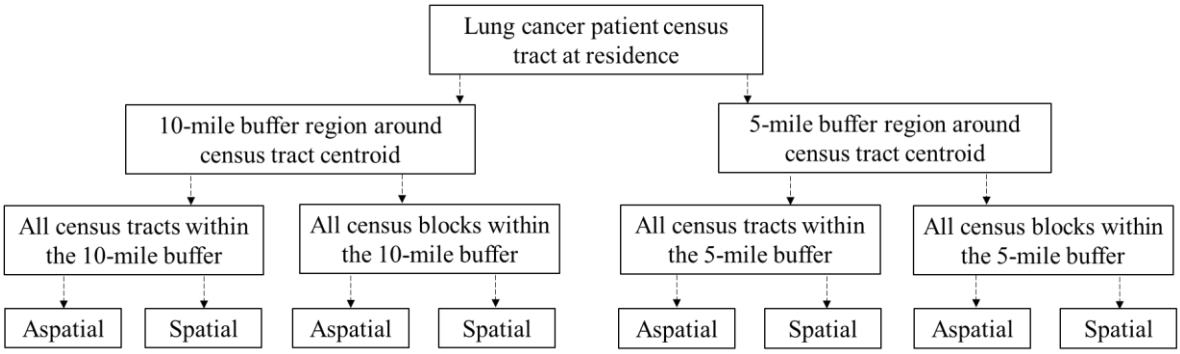

Figure 4.4 Flow chart depicting the calculation procedure for eight dissimilarity indices.

Chapter 4, in full, has been submitted for publication of the material as it may appear in Cancer Epidemiology Biomarkers and Prevention. The dissertation author was the primary investigator and author of this paper

5 Discussion

5.1 Summary of dissertation research

An interplay of factors, including an individual's physical and social environment, can influence health outcomes (206). Air pollution and racial residential segregation are two neighborhood factors that may explain heterogeneity in lung cancer survival across space and time. Because of this, they have received attention as potential risk factors that may contribute to adverse cancer outcomes (207-209). Several recent studies have examined how air pollution impacts cancer survival and found strong associations between air pollution and survival. Similarly, studies have examined the impact of racial residential segregation and cancer survival, with varying conclusions. While this evidence can aid in our understanding of contextual factors in lung cancer survival, methodological concerns have not been adequately addressed. Addressing these concerns will help shape our understanding of lung cancer etiology and potential interventions to improve cancer survival and target geographical disparities. Further, illustrating the methodological concerns and providing practical solutions will be valuable for applications in other research areas.

This dissertation aimed to evaluate the role of ambient air pollution exposure and racial residential segregation on lung cancer survival. This dissertation expands on the current literature by addressing methodological challenges in epidemiologic studies that evaluate neighborhood-level exposures. Neighborhood-level exposures are unique because they have both spatial and temporal components that need to be appropriately considered in an analysis. The first aim of this dissertation underscores the importance of evaluating patterns in lung cancer survival both spatially and temporally, as global assessments can miss important heterogeneity across space and time. This study demonstrated that there is substantial heterogeneity over space in California and that this heterogeneity changes across time. The spatial patterning in survival provides evidence

for the existence of neighborhood-level risk factors that may vary concomitantly. Further, this aim provides evidence that local policies may effectively improve lung cancer survival equity in California.

The second aim of this dissertation examined the impact of ambient air pollution after a lung cancer diagnosis on survival. Several existing papers had previously examined the impact of air pollution on cancer survival, but these studies consistently overlooked a methodological flaw that results in immortal time bias (ITB). This aim expands on previous work by identifying how this bias is introduced in this context and providing an accessible solution to assess the relationship between a time-varying exposure and time-to-event outcome, such as survival while avoiding ITB. The association between air pollution and lung cancer survival was estimated using the naïve approach seen in existing studies and compared to estimates under the proposed approach using discrete-time models. We illustrate that after using the discrete-time approach, there is little evidence of an effect of air pollution on survival after a lung cancer diagnosis, in contrast with the results from the naïve approach. This aim demonstrates that avoidable systematic errors can be introduced when time-varying exposures in a time-to-event context are not handled appropriately.

The final aim of this dissertation addressed inconsistencies in the measurement of segregation in the racial residential segregation and cancer outcomes literature. In this aim, the relationship between racial residential segregation and lung cancer survival was examined using the dissimilarity index, a popular segregation index that measures the evenness of two population groups across space. To illustrate the contribution of considering both spatial scale and spatial relationships in the measurement of segregation, the dissimilarity index was calculated eight ways. In these eight calculations, four spatial scales and both spatial and aspatial measures were used. These eight dissimilarity indices were then used to estimate the impact of racial residential

segregation on lung cancer survival. We found that spatial scale and using indices that account for spatial relationships impact both the absolute estimate of segregation and the effect estimate for the association between segregation and lung cancer survival.

This dissertation advances the field of social and environmental epidemiology in lung cancer in several ways. First, Chapter 2 illustrates how spatial epidemiology can be used to investigate patterns in lung cancer (or other outcomes) to elucidate new potential modifiable risk factors and design targeted interventions to reduce geographic inequalities. Next, Chapter 3 illustrates how time-varying exposures in time-to-event contexts require specific considerations to ensure that systematic error is not introduced through mishandling of these exposures. This illustration will increase the understanding of this error and methods that can be used to avoid it. Finally, Chapter 4 of the dissertation illustrates the importance of explicitly considering spatial relationships and spatial scale in measuring inherently spatial exposures, such as residential segregation. This chapter emphasizes that when exposures such as racial residential segregation are being used, it is necessary to explicitly define the choice of spatial scale and measures that account for spatial relationships. If there is no specific reason for the analytic choices for measuring segregation, it is necessary to conduct sensitivity analyses to ensure these decisions are not driving conclusions.

5.2 Implications

Lung cancer survival remains lower than all other leading cancers, despite improvements in treatment and prevention over several decades. Recently, there has been an increased interest in examining how place-based factors may be contributing to poor survival and may inform potential targeted interventions. In order to identify and explore these place-based factors, it is necessary to consider specific methodological considerations. Understanding the role of neighborhood

contextual factors in lung cancer and corresponding methodological considerations is an important area of research for several reasons.

First, disease mapping is an invaluable tool in spatial epidemiology to assess geographic patterns and spatial anomalies. Studies in lung cancer and other cancers commonly only examine survival changes over time, but not over space (128, 137, 210). Examining the extent of small-area variation in cancer survival can aid in hypothesis-generating for risk factors that similarly vary over space and time, such as neighborhood factors and environmental determinants. Further, policymakers that make evidence-based decisions on resource allocation or intervention planning to address disparities in cancer survival may rely on information about small-area heterogeneity.

Second, as the climate changes, research on the impacts of air pollutants and other environmental factors will increase. Air pollution is a unique risk factor in that it can be modified by both individual-level behavior changes and at the population level through policies (211, 212). Specific methodological concerns arise with air pollution or other environmental risk factors that vary over space and time. Often, studies simplify time-varying exposures by using maximums, minimums, or averages. Until now, there has been little discussion of how this simplification of a time-varying exposure in a time-to-event context can introduce ITB in the context of environmental exposures. The impact of this bias can be significant, as illustrated in this dissertation. As these studies increase in frequency, it is necessary to understand how to avoid potential bias in environmental epidemiology studies. As climate change continues, epidemiologists will more often seek to understand the impact of environmental exposures. It is necessary to elucidate these avoidable errors and provide approachable solutions.

Finally, it is necessary to understand how racial residential segregation impacts lung cancer to design and implement targeted local interventions to improve equity in cancer outcomes.

Policies designed to improve equity in health care access and other factors related to unequal access to economic resources can impact lung cancer survival. As more studies seek to understand the impact of residential segregation and other neighborhood factors in cancer outcomes, it is important to illustrate potential challenges in measuring segregation or related exposures. Studies that use segregation measures without regard for the spatial structure of populations or the consideration of how spatial size impacts estimates can lead to inconsistent conclusions driven in part by discrepancies in how the segregation index is measured and what it is capturing.

