

**Using Residential Histories to Investigate the
Role of Segregation in Racial Cancer Inequities**

BY

EMMA E. BOYLAN

B.A., University of Illinois at Urbana-Champaign, 2009
M.S., University of Illinois at Chicago, 2016

THESIS

Submitted as partial fulfillment of the requirements
for the Doctor of Philosophy in Epidemiology
in the Graduate College of the
University of Illinois at Chicago, 2023

Chicago, Illinois

Defense Committee:

Vincent Freeman, Chair and Advisor
Sanjib Basu, Biostatistics
Lisa Sharp, Pharmacy Systems, Outcomes, and Policy
Leslie Stayner, Epidemiology
Nebiyu Tilahun, Urban Planning and Policy

ACKNOWLEDGMENTS

This work would not have been possible without the contributions of many others.

My dissertation chair, Vincent Freeman, MD, MPH, has supported me tirelessly since I began studying epidemiology. His insights and advice were invaluable not only to every aspect of this study, but also to our other work together on the relationship between place and health, which deeply informed this research. Dr. Freeman secured funding and data for the study conducted using the University of Illinois Hospital tumor registry, conducted the matching process with LexisNexis and the National Death Index, led research on residential history data quality that formed the foundation for this dissertation, first suggested to me that surveys such as the Panel Study of Income Dynamics (PSID) might present an alternative, and supported me during a very personally challenging period of this program. Clearly, this research would not have occurred without him. I am lucky to have worked with him and to have benefited from his mentorship.

The contributions from each member of my committee greatly improved this research. Sanjib Basu, PhD, and Leslie Stayner, PhD, provided helpful advice about how to analyze the unusual data sources required for this research and clearly communicate the chosen methods. Lisa Sharp, MA, PhD, was always thoughtful, encouraging, and informative as I planned this study and thought through the possible interpretations of its results. With Dr. Freeman, she also supported my applications to fund this research during the summer and fall of 2020. Nebiyu Tilahun, PhD, provided valuable advice from his own experience in working with the very complex PSID data. I am especially grateful for his sharing tips and readings that helped me understand relevant points of the sample design.

Xiaohan Mei, Yan Gao, PhD, and Ifeanyi Beverly Chukwudozie, MBA, MPH, all contributed to the preparation and analysis of data used in this study. Beverly and Yan helped assemble the historical Census data used for both studies presented here, and linked it to records from the University of Illinois Hospital tumor registry. Xiaohan and Yan performed important preliminary

ACKNOWLEDGMENTS (CONTINUED)

analyses of the patient data and histories that informed my own interpretation and made my job much easier. Beverly's friendship and support during our regular Zoom writing sessions broke the isolation of 2020 and beyond, and made it possible to push this project forward. So important has Beverly been to my completing this project, she is actually on Zoom with me now as I write these words.

Many individuals I worked with during this program provided personal and professional support for which I am very grateful. Amber Uskali, MPH, and Lisa Powell, PhD, of the Policy, Practice, and Prevention Research Center (P3RC) supervised me from 2015-2022 and were endlessly supportive, flexible, and kind. I learned so much from working with them that has informed my work in public health. My colleagues at the Chicago Department of Public Health, especially in Chronic Disease Prevention & Health Promotion and the Office of Epidemiology, were a constant source of encouragement and perspective.

Multiple organizations provided funding and related support for this project. The P3RC, which is funded by the CDC Prevention Research Centers program, provided my tuition support from 2015-2022. The Institute for Research on Race and Public Policy at UIC provided dissertation grant funding that allowed me to access the PSID restricted datasets. The University of Illinois Cancer Center funded the search for commercial residential histories used in Chapter V. The Office of the Dean at UIC School of Public Health provided funding support for the death certificate matching done in the same chapter. Mary Arey MBA, MSN, RN and Karen Snitchler, RHIT, CTR assisted with data collection from the University of Illinois Hospital Tumor Registry. Many staff and investigators contributed to the design, data collection, distribution, and extensive documentation of the PSID. The collection of PSID data used in this study was partly supported by the National Institutes of Health under grant number R01 HD069609 and R01 AG040213, and the National Science Foundation under award numbers SES 1157698 and 1623684. Their support is gratefully acknowledged.