5.3 On the relationship between racial residential segregation and air pollution

Although the present work and previous studies often examine the impacts of racial residential segregation and air pollution separately, they are interrelated factors that may have multiplicative impacts on adverse health and racial disparities. Black and Hispanic Americans are disproportionately exposed to pollutants of all kinds, which may be a partial driver of unequal health outcomes (213). This unequal exposure is most pronounced in metropolitan areas with high levels of racial residential segregation (214). This inequity is further exacerbated because air pollution is primarily caused by NH white populations' consumption of goods and services but is disproportionately experienced by Black and Hispanic populations (215). The concept of disproportionate exposure to pollutants is known as environmental injustice, which acknowledges the racial and economic disparities among communities bearing disproportionate environmental costs, such as increased exposure to lead, particulate air pollution, ozone, or other health hazards. Expanding public health research to examine the joint effects of air pollution and racial residential segregation on adverse health outcomes such as lung cancer would have benefits for improving population health and health disparities.

5.4 Limitations and recommendations for future work

This dissertation has several limitations worth noting. While population-based cancer registries are well-recognized for containing quality information on demographics and cancer incidence, there is missing relevant information that may be important in lung cancer survival. California Cancer Registry (CCR) does not contain detailed information on treatment, diagnostics, comorbidities, smoking status, or functional status. Previous studies have demonstrated that linkages with claims or electronic health records can supplement missing information in cancer registries (216). Another limitation in the CCR is that information on residential mobility is not collected. For this reason, all analyses in this dissertation use zip code at diagnosis for longitudinal exposure ascertainment. This may result in misclassification of exposure status for those whose mobility is related to neighborhood conditions related to social or environmental exposures. Previous studies have demonstrated that cancer survivors have shorter housing tenure than their healthy counterparts (217). One study found that 65% of colorectal cancer patients remained at the same residence during a five-year period, while 22% changed the census tract, and 12% moved out of state (218). Future studies could utilize corrections to adjust for and assess the extent of misclassification (219).

As the field of spatial epidemiology is growing, particularly in cancer research, current methods have been developed to improve small-area estimation (220). Small-area analysis can help describe local variations and highlight geographic disparities. This estimation can result in instability in estimates, especially for rare diseases. Recent studies have illustrated how Bayesian modeling can be well suited to employ spatial and temporal smoothing to produce more stable estimates (221). Two studies demonstrated how these methods are being extended to improve small area estimates (222) and improve modeling or rare outcomes through spatiotemporal mixture

modeling (123). As small-area level data is becoming more readily available, there is more possibility for analyses that use a geospatial perspective. Further, future work could investigate how other neighborhood factors that vary over space and time may contribute to lung cancer survival heterogeneity. Previous studies have demonstrated that factors such as green space, neighborhood deprivation, and distance to medical centers may be important determinants in cancer survival (223-225).

In the field of lung cancer survival, little is known about potential interactions between air pollution or other environmental hazards and cancer treatment. No existing studies have examined the interaction between air pollution exposure and lung cancer treatment(s) on survival. There is established evidence that smoking after a lung cancer diagnosis interacts with cancer treatment to reduce survival through increased risk for second primary cancers, risk of recurrence, poor response to treatment, and increased treatment-related toxicity (226-230). Both tobacco smoke and air pollution consist of a complex mixture of gaseous and particulate components that vary in chemical composition. While PM exposure occurs at much lower exposure concentrations than cigarette smoke, similar biologic mechanisms are documented (231, 232). For patients receiving chemotherapy and other pharmacologic compounds, air pollution may result in reduced efficacy of chemotherapy and reduced drug metabolism (233) (234). Among patients receiving radiation therapy, air pollution may interact with radiation to increase second lung cancer risk, increase risk of complications, and decrease survival. Second lung cancer risk is higher among smokers who received chest irradiation, and smokers experienced an increased likelihood of complications and decreased survival than nonsmokers (235, 236). In patients that receive surgical resection, evidence has shown that smokers are at an increased risk of postoperative complications (237). Finally, lung cancer patients who quit smoking after diagnosis have a decreased risk of multiple

primary lung tumors (238). While the available evidence is specific to cigarette smokers, it is plausible that similar interactions may exist between air pollution and lung cancer treatment to affect survival.

Racial residential segregation is one of the most commonly studied structural racism domains and has been demonstrated to influence health across various outcomes. However, other areas of structural racism extend beyond neighborhood context and can include national, state, and local laws. It is currently unknown how these racist laws and policies impact lung cancer incidence and survival. For example, voter suppression laws are known to disproportionately affect people of color. As a result, areas with low voter turnout may have decreased access to opportunities that offer equity-focused health benefits, such as Medicaid expansion. These opportunities may have a meaningful impact on lung cancer survival through access to primary care providers, specialty providers, and specialized cancer centers. Similarly, racist policies such as anti-drug legislation have resulted in mandatory minimum sentences and subsequent increasing numbers of Black Americans being incarcerated. Incarceration and other forms of disenfranchisement can lead to increased stress resulting in community and individual health effects (239). Future studies could consider how these other structural and environmental racism domains impact lung cancer survival and disparities.

5.5 Concluding remarks

In conclusion, this dissertation examines the role of neighborhood contextual determinants in lung cancer survival. This work expands on the current evidence by examining spatial heterogeneity in lung cancer survival, illuminating a systematic error in work examining the role of time-varying air pollution on lung cancer survival, and illustrating the considerations needed for measuring racial residential segregation for consistent and robust estimation of effect in lung

cancer survival. This body of work advances the field by addressing common methodological issues in epidemiologic studies of social and environmental exposures in lung cancer survival and provides detailed recommendations to guide future research.

References

1. American Cancer Society, Cancer Facts and Figures 2021. https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2021/cancer-facts-and-figures-2021.pdf.

2. Cancer IAftSoL. 8th Edition of the TNM Classification for Lung Cancer2019.

3. Atkins GT, Kim T, Munson J. Residence in Rural Areas of the United States and Lung Cancer Mortality. Disease Incidence, Treatment Disparities, and Stage-Specific Survival. Ann Am Thorac Soc. 2017;14(3):403-11. doi:10.1513/AnnalsATS.201606-469OC

4. Kehm RD, Yang W, Tehranifar P, Terry MB. 40 Years of Change in Age- and Stage-Specific Cancer Incidence Rates in US Women and Men. JNCI Cancer Spectr. 2019;3(3):pkz038-pkz. doi:10.1093/jncics/pkz038

5. Siegel RL, Jacobs EJ, Newton CC, et al. Deaths Due to Cigarette Smoking for 12 Smoking-Related Cancers in the United States. JAMA internal medicine. 2015;175(9):1574-6. doi:10.1001/jamainternmed.2015.2398

6. Ebrahimi H, Aryan Z, Saeedi Moghaddam S, et al. Global, regional, and national burden of respiratory tract cancers and associated risk factors from 1990 to 2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet Respiratory Medicine. 2021;9(9):1030-49. doi:https://doi.org/10.1016/S2213-2600(21)00164-8

7. Guidelines on the selection of patients with lung cancer for surgery. Thorax. 2001;56(2):89. doi:10.1136/thorax.56.2.89

8. National Cancer Institute S, Epidemiology and End Results Program. Cancer Stat Facts: Lung and Bronchus Cancer. 2019. https://seer.cancer.gov/statfacts/html/lungb.html. 2019.

9. Visbal AL, Williams BA, Nichols III FC, et al. Gender differences in non–small-cell lung cancer survival: an analysis of 4,618 patients diagnosed between 1997 and 2002. The Annals of thoracic surgery. 2004;78(1):209-15.

10. Bilello KS, Murin S, Matthay RA. Epidemiology, etiology, and prevention of lung cancer. Clinics in chest medicine. 2002;23(1):1-25.

11. Radkiewicz C, Dickman PW, Johansson ALV, Wagenius G, Edgren G, Lambe M. Sex and survival in non-small cell lung cancer: A nationwide cohort study. PloS one. 2019;14(6):e0219206. doi:10.1371/journal.pone.0219206

12. Sagerup CMT, Småstuen M, Johannesen TB, Helland Å, Brustugun OT. Sex-specific trends in lung cancer incidence and survival: a population study of 40 118 cases. Thorax. 2011;66(4):301-7. doi:10.1136/thx.2010.151621

13. Lu T, Yang X, Huang Y, et al. Trends in the incidence, treatment, and survival of patients with lung cancer in the last four decades. Cancer Manag Res. 2019;11:943-53. doi:10.2147/CMAR.S187317

14. Fuertes M, Castilla J, Alonso C, Prez J. Cisplatin biochemical mechanism of action: from cytotoxicity to induction of cell death through interconnections between apoptotic and necrotic pathways. Current medicinal chemistry. 2003;10(3):257-66.