ACKNOWLEDGMENTS (CONTINUED)

My parents, John and Jennifer Boylan, and my aunt, Nancy L. Webb, made graduate school sound fun and then never pressed me for details when I was in the thick of dissertation research. Their support meant that I always knew I would finish, even when it felt unlikely. Samantha Evy Schoeneman Patel, MD has believed in me since long before I knew how to do anything this involved. I am forever grateful for her friendship and encouragement.

The contributions of my partner, Matthew Strong, continue to defy description. I am so grateful for his love and support, which transcend this or any project.

EEB

TABLE OF CONTENTS

<u>CHAPTER</u>	<u>PAGE</u>
I. INTRODUCTION	1
A. Background	1
B. Research Objective	2
C. Hypothesis	2
D. Study Aims	3
1. Aim 1	3
2. Aim 2	3
3. Aim 3	3
E. Evidence Linking Segregation and Racial Cancer Inequities	3
F. Barriers to Studying Fundamental Causes of Cancer Inequities	7
G. Research Contribution	8
II. METHODS	11
A. Analytic Approach	11
B. Data	12
1. Panel Study of Income Dynamics	12
2. University of Illinois Cancer Center Tumor Registry	13
3. LexisNexis Accurint	13
4. Census Data	16
5. National Death Index	16
6. Data Preparation	16
i. Panel Study of Income Dynamics	16
ii. Census Data	17
C. Statistical Analysis	18
1. Panel Study of Income Dynamics	18
2. University of Illinois Hospital Tumor Registry	19
III. EXPOSURE TO NEIGHBORHOOD CONCENTRATED DISADVANTAGE ACROSS THE LIFE COURSE AND SURVIVAL AFTER CANCER DIAGNOSIS IN THE PANEL STUDY OF INCOME DYNAMICS COHORT	20
A. Introduction	20
B. Materials and Methods	21
1. Primary Data Source	21
2. Sampling Frame	22
3. Vital Status Ascertainment	24
4. Residential Exposure Measurement	24
5. Data Preparation	25
6. Statistical Analysis	25
C. Results	27
1. Study Participant Characteristics	27
2. Distribution and Availability of CD Score	27
3. Survival Analysis	27
D. Discussion	32
E. Conclusion	36

TABLE OF CONTENTS (continued)

<u>CHAPTER</u>	<u>PAGE</u>
IV. PLANNED MEDIATION ANALYSIS IN THE PSID COHORT IS NOT JUSTIFIED BY PRIOR RESULTS	38
V. NEIGHBORHOOD CONCENTRATED DISADVANTAGE PREDICTS SURVIVAL BUT NOT STAGE AT DIAGNOSIS IN A COLORECTAL CANCER COHORT	39
A. Introduction.....	39
B. Methods.....	40
1. Setting.....	40
2. Case Data, Ascertainment and Inclusion Criteria	40
3. Tumor Characteristics	41
4. Residential History Data Collection and Preparation.....	41
5. Statistical Analysis.....	44
C. Results.....	45
D. Discussion.....	55
E. Conclusion	57
VI. CONCLUSION	58
A. Summary of Findings	58
B. Comparison of Data and Results.....	59
1. Lack of Observed Racial Disparities in Either Data Set.....	59
2. Role of Residential Histories in Producing Selection or Attrition Bias..	59
C. Role of Missing Case Data.....	61
1. Panel Study of Income Dynamics.....	61
i. Death Data.....	61
ii. Exact Dates and Self-Reported Diagnosis Timing.....	61
iii. Disease Information	62
2. University of Illinois Hospital Tumor Registry.....	63
D. Role of Exposure Specification	64
E. Future Directions.....	66
1. Using PSID for Social Epidemiology	66
2. Weighing the Probability of Data Loss and Selection Bias When Using Residential Histories to Study Marginalized Populations.....	67
3. Acting on Evidence That Neighborhood Disadvantage at Diagnosis Contributes to Preventable Cancer Deaths Due to Systemic Racism.....	67
CITED LITERATURE	69
VITA.....	78