15. Paesmans M. Prognostic and predictive factors for lung cancer. Breathe. 2012;9(2):112-21. doi:10.1183/20734735.006911

16. Giroux Leprieur E, Lavole A, Ruppert AM, et al. Factors associated with long-term survival of patients with advanced non-small cell lung cancer. Respirology. 2012;17(1):134-42. doi:10.1111/j.1440-1843.2011.02070.x

17. Davis JS, Prophet E, Peng H-L, et al. Potential Influence on Clinical Trials of Long-Term Survivors of Stage IV Non-small cell Lung Cancer. JNCI Cancer Spectr. 2019;3(2). doi:10.1093/jncics/pkz010

18. Huang C-Y, Chen B-H, Chou W-C, Yang C-T, Chang JW-C. Factors associated with the prognosis and long-term survival of patients with metastatic lung adenocarcinoma: a retrospective analysis. J Thorac Dis. 2018;10(4):2070-8.

19. National Cancer Institute, Lung Cancer Screening. https://progressreport.cancer.gov/detection/lung_cancer. Accessed September 24 2019.

20. Cataldo JK, Dubey S, Prochaska JJ. Smoking cessation: an integral part of lung cancer treatment. Oncology. 2010;78(5-6):289-301. doi:10.1159/000319937

21. Walker MS, Vidrine DJ, Gritz ER, et al. Smoking relapse during the first year after treatment for early-stage non-small-cell lung cancer. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2006;15(12):2370-7. doi:10.1158/1055-9965.Epi-06-0509

22. Mayer DK, Carlson J. Smoking patterns in cancer survivors. Nicotine & Tobacco Research. 2010;13(1):34-40.

23. Gallaway MS, Glover-Kudon R, Momin B, et al. Smoking cessation attitudes and practices among cancer survivors – United States, 2015. Journal of Cancer Survivorship. 2019;13(1):66-74. doi:10.1007/s11764-018-0728-2

24. Jones GS, Baldwin DR. Recent advances in the management of lung cancer. Clin Med (Lond). 2018;18(Suppl 2):s41-s6. doi:10.7861/clinmedicine.18-2-s41

25. Diez Roux AV. Investigating neighborhood and area effects on health. American journal of public health. 2001;91(11):1783-9.

26. Bennett KJ, Probst JC, Bellinger JD. Receipt of cancer screening services: surprising results for some rural minorities. The Journal of Rural Health. 2012;28(1):63-72.

27. Gordon-Larsen P, Nelson MC, Page P, Popkin BM. Inequality in the built environment underlies key health disparities in physical activity and obesity. Pediatrics. 2006;117(2):417-24.

28. Cummins S, Macintyre S. Food environments and obesity—neighbourhood or nation? International journal of epidemiology. 2006;35(1):100-4.

29. McEntee JC, Ogneva-Himmelberger Y. Diesel particulate matter, lung cancer, and asthma incidences along major traffic corridors in MA, USA: A GIS analysis. Health & place. 2008;14(4):817-28.

30. Richards TB, Henley SJ, Puckett MC, et al. Lung cancer survival in the United States by race and stage (2001-2009): Findings from the CONCORD-2 study. Cancer. 2017;123 Suppl 24(Suppl 24):5079-99. doi:10.1002/cncr.31029

31. Lelieveld J, Evans JS, Fnais M, Giannadaki D, Pozzer A. The contribution of outdoor air pollution sources to premature mortality on a global scale. Nature. 2015;525(7569):367-71. doi:10.1038/nature15371

32. Lim SS, Vos T, Flaxman AD, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. The lancet. 2012;380(9859):2224-60.

33. Gualtieri M, Øvrevik J, Mollerup S, et al. Airborne urban particles (Milan winter-PM2. 5) cause mitotic arrest and cell death: effects on DNA, mitochondria, AhR binding and spindle organization. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis. 2011;713(1-2):18-31.

34. Wang G, Zhen L, Lü P, Jiang R, Song W. Effects of ozone and fine particulate matter (PM2. 5) on rat cardiac autonomic nervous system and systemic inflammation. Wei sheng yan jiu= Journal of hygiene research. 2013;42(4):554-60.

35. Deng X, Rui W, Zhang F, Ding W. PM 2.5 induces Nrf2-mediated defense mechanisms against oxidative stress by activating PIK3/AKT signaling pathway in human lung alveolar epithelial A549 cells. Cell biology and toxicology. 2013;29(3):143-57.

36. Valavanidis A, Fiotakis K, Vlachogianni T. Airborne particulate matter and human health: toxicological assessment and importance of size and composition of particles for oxidative damage and carcinogenic mechanisms. Journal of environmental science and health. Part C, Environmental carcinogenesis & ecotoxicology reviews. 2008;26(4):339-62. doi:10.1080/10590500802494538

37. Hajat A, Hsia C, O'Neill MS. Socioeconomic Disparities and Air Pollution Exposure: a Global Review. Current environmental health reports. 2015;2(4):440-50. doi:10.1007/s40572-015-0069-5

38. Research TNBoE. Recent Increases in Air Pollution: Evidence and Implications for Mortality. 2019.

39. Brook RD, Rajagopalan S, Pope CA, 3rd, et al. Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association. Circulation. 2010;121(21):2331-78. doi:10.1161/CIR.0b013e3181dbece1

40. Newby DE, Mannucci PM, Tell GS, et al. Expert position paper on air pollution and cardiovascular disease. European heart journal. 2015;36(2):83-93b. doi:10.1093/eurheartj/ehu458

41. Crouse DL, Goldberg MS, Ross NA, Chen H, Labreche F. Postmenopausal breast cancer is associated with exposure to traffic-related air pollution in Montreal, Canada: a case-control study. Environmental health perspectives. 2010;118(11):1578-83. doi:10.1289/ehp.1002221

42. Cho C-C, Hsieh W-Y, Tsai C-H, Chen C-Y, Chang H-F, Lin C-S. In Vitro and In Vivo Experimental Studies of PM(2.5) on Disease Progression. International journal of environmental research and public health. 2018;15(7):1380. doi:10.3390/ijerph15071380

43. Di Q, Wang Y, Zanobetti A, et al. Air Pollution and Mortality in the Medicare Population. The New England journal of medicine. 2017;376(26):2513-22. doi:10.1056/NEJMoa1702747

44. Liu C, Sun J, Liu Y, et al. Different exposure levels of fine particulate matter and preterm birth: a meta-analysis based on cohort studies. Environmental science and pollution research international. 2017;24(22):17976-84. doi:10.1007/s11356-017-9363-0

45. Chen G, Guo Y, Abramson MJ, Williams G, Li S. Exposure to low concentrations of air pollutants and adverse birth outcomes in Brisbane, Australia, 2003-2013. The Science of the total environment. 2018;622-623:721-6. doi:10.1016/j.scitotenv.2017.12.050

46. Pope CA, 3rd. Mortality effects of longer term exposures to fine particulate air pollution: review of recent epidemiological evidence. Inhalation toxicology. 2007;19 Suppl 1:33-8. doi:10.1080/08958370701492961

47. Kim H-B, Shim J-Y, Park B, Lee Y-J. Long-Term Exposure to Air Pollutants and Cancer Mortality: A Meta-Analysis of Cohort Studies. International journal of environmental research and public health. 2018;15(11):2608. doi:10.3390/ijerph15112608

48. Cui P, Huang Y, Han J, Song F, Chen K. Ambient particulate matter and lung cancer incidence and mortality: a meta-analysis of prospective studies. European journal of public health. 2015;25(2):324-9. doi:10.1093/eurpub/cku145

49. Hamra GB, Guha N, Cohen A, et al. Outdoor particulate matter exposure and lung cancer: a systematic review and meta-analysis. Environmental health perspectives. 2014;122(9):906-11. doi:10.1289/ehp.1408092

10.1289/ehp/1408092

50. Yang WS, Zhao H, Wang X, Deng Q, Fan WY, Wang L. An evidence-based assessment for the association between long-term exposure to outdoor air pollution and the risk of lung cancer. European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP). 2016;25(3):163-72. doi:10.1097/cej.0000000000000158

51. Li J, Li WX, Bai C, Song Y. Particulate matter-induced epigenetic changes and lung cancer. The clinical respiratory journal. 2017;11(5):539-46. doi:10.1111/crj.12389

52. International Agency for Research on Cancer WHO. IARC: Outdoor air pollution a leading enviornmental cause of cancer deaths. 2013.

53. Loomis D, Grosse Y, Lauby-Secretan B, et al. The carcinogenicity of outdoor air pollution. The Lancet. Oncology. 2013;14(13):1262-3.

54. Jin X, Xue B, Zhou Q, Su R, Li Z. Mitochondrial damage mediated by ROS incurs bronchial epithelial cell apoptosis upon ambient PM2. 5 exposure. The Journal of toxicological sciences. 2018;43(2):101-11.

55. Weichenthal SA, Pollitt KG, Villeneuve PJ. PM 2.5, oxidant defence and cardiorespiratory health: a review. Environmental Health. 2013;12(1):40.

56. Becker S, Soukup J. Coarse(PM(2.5-10)), fine(PM(2.5)), and ultrafine air pollution particles induce/increase immune costimulatory receptors on human blood-derived monocytes but not on alveolar macrophages. Journal of toxicology and environmental health. Part A. 2003;66(9):847-59. doi:10.1080/15287390306381

57. Wang G, Zhao J, Jiang R, Song W. Rat lung response to ozone and fine particulate matter (PM2.5) exposures. Environmental toxicology. 2015;30(3):343-56. doi:10.1002/tox.21912

58. Habre R, Moshier E, Castro W, et al. The effects of PM2.5 and its components from indoor and outdoor sources on cough and wheeze symptoms in asthmatic children. Journal of exposure science & environmental epidemiology. 2014;24(4):380-7. doi:10.1038/jes.2014.21

59. Darrow LA, Klein M, Flanders WD, Mulholland JA, Tolbert PE, Strickland MJ. Air pollution and acute respiratory infections among children 0-4 years of age: an 18-year time-series study. American journal of epidemiology. 2014;180(10):968-77. doi:10.1093/aje/kwu234

60. Zhao H, Li W, Gao Y, Li J, Wang H. Exposure to particular matter increases susceptibility to respiratory Staphylococcus aureus infection in rats via reducing pulmonary natural killer cells. Toxicology. 2014;325:180-8. doi:10.1016/j.tox.2014.09.006

61. Wu S, Deng F, Hao Y, et al. Chemical constituents of fine particulate air pollution and pulmonary function in healthy adults: the Healthy Volunteer Natural Relocation study. Journal of hazardous materials. 2013;260:183-91. doi:10.1016/j.jhazmat.2013.05.018

62. Franchini M, Mannucci PM. Thrombogenicity and cardiovascular effects of ambient air pollution. Blood. 2011;118(9):2405-12. doi:10.1182/blood-2011-04-343111

63. Hicken MT, Adar SD, Hajat A, et al. Air Pollution, Cardiovascular Outcomes, and Social Disadvantage: The Multi-ethnic Study of Atherosclerosis. Epidemiology (Cambridge, Mass.). 2016;27(1):42-50. doi:10.1097/EDE.0000000000000367

64. Kurt OK, Zhang J, Pinkerton KE. Pulmonary health effects of air pollution. Current opinion in pulmonary medicine. 2016;22(2):138-43. doi:10.1097/mcp.0000000000000248

65. Coussens LM, Werb Z. Inflammation and cancer. Nature. 2002;420(6917):860.

66. Ballaz S, Mulshine JL. The potential contributions of chronic inflammation to lung carcinogenesis. Clinical lung cancer. 2003;5(1):46-62. doi:10.3816/CLC.2003.n.021

67. Enewold L, Mechanic LE, Bowman ED, et al. Serum concentrations of cytokines and lung cancer survival in African Americans and Caucasians. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2009;18(1):215-22. doi:10.1158/1055-9965.EPI-08-0705

68. Kang D, Hamasaki N. Mitochondrial oxidative stress and mitochondrial DNA. Clinical chemistry and laboratory medicine. 2003;41(10):1281-8. doi:10.1515/cclm.2003.195

69. Kong SY, Goodman M, Judd S, Bostick RM, Flanders WD, McClellan W. Oxidative balance score as predictor of all-cause, cancer, and noncancer mortality in a biracial US cohort. Ann Epidemiol. 2015;25(4):256-62.e1. doi:10.1016/j.annepidem.2015.01.004

70. Eckel SP, Cockburn M, Shu YH, et al. Air pollution affects lung cancer survival. Thorax. 2016;71(10):891-8. doi:10.1136/thoraxjnl-2015-207927

71. Deng H, Eckel SP, Liu L, Lurmann FW, Cockburn MG, Gilliland FD. Particulate matter air pollution and liver cancer survival. International journal of cancer. 2017;141(4):744-9. doi:10.1002/ijc.30779

72. Hu H, Dailey AB, Kan H, Xu X. The effect of atmospheric particulate matter on survival of breast cancer among US females. Breast cancer research and treatment. 2013;139(1):217-26. doi:10.1007/s10549-013-2527-9

73. Xu X, Ha S, Kan H, Hu H, Curbow BA, Lissaker CT. Health effects of air pollution on length of respiratory cancer survival. BMC public health. 2013;13:800. doi:10.1186/1471-2458-13-800

74. Villanueva C, Chang J, Ziogas A, Bristow RE, Vieira VM. Ambient air pollution and ovarian cancer survival in California. Gynecologic oncology. 2021. doi:https://doi.org/10.1016/j.ygyno.2021.07.036

75. Massey DS, Denton NA. The dimensions of residential segregation. Social forces. 1988;67(2):281-315.

76. Silverman RM, Patterson KL. The four horsemen of the fair housing apocalypse: a critique of fair housing policy in the USA. Critical Sociology. 2012;38(1):123-40.

77. Krysan M, Farley R. The residential preferences of blacks: Do they explain persistent segregation? Social Forces. 2002;80(3):937-80.

78. Massey D, Denton NA. American apartheid: Segregation and the making of the underclass: Harvard university press; 1993.

79. Williams DR, Collins C. Racial residential segregation: a fundamental cause of racial disparities in health. Public health reports (Washington, D.C. : 1974). 2001;116(5):404-16. doi:10.1093/phr/116.5.404

80. Landrine H, Corral I. Separate and unequal: residential segregation and black health disparities. Ethnicity & disease. 2009;19(2):179.

81. Kirby JB, Kaneda T. Neighborhood socioeconomic disadvantage and access to health care. Journal of health and social behavior. 2005;46(1):15-31.

82. Gaskin DJ, Dinwiddie GY, Chan KS, McCleary R. Residential segregation and disparities in health care services utilization. Medical Care Research and Review. 2012;69(2):158-75.

83. Arnett MJ, Thorpe RJ, Gaskin DJ, Bowie JV, LaVeist TA. Race, Medical Mistrust, and Segregation in Primary Care as Usual Source of Care: Findings from the Exploring Health Disparities in Integrated Communities Study. Journal of Urban Health. 2016;93(3):456-67. doi:10.1007/s11524-016-0054-9

84. Auchincloss AH, Van Nostrand JF, Ronsaville D. Access to health care for older persons in the United States: personal, structural, and neighborhood characteristics. Journal of Aging and Health. 2001;13(3):329-54.

85. White K, Haas JS, Williams DR. Elucidating the role of place in health care disparities: the example of racial/ethnic residential segregation. Health services research. 2012;47(3pt2):1278-99.