LIST OF TABLES

<u>TABLE</u>	<u>PAGE</u>
I. VARIABLES FROM THE PSID PUBLIC DATA INDEX	14
II. CENSUS AND AMERICAN COMMUNITY SURVEY VARIABLES USED TO CONSTRUCT THE CONCENTRATED DISADVANTAGE SCORE.....	17
III. CHARACTERISTICS OF PSID PARTICIPANTS WHO REPORTED A HISTORY OF BREAST, COLORECTAL, LUNG, OR PROSTATE CANCER, BY VITAL STATUS, 2005-2019.....	28
IV. DISTRIBUTIONS AND AVAILABILITY OF CD SCORES AMONG PSID PARTICIPANTS WHO REPORTED A HISTORY OF BREAST, COLORECTAL, LUNG, OR PROSTATE CANCER, 2005-2019.....	29
V. ADJUSTED HAZARD RATIOS FOR THE ASSOCIATION OF CENSUS TRACT- LEVEL CONCENTRATED DISADVANTAGE AT DIFFERENT POINTS IN TIME AND ALL-CAUSE MORTALITY AMONG PSID PARTICIPANTS WHO REPORTED A HISTORY OF BREAST, COLORECTAL, LUNG, OR PROSTATE CANCER, 2005- 2019†	31
VI. CHARACTERISTICS OF UI HEALTH PATIENTS DIAGNOSED WITH COLORECTAL CANCER, 1995-2004, BY VITAL STATUS.....	47
VII. CHARACTERISTICS OF UI HEALTH PATIENTS DIAGNOSED WITH COLORECTAL CANCER, 1995-2004, BY CONCENTRATED DISADVANTAGE AT DIAGNOSIS	48
VIII. ASSOCIATION BETWEEN TUMOR CHARACTERISTICS AND RACE/ETHNICITY AMONG PEOPLE DIAGNOSED WITH COLORECTAL CANCER AT UI HEALTH, 1995-2004.....	49
IX. MULTINOMIAL MODEL 1: PREDICTORS OF STAGE AT DIAGNOSIS WITH LAGGED CONCENTRATED DISADVANTAGE AMONG PEOPLE DIAGNOSED WITH COLORECTAL CANCER AT UI HEALTH, 1995-2004.....	50
X. MULTINOMIAL MODEL 2: PREDICTORS OF STAGE AT DIAGNOSIS WITH CONTEMPORARY CONCENTRATED DISADVANTAGE AMONG PEOPLE DIAGNOSED WITH COLORECTAL CANCER AT UI HEALTH, 1995-2004.....	50
XI. LOGISTIC REGRESSION MODELS 1-2: PREDICTORS OF ADVANCED STAGE AT DIAGNOSIS AMONG PEOPLE DIAGNOSED WITH COLORECTAL CANCER AT UI HEALTH, 1995-2004, BY CHOICE OF CONCENTRATED DISADVANTAGE SCORE	52
XII. SURVIVAL MODELS 1-2: PREDICTORS OF DISEASE-SPECIFIC DEATH AMONG PEOPLE DIAGNOSED WITH COLORECTAL CANCER AT UI HEALTH, 1995-2004, BY CHOICE OF CONCENTRATED DISADVANTAGE SCORE	53

LIST OF FIGURES

<u>FIGURE</u>	<u>PAGE</u>
1. Selection criteria and sample history applied to the Panel Study of Income Dynamics data set.....	23
2. Kaplan-Meier survival estimates stratified by mean historical concentrated disadvantage among PSID participants with each type of cancer: breast (A), prostate (B), colorectal (C), or lung (D).....	30
3. Relationship between CD score and survival at each point in time for PSID participants with a history of breast, colorectal, lung, or prostate cancer	32
4. Selection of the analytic cohort.....	46
5. Distribution of cancer stage by race among UI Health patients diagnosed with colorectal cancer, 1995-2004	46
6. Kaplan-Meier survival estimates stratified by stage at diagnosis among UI Health patients diagnosed with colorectal cancer, 1995-2004.....	51
7. Trend in association between CD scores at different lag times and hazard of CRC-specific death in UI Health patients diagnosed with colorectal cancer, 1995-2004.....	54