86. Asch DA, Islam MN, Sheils NE, et al. Patient and Hospital Factors Associated With Differences in Mortality Rates Among Black and White US Medicare Beneficiaries Hospitalized With COVID-19 Infection. JAMA Network Open. 2021;4(6):e2112842-e. doi:10.1001/jamanetworkopen.2021.12842

87. Pollack CE, Bekelman JE, Liao KJ, Armstrong K. Hospital racial composition and the treatment of localized prostate cancer. Cancer. 2011;117(24):5569-78. doi:10.1002/cncr.26232

88. Sarrazin MSV, Campbell ME, Richardson KK, Rosenthal GE. Racial segregation and disparities in health care delivery: conceptual model and empirical assessment. Health services research. 2009;44(4):1424-44. doi:10.1111/j.1475-6773.2009.00977.x

89. Rosero EB, Joshi GP, Minhajuddin A, Timaran CH, Modrall JG. Effects of hospital safety-net burden and hospital volume on failure to rescue after open abdominal aortic surgery. J Vasc Surg. 2017;66(2):404-12. doi:10.1016/j.jvs.2016.12.146

90. Werner RM, Goldman LE, Dudley RA. Comparison of change in quality of care between safety-net and non–safety-net hospitals. Jama. 2008;299(18):2180-7.

91. Haas JS, Earle CC, Orav JE, et al. Racial segregation and disparities in breast cancer care and mortality. Cancer. 2008;113(8):2166-72.

92. Chan KS, Gaskin DJ, Dinwiddie GY, McCleary R. Do diabetic patients living in racially segregated neighborhoods experience different access and quality of care? Medical care. 2012;50(8):692.

93. Johnston R, Poulsen M, Forrest J. Moving On from Indices, Refocusing on Mix: On Measuring and Understanding Ethnic Patterns of Residential Segregation. Journal of Ethnic and Migration Studies. 2010;36(4):697-706. doi:10.1080/13691830903505045

94. Reardon SF, O'Sullivan D. Measures of spatial segregation. Sociological methodology. 2004;34(1):121-62.

95. Massey DS, Denton NA. Hypersegregation in U.S. metropolitan areas: black and Hispanic segregation along five dimensions. Demography. 1989;26(3):373-91.

96. Booth A, Crouter AC. Does it take a village?: Community effects on children, adolescents, and families: Psychology Press; 2001.

97. Massey DS, White MJ, Phua V-C. The dimensions of segregation revisited. Sociological methods & research. 1996;25(2):172-206.

98. Singh GK. Area socioeconomic variations in US cancer incidence, mortality, stage, treatment, and survival, 1975-1999: US Department of Health and Human Services, National Institutes of Health …; 2003.

99. Coughlin SS, Richards TB, Thompson T, et al. Rural/nonrural differences in colorectal cancer incidence in the United States, 1998–2001. Cancer. 2006;107(S5):1181-8.

100. Ellis L, Canchola AJ, Spiegel D, Ladabaum U, Haile R, Gomez SL. Racial and Ethnic Disparities in Cancer Survival: The Contribution of Tumor, Sociodemographic, Institutional, and Neighborhood Characteristics. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2018;36(1):25-33. doi:10.1200/JCO.2017.74.2049

101. Hayanga AJ, Zeliadt SB, Backhus LM. Residential segregation and lung cancer mortality in the United States. JAMA surgery. 2013;148(1):37-42.

102. Haas JS, Earle CC, Orav JE, Brawarsky P, Neville BA, Williams DR. Racial segregation and disparities in cancer stage for seniors. Journal of general internal medicine. 2008;23(5):699-705.

103. Johnson AM, Johnson A, Hines RB, Bayakly R. The effects of residential segregation and neighborhood characteristics on surgery and survival in patients with early-stage non–small cell lung cancer. Cancer Epidemiology and Prevention Biomarkers. 2016;25(5):750-8.

104. Warner ET, Gomez SL. Impact of neighborhood racial composition and metropolitan residential segregation on disparities in breast cancer stage at diagnosis and survival between black and white women in California. Journal of community health. 2010;35(4):398-408.

105. Zhou Y, Bemanian A, Beyer KM. Housing discrimination, residential racial segregation, and colorectal cancer survival in southeastern Wisconsin. AACR; 2017.

106. Poulson M, Cornell E, Madiedo A, et al. The Impact of Racial Residential Segregation on Colorectal Cancer Outcomes and Treatment. Ann Surg. 2020. doi:10.1097/sla.0000000000004653

107. Poulson MR, Beaulieu-Jones BR, Kenzik KM, et al. Residential racial segregation and disparities in breast cancer presentation, treatment, and survival. Annals of surgery. 2020;273(1):3-9.

108. Westrick AC, Bailey ZD, Schlumbrecht M, et al. Residential segregation and overall survival of women with epithelial ovarian cancer. Cancer. 2020;126(16):3698-707.

109. Harris JA, Hunter WP, Ji YD, Hanna GJ. Effects of racial residential segregation on oral squamous cell carcinoma prognosis and survival. Oral Oncology. 2021;119:105364.

110. Openshaw S. The modifiable areal unit problem. Quantitative geography: A British view. 1981:60-9.

111. Johnston R, Poulsen M, Forrest J. Ethnic and racial segregation in US metropolitan areas, 1980-2000: The dimensions of segregation revisited. Urban Affairs Review. 2007;42(4):479-504.

112. O'Sullivan D, Wong DWS. A Surface-Based Approach to Measuring Spatial Segregation. Geographical Analysis. 2007;39(2):147-68. doi:https://doi.org/10.1111/j.1538-4632.2007.00699.x

113. Wong DWS. Comparing Traditional and Spatial Segregation Measures: A Spatial Scale Perspective1. Urban Geography. 2004;25(1):66-82. doi:10.2747/0272-3638.25.1.66

114. White MJ. The measurement of spatial segregation. American journal of sociology. 1983;88(5):1008-18.

115. Wong DWS. Spatial Decomposition of Segregation Indices: A Framework Toward Measuring Segregation at Multiple Levels. Geographical Analysis. 2003;35(3):179-94. doi:https://doi.org/10.1111/j.1538-4632.2003.tb01109.x

116. Lee BA, Reardon SF, Firebaugh G, Farrell CR, Matthews SA, O'Sullivan D. Beyond the census tract: Patterns and determinants of racial segregation at multiple geographic scales. American Sociological Review. 2008;73(5):766-91.

117. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. CA: a cancer journal for clinicians. 2021;71(1):7-33. doi:https://doi.org/10.3322/caac.21654

118. Cheng T-YD, Cramb SM, Baade PD, Youlden DR, Nwogu C, Reid ME. The International Epidemiology of Lung Cancer: Latest Trends, Disparities, and Tumor Characteristics. Journal of Thoracic Oncology. 2016;11(10):1653-71. doi:https://doi.org/10.1016/j.jtho.2016.05.021

119. Yen IH, Syme SL. The social environment and health: a discussion of the epidemiologic literature. Annual review of public health. 1999;20(1):287-308.

120. Gomez SL, Shariff-Marco S, DeRouen M, et al. The impact of neighborhood social and built environment factors across the cancer continuum: Current research, methodological considerations, and future directions. Cancer. 2015;121(14):2314-30. doi:https://doi.org/10.1002/cncr.29345

121. Ellen IG, Turner MA. Do neighborhoods matter and why? Choosing a better life. 2003:313-38.

122. Diez Roux AV, Mair C. Neighborhoods and health. 2010.

123. Carroll R, Lawson AB, Kirby RS, Faes C, Aregay M, Watjou K. Space-time variation of respiratory cancers in South Carolina: a flexible multivariate mixture modeling approach to risk estimation. Annals of epidemiology. 2017;27(1):42-51.

124. Cramb SM, Moraga P, Mengersen KL, Baade PD. Spatial variation in cancer incidence and survival over time across Queensland, Australia. Spatial and Spatio-temporal Epidemiology. 2017;23:59-67. doi:https://doi.org/10.1016/j.sste.2017.09.002

125. Haynes R, Pearce J, Barnett R. Cancer survival in New Zealand: Ethnic, social and geographical inequalities. Social Science & Medicine. 2008;67(6):928-37. doi:https://doi.org/10.1016/j.socscimed.2008.05.005

126. Wang M, Matthews SA, Iskandarani K, et al. Spatial-temporal analysis of prostate cancer incidence from the Pennsylvania Cancer Registry, 2000-2011. Geospatial health. 2017;12(2):611. doi:10.4081/gh.2017.611

127. Lee Y-C, Calderon-Candelario RA, Holt GE, Campos MA, Mirsaeidi M. State-Level Disparity in Lung Cancer Survival in the United States. Frontiers in Oncology. 2020;10(1449). doi:10.3389/fonc.2020.01449

128. Li T, Pan K, Ellinwood AK, Cress RD. Survival Trends of Metastatic Lung Cancer in California by Age at Diagnosis, Gender, Race/Ethnicity, and Histology, 1990-2014. Clinical lung cancer. 2020. doi:10.1016/j.cllc.2020.11.005

129. Corazziari I, Quinn M, Capocaccia R. Standard cancer patient population for age standardising survival ratios. Eur J Cancer. 2004;40(15):2307-16. doi:10.1016/j.ejca.2004.07.002

130. Brenner H, Arndt V, Gefeller O, Hakulinen T. An alternative approach to age adjustment of cancer survival rates. European Journal of Cancer. 2004;40(15):2317-22. doi:https://doi.org/10.1016/j.ejca.2004.07.007

131. Powell HA, Tata LJ, Baldwin DR, et al. Treatment decisions and survival for people with small-cell lung cancer. British journal of cancer. 2014;110(4):908-15. doi:10.1038/bjc.2013.812

132. Blom EF, Ten Haaf K, Arenberg DA, de Koning HJ. Disparities in Receiving Guideline-Concordant Treatment for Lung Cancer in the United States. Ann Am Thorac Soc. 2020;17(2):186-94. doi:10.1513/AnnalsATS.201901-094OC

133. Forrest LF, Adams J, Wareham H, Rubin G, White M. Socioeconomic inequalities in lung cancer treatment: systematic review and meta-analysis. PLoS Med. 2013;10(2):e1001376. doi:10.1371/journal.pmed.1001376

134. Ambroggi M, Biasini C, Del Giovane C, Fornari F, Cavanna L. Distance as a Barrier to Cancer Diagnosis and Treatment: Review of the Literature. The oncologist. 2015;20(12):1378-85. doi:10.1634/theoncologist.2015-0110

135. Campbell N, Elliott A, Sharp L, Ritchie L, Cassidy J, Little J. Rural factors and survival from cancer: analysis of Scottish cancer registrations. British journal of cancer. 2000;82(11):1863-6.

136. Elliott TE, Elliott BA, Renier CM, Haller IV. Rural-urban differences in cancer care: results from the Lake Superior Rural Cancer Care Project. Minnesota medicine. 2004;87(9):44-50.

137. Lou Y, Dholaria B, Soyano A, et al. Survival trends among non-small-cell lung cancer patients over a decade: Impact of initial therapy at academic centers. Cancer medicine. 2018;7(10):4932-42.

138. Tannenbaum SL, Koru-Sengul T, Zhao W, Miao F, Byrne MM. Survival Disparities in Non–Small Cell Lung Cancer by Race, Ethnicity, and Socioeconomic Status. The Cancer Journal. 2014;20(4):237-45. doi:10.1097/ppo.0000000000000058

139. Sugimura H, Yang P. Long-term survivorship in lung cancer: a review. Chest. 2006;129(4):1088-97.

140. Wang N, Mengersen K, Kimlin M, et al. Lung cancer and particulate pollution: A critical review of spatial and temporal analysis evidence. Environmental research. 2018;164:585-96. doi:10.1016/j.envres.2018.03.034

141. Hart JE. Air pollution affects lung cancer survival. Thorax. 2016;71(10):875-6. doi:10.1136/thoraxjnl-2016-208967

142. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. American journal of epidemiology. 2016;183(8):758-64.

143. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. Journal of clinical epidemiology. 2016;79:70-5.

144. Gail MH. Does cardiac transplantation prolong life? A reassessment. Annals of internal medicine. 1972;76(5):815-7.

145. Suissa S. Immortal time bias in pharmacoepidemiology. American journal of epidemiology. 2008;167(4):492-9.

146. Dickerman BA, García-Albéniz X, Logan RW, Denaxas S, Hernán MA. Avoidable flaws in observational analyses: an application to statins and cancer. Nat Med. 2019;25(10):1601-6. doi:10.1038/s41591-019-0597-x

147. Adams RJ, Fuhlbrigge AL, Finkelstein JA, Weiss ST. Intranasal steroids and the risk of emergency department visits for asthma. Journal of Allergy and Clinical Immunology. 2002;109(4):636-42.

148. Donahue JG, Weiss ST, Livingston JM, Goetsch MA, Greineder DK, Platt R. Inhaled steroids and the risk of hospitalization for asthma. Jama. 1997;277(11):887-91.

149. Khurana V, Bejjanki HR, Caldito G, Owens MW. Statins reduce the risk of lung cancer in humans: a large case-control study of US veterans. Chest. 2007;131(5):1282-8.

150. Suissa S. Immortal time bias in observational studies of drug effects. Pharmacoepidemiology and drug safety. 2007;16(3):241-9.

151. Jones M, Fowler R. Immortal time bias in observational studies of time-to-event outcomes. J Crit Care. 2016;36:195-9. doi:10.1016/j.jcrc.2016.07.017

152. Bullock C, Ard K, Saalman G. Measuring the relationship between state environmental justice action and air pollution inequality, 1990–2009. Review of Policy Research. 2018;35(3):466-90.

153. McClure CD, Jaffe DA. US particulate matter air quality improves except in wildfire-prone areas. Proceedings of the National Academy of Sciences. 2018;115(31):7901-6.

154. Hernán MA. The hazards of hazard ratios. Epidemiology (Cambridge, Mass.). 2010;21(1):13-5. doi:10.1097/EDE.0b013e3181c1ea43

155. Buteau S, Hatzopoulou M, Crouse DL, et al. Comparison of spatiotemporal prediction models of daily exposure of individuals to ambient nitrogen dioxide and ozone in Montreal, Canada. Environmental research. 2017;156:201-30. doi:10.1016/j.envres.2017.03.017

156. Yost K, Perkins C, Cohen R, Morris C, Wright W. Socioeconomic status and breast cancer incidence in California for different race/ethnic groups. Cancer causes & control : CCC. 2001;12(8):703-11. doi:10.1023/a:1011240019516

157. Murray EJ, Caniglia EC, Petito LC. Causal survival analysis: A guide to estimating intention-to-treat and per-protocol effects from randomized clinical trials with non-adherence. Research Methods in Medicine & Health Sciences. 2021;2(1):39-49. doi:10.1177/2632084320961043

158. Austin PC. A Tutorial on Multilevel Survival Analysis: Methods, Models and Applications. International Statistical Review. 2017;85(2):185-203. doi:https://doi.org/10.1111/insr.12214

159. Mansournia MA, Nazemipour M, Etminan M. Causal diagrams for immortal time bias. International journal of epidemiology. 2021. doi:10.1093/ije/dyab157

160. Breskin A, Edmonds A, Cole SR, et al. G-computation for policy-relevant effects of interventions on time-to-event outcomes. International journal of epidemiology. 2020.

161. Keil AP, Edwards JK, Richardson DB, Naimi AI, Cole SR. The parametric g-formula for time-to-event data: intuition and a worked example. Epidemiology. 2014;25(6):889-97. doi:10.1097/ede.0000000000000160

162. Naimi AI, Cole SR, Kennedy EH. An introduction to g methods. International journal of epidemiology. 2016;46(2):756-62. doi:10.1093/ije/dyw323

163. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000;11(5):550-60. doi:10.1097/00001648-200009000-00011

164. He J-R, Liu Y, Xia X-Y, et al. Ambient Temperature and the Risk of Preterm Birth in Guangzhou, China (2001–2011). Environmental health perspectives. 2016;124(7):1100-6. doi:doi:10.1289/ehp.1509778

165. Bentayeb M, Wagner V, Stempfelet M, et al. Association between long-term exposure to air pollution and mortality in France: A 25-year follow-up study. Environment International. 2015;85:5-14. doi:https://doi.org/10.1016/j.envint.2015.08.006

166. Strand LB, Barnett AG, Tong S. Maternal Exposure to Ambient Temperature and the Risks of Preterm Birth and Stillbirth in Brisbane, Australia. American journal of epidemiology. 2011;175(2):99-107. doi:10.1093/aje/kwr404

167. Di Q, Wang Y, Zanobetti A, et al. Air Pollution and Mortality in the Medicare Population. The New England journal of medicine. 2017;376(26):2513-22. doi:10.1056/NEJMoa1702747

168. Tian F, Liu X, Chao Q, et al. Ambient air pollution and low temperature associated with case fatality of COVID-19: A nationwide retrospective cohort study in China. Innovation (N Y). 2021;2(3):100139. doi:10.1016/j.xinn.2021.100139

169. Acevedo-Garcia D, Lochner KA, Osypuk TL, Subramanian SV. Future Directions in Residential Segregation and Health Research: A Multilevel Approach. American journal of public health. 2003;93(2):215-21. doi:10.2105/ajph.93.2.215

170. Kramer MR, Hogue CR. Is segregation bad for your health? Epidemiologic reviews. 2009;31(1):178-94.

171. Krysan M, Crowder K. Cycle of segregation: Social processes and residential stratification: Russell Sage Foundation; 2017.

172. Yang T-C, Park K, Matthews SA. Racial/ethnic segregation and health disparities: Future directions and opportunities. Sociol Compass. 2020;14(6):e12794. doi:10.1111/soc4.12794

173. Caldwell JT, Ford CL, Wallace SP, Wang MC, Takahashi LM. Racial and ethnic residential segregation and access to health care in rural areas. Health & place. 2017;43:104-12. doi:10.1016/j.healthplace.2016.11.015

174. Woo B, Kravitz-Wirtz N, Sass V, Crowder K, Teixeira S, Takeuchi DT. Residential Segregation and Racial/Ethnic Disparities in Ambient Air Pollution. Race and Social Problems. 2019;11(1):60-7. doi:10.1007/s12552-018-9254-0

175. Xiao YK, Graham G. Where we live: The impact of neighborhoods and community factors on cardiovascular health in the United States. Clin Cardiol. 2019;42(1):184-9. doi:10.1002/clc.23107

176. Soneji S, Tanner NT, Silvestri GA, Lathan CS, Black W. Racial and Ethnic Disparities in Early-Stage Lung Cancer Survival. Chest. 2017;152(3):587-97. doi:10.1016/j.chest.2017.03.059

177. Morello-Frosch R, Jesdale BM. Separate and unequal: residential segregation and estimated cancer risks associated with ambient air toxics in US metropolitan areas. Environmental health perspectives. 2006;114(3):386-93.

178. Hicken MT, Adar SD, Hajat A, et al. Air Pollution, Cardiovascular Outcomes, and Social Disadvantage: The Multi-ethnic Study of Atherosclerosis. Epidemiology. 2016;27(1):42-50. doi:10.1097/ede.0000000000000367

179. Jones MR, Diez-Roux AV, Hajat A, et al. Race/ethnicity, residential segregation, and exposure to ambient air pollution: the Multi-Ethnic Study of Atherosclerosis (MESA). American journal of public health. 2014;104(11):2130-7. doi:10.2105/AJPH.2014.302135

180. Massey DS, Fischer MJ. How segregation concentrates poverty. Ethnic and racial studies. 2000;23(4):670-91.

181. Gaskin DJ, Dinwiddie GY, Chan KS, McCleary RR. Residential Segregation and the Availability of Primary Care Physicians. Health services research. 2012;47(6):2353-76. doi:https://doi.org/10.1111/j.1475-6773.2012.01417.x

182. Bailey ZD, Krieger N, Agénor M, Graves J, Linos N, Bassett MT. Structural racism and health inequities in the USA: evidence and interventions. The Lancet. 2017;389(10077):1453-63.

183. Elk R, Landrine H. Cancer disparities: causes and evidence-based solutions: Springer Publishing Company; 2012.

184. Woolf SH, Braveman P. Where health disparities begin: the role of social and economic determinants—and why current policies may make matters worse. Health affairs. 2011;30(10):1852-9.

185. Landrine H, Corral I, Lee JGL, Efird JT, Hall MB, Bess JJ. Residential Segregation and Racial Cancer Disparities: A Systematic Review. J Racial Ethn Health Disparities. 2017;4(6):1195-205. doi:10.1007/s40615-016-0326-9

186. Richards TB, Henley SJ, Puckett MC, et al. Lung cancer survival in the United States by race and stage (2001-2009): Findings from the CONCORD-2 study. Cancer. 2017;123 Suppl 24:5079-99. doi:10.1002/cncr.31029

187. Bach PB, Schrag D, Brawley OW, Galaznik A, Yakren S, Begg CB. Survival of blacks and whites after a cancer diagnosis. Jama. 2002;287(16):2106-13.

188. Ryan BM. Lung cancer health disparities. Carcinogenesis. 2018;39(6):741-51. doi:10.1093/carcin/bgy047

189. Wolf A, Alpert N, Tran BV, Liu B, Flores R, Taioli E. Persistence of racial disparities in early-stage lung cancer treatment. The Journal of thoracic and cardiovascular surgery. 2019;157(4):1670-9.e4. doi:10.1016/j.jtcvs.2018.11.108

190. Williams CD, Salama JK, Moghanaki D, Karas TZ, Kelley MJ. Impact of Race on Treatment and Survival among U.S. Veterans with Early-Stage Lung Cancer. Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer. 2016;11(10):1672-81. doi:10.1016/j.jtho.2016.05.030

191. Paro A, Dalmacy D, Madison Hyer J, Tsilimigras DI, Diaz A, Pawlik TM. Impact of Residential Racial Integration on Postoperative Outcomes Among Medicare Beneficiaries Undergoing Resection for Cancer. Annals of surgical oncology. 2021. doi:10.1245/s10434-021-10034-w

192. Oka M, Wong DWS. Capturing the two dimensions of residential segregation at the neighborhood level for health research. Frontiers in public health. 2014;2:118-. doi:10.3389/fpubh.2014.00118

193. Johnston R, Poulsen M, Forrest J. Ethnic and Racial Segregation in U.S. Metropolitan Areas, 1980-2000:The Dimensions of Segregation Revisited. Urban Affairs Review. 2007;42(4):479-504. doi:10.1177/1078087406292701

194. Krupka D. Are Big Cities More Segregated? Neighbourhood Scale and the Measurement of Segregation. Urban Studies. 2007;44:187-97. doi:10.1080/00420980601023828

195. Ellis L, Canchola AJ, Spiegel D, Ladabaum U, Haile R, Gomez SL. Racial and ethnic disparities in cancer survival: the contribution of tumor, sociodemographic, institutional, and neighborhood characteristics. Journal of Clinical Oncology. 2018;36(1):25.

196. Bureau UC. 2010 American Community Survey 5-year detailed tables. 2021.

197. Cohn MJ, Jackman SP. A comparison of aspatial and spatial measures of segregation. Transactions in GIS. 2011;15:47-66.

198. Yao J, Wong DW, Bailey N, Minton J. Spatial segregation measures: A methodological review. Tijdschrift voor economische en sociale geografie. 2019;110(3):235-50.

199. Morrill RL, editor. On the measure of geographic segregation. Geography research forum; 1991.

200. Wong DW. Formulating a General Spatial Segregation Measure. The Professional Geographer. 2005;57(2):285-94. doi:10.1111/j.0033-0124.2005.00478.x

201. Barros J, Feitosa FF. Uneven geographies: Exploring the sensitivity of spatial indices of residential segregation. Environment and Planning B: Urban Analytics and City Science. 2018;45(6):1073-89.

202. Kramer MR, Cooper HL, Drews-Botsch CD, Waller LA, Hogue CR. Do measures matter? Comparing surface-density-derived and census-tract-derived measures of racial residential segregation. International journal of health geographics. 2010;9(1):1-15.

203. Wong DW. Spatial dependency of segregation indices. Canadian Geographer/Le Geographe Canadien. 1997;41(2):128-36.

204. Allen R, Burgess S, Davidson R, Windmeijer F. More reliable inference for the dissimilarity index of segregation. Econom J. 2015;18(1):40-66. doi:10.1111/ectj.12039

205. Fotheringham AS, Wong DWS. The Modifiable Areal Unit Problem in Multivariate Statistical Analysis. Environment and Planning A: Economy and Space. 1991;23(7):1025-44. doi:10.1068/a231025

206. Kwan M-P. Geographies of health, disease and well-being: recent advances in theory and method. 2016.

207. Vieira VM, Villanueva C, Chang J, Ziogas A, Bristow RE. Impact of community disadvantage and air pollution burden on geographic disparities of ovarian cancer survival in California. Environmental research. 2017;156:388-93. doi:https://doi.org/10.1016/j.envres.2017.03.057

208. Johnson AM, Hines RB, Johnson JA, Bayakly AR. Treatment and survival disparities in lung cancer: The effect of social environment and place of residence. Lung cancer (Amsterdam, Netherlands). 2014;83(3):401-7. doi:https://doi.org/10.1016/j.lungcan.2014.01.008

209. Krieger N, Wright E, Chen JT, Waterman PD, Huntley ER, Arcaya M. Cancer Stage at Diagnosis, Historical Redlining, and Current Neighborhood Characteristics: Breast, Cervical, Lung, and Colorectal Cancer, Massachusetts, 2001-2015. American journal of epidemiology. 2020.

210. Patel MI, McKinley M, Cheng I, Haile R, Wakelee H, Gomez SL. Lung cancer incidence trends in California by race/ethnicity, histology, sex, and neighborhood socioeconomic status: An analysis spanning 28 years. Lung cancer (Amsterdam, Netherlands). 2017;108:140-9. doi:10.1016/j.lungcan.2017.03.014

211. Henneman LRF, Liu C, Mulholland JA, Russell AG. Evaluating the effectiveness of air quality regulations: A review of accountability studies and frameworks. Journal of the Air & Waste Management Association. 2017;67(2):144-72. doi:10.1080/10962247.2016.1242518

212. Giles LV, Barn P, Künzli N, et al. From Good Intentions to Proven Interventions: Effectiveness of Actions to Reduce the Health Impacts of Air Pollution. Environmental health perspectives. 2011;119(1):29-36. doi:doi:10.1289/ehp.1002246

213. Brulle RJ, Pellow DN. Environmental justice: human health and environmental inequalities. Annual review of public health. 2006;27:103-24. doi:10.1146/annurev.publhealth.27.021405.102124

214. Woo B, Kravitz-Wirtz N, Sass V, Crowder K, Teixeira S, Takeuchi DT. Residential Segregation and Racial/Ethnic Disparities in Ambient Air Pollution. Race Soc Probl. 2019;11(1):60-7. doi:10.1007/s12552-018-9254-0

215. Tessum CW, Apte JS, Goodkind AL, et al. Inequity in consumption of goods and services adds to racial–ethnic disparities in air pollution exposure. Proceedings of the National Academy of Sciences. 2019;116(13):6001-6. doi:10.1073/pnas.1818859116

216. Kurian AW, Mitani A, Desai M, et al. Breast cancer treatment across health care systems: linking electronic medical records and state registry data to enable outcomes research. Cancer. 2014;120(1):103-11.

217. Liu B, Lee FF, Boscoe F. Residential mobility among adult cancer survivors in the United States. BMC public health. 2020;20(1):1601. doi:10.1186/s12889-020-09686-2

218. Wiese D, Stroup AM, Maiti A, et al. Residential mobility and geospatial disparities in colon cancer survival. Cancer Epidemiology and Prevention Biomarkers. 2020;29(11):2119-25.

219. Bryere J, Pornet C, Dejardin O, Launay L, Guittet L, Launoy G. Correction of misclassification bias induced by the residential mobility in studies examining the link between socioeconomic environment and cancer incidence. Cancer epidemiology. 2015;39(2):256-64.

220. Kirby RS, Delmelle E, Eberth JM. Advances in spatial epidemiology and geographic information systems. Annals of Epidemiology. 2017;27(1):1-9. doi:https://doi.org/10.1016/j.annepidem.2016.12.001

221. Jay M, Oleson J, Charlton M, Arab A. A Bayesian approach for estimating age-adjusted rates for low-prevalence diseases over space and time. Statistics in medicine. 2021;40(12):2922-38. doi:https://doi.org/10.1002/sim.8948

222. Neyens T, Lawson AB, Kirby RS, et al. Disease mapping of zero-excessive mesothelioma data in Flanders. Annals of epidemiology. 2017;27(1):59-66. e3.

223. Miki Y, Inoue M, Ikeda A, et al. Neighborhood deprivation and risk of cancer incidence, mortality and survival: results from a population-based cohort study in Japan. PloS one. 2014;9(9):e106729.

224. Wiese D, Stroup AM, Maiti A, et al. Measuring Neighborhood Landscapes: Associations between a Neighborhood's Landscape Characteristics and Colon Cancer Survival. International journal of environmental research and public health. 2021;18(9):4728.

225. Russell E, Kramer MR, Cooper HL, Thompson WW, Arriola KRJ. Residential racial composition, spatial access to care, and breast cancer mortality among women in Georgia. Journal of Urban Health. 2011;88(6):1117-29.

226. Warren GW, Sobus S, Gritz ER. The biological and clinical effects of smoking by patients with cancer and strategies to implement evidence-based tobacco cessation support. The Lancet Oncology. 2014;15(12):e568-e80.

227. Tanner NT, Kanodra NM, Gebregziabher M, et al. The association between smoking abstinence and mortality in the National Lung Screening Trial. American journal of respiratory and critical care medicine. 2016;193(5):534-41.

228. Shiono S, Katahira M, Abiko M, Sato T. Smoking is a perioperative risk factor and prognostic factor for lung cancer surgery. General thoracic and cardiovascular surgery. 2015;63(2):93-8. doi:10.1007/s11748-014-0461-3

229. Petros WP, Younis IR, Ford JN, Weed SA. Effects of tobacco smoking and nicotine on cancer treatment. Pharmacotherapy. 2012;32(10):920-31. doi:10.1002/j.1875-9114.2012.01117

230. Centers for Disease Control and Prevention SaTU. 2014 Surgeon General's Report: The Health Consequences of Smoking - 50 Years of Progress2014.

231. Stedman RL. The chemical composition of tobacco and tobacco smoke. Chem Rev. 1968;68(2):153-207. doi:10.1021/cr60252a002

232. Talhout R, Schulz T, Florek E, van Benthem J, Wester P, Opperhuizen A. Hazardous compounds in tobacco smoke. Int J Environ Res Public Health. 2011;8(2):613-28. doi:10.3390/ijerph8020613

233. Zevin S, Benowitz NL. Drug interactions with tobacco smoking. An update. Clin Pharmacokinet. 1999;36(6):425-38. doi:10.2165/00003088-199936060-00004

234. Cooley ME, Sipples RL, Murphy M, Sarna L. Smoking cessation and lung cancer: oncology nurses can make a difference. Semin Oncol Nurs. 2008;24(1):16-26. doi:10.1016/j.soncn.2007.11.008

235. Tucker MA, Murray N, Shaw EG, et al. Second Primary Cancers Related to Smoking and Treatment of Small-Cell Lung Cancer. JNCI: Journal of the National Cancer Institute. 1997;89(23):1782-8. doi:10.1093/jnci/89.23.1782

236. Fox JL, Rosenzweig KE, Ostroff JS. The effect of smoking status on survival following radiation therapy for non-small cell lung cancer. Lung cancer (Amsterdam, Netherlands). 2004;44(3):287-93. doi:10.1016/j.lungcan.2003.11.012

237. Moller AM, Villebro N, Pedersen T, Tonnesen H. Effect of preoperative smoking intervention on postoperative complications: a randomised clinical trial. Lancet (London, England). 2002;359(9301):114-7. doi:10.1016/s0140-6736(02)07369-5

238. Trousse D, Barlesi F, Loundou A, et al. Synchronous multiple primary lung cancer: an increasing clinical occurrence requiring multidisciplinary management. The Journal of thoracic and cardiovascular surgery. 2007;133(5):1193-200. doi:10.1016/j.jtcvs.2007.01.012

239. Wildeman C, Wang EA. Mass incarceration, public health, and widening inequality in the USA. Lancet (London, England). 2017;389(10077):1464-74. doi:10.1016/s0140-6736(17)30259-3