LIST OF ABBREVIATIONS

ACS	American Community Survey
AJCC	American Joint Committee on Cancer
CD	Concentrated disadvantage
CI	Confidence interval
CPI-U	Consumer Price Index for All Urban Consumers
CRC	Colorectal cancer
EHR	Electronic health record
HR	Hazard ratio
ICD-O-3	International Classification of Diseases for Oncology 3 rd Edition
IQR	Interquartile range
LTDB	Longitudinal Tract Database
NDI	National Death Index
NHIS	National Health Interview Survey
NHW	Non-Hispanic White
NHB	Non-Hispanic Black
OR	Odds ratio
PSID	Panel Study of Income Dynamics
SD	Standard deviation
SEER	Surveillance, Epidemiology, and End Results Program
SES	Socioeconomic status

SUMMARY

Segregation and neighborhood socioeconomic disadvantage are associated with racial disparities in cancer prognosis and survival. However, limited availability of historical data means that few epidemiologists have been able to study the health effects of experiencing either segregation or neighborhood disadvantage in the past. To address this need, this study explores the use of residential history data obtained in two settings among people with cancer. In the Panel Study of Income Dynamics sample, participants in a long-running sociological survey who reported a personal history of cancer were studied. Residential history data were obtained from the past addresses used to contact participants for the study and used to estimate the degree of neighborhood disadvantage that participants experienced before being diagnosed with cancer. The outcome of interest was all cause mortality. In the University of Illinois Hospital sample, people diagnosed with colorectal cancer between 1995-2004 were studied. Disease and demographic data were obtained from the hospital tumor registry, while residential history data were obtained from a commercial public records database. Because more detailed death certificate data were available in this group, the outcome of interest was cancer-specific mortality.

Results appeared to confirm prior research indicating that residential history data are systematically missing in people who are any race or ethnicity other than non-Hispanic white, who are older, who have died, or who have lower incomes. This had previously been demonstrated for commercial public records databases only. Both study samples, when limited to participants who had between 10 and 20 years of residential history data available, were not representative of racial inequities in cancer prognosis and survival prevailing in the general population at the time the data were collected. Once each dataset was prepared, there were no racial inequities in stage at diagnosis, all-cause mortality, or cancer-specific mortality among the included cases.

Among PSID participants, there was no evidence that neighborhood concentrated disadvantage was associated with mortality. Data from the PSID may be useful to social or health

SUMMARY (CONTINUED)

researchers interested in studying more common conditions or using a more limited time period of residential history. Lessons learned from working with this complex dataset are presented.

In the University of Illinois Hospital sample, neighborhood concentrated disadvantage was not associated with cancer stage at diagnosis, regardless of when it was measured. Neighborhood concentrated disadvantage did predict cancer-specific mortality risk. However, the effect was strongest when the neighborhood environment was measured around the time of diagnosis.

Reliance on commercial residential history data resulted in restriction of the study sample without adding new prognostic information.

This study concludes with a comparison of the quality, availability, and limitations of residential history data for researchers interested in the relationship between socioeconomic conditions and health.

I. INTRODUCTION

A. Background

Racial inequities in the risk of suffering and death from colorectal cancer (CRC) are preventable and caused primarily by racism. Present-day racial disparities emerged in the US within living memory: until the late 1970s, CRC-specific mortality was higher in whites (1). Today, non-Hispanic Black Americans are burdened by CRC incidence that is 20% greater, and CRC mortality more than 40% greater, than that experienced by non-Hispanic white people (1). These disparities occur across the cancer continuum and reflect differences in stress and discrimination, exposures, access to resources, attitudes, and beliefs across the life course. In the US, non-Hispanic Black people are also more segregated than other groups regardless of individual socioeconomic status (SES). As a result, segregation, and the conditions of segregated neighborhoods, may be more important to the health of non-Hispanic Black people and contribute to health disparities affecting them (2). Geographic variation in CRC burden operates at multiple scales, in patterns that reflect racial segregation (3–7). Segregation is known to contribute to racial inequities in other health conditions and behaviors, including obesity and smoking, which are related to CRC; but its effect is strongest at the local level and may be state-specific (2). However, further research is needed to confirm that racial segregation contributes to CRC as it does to other common cancers; and to identify features of neighborhoods that mediate the relationship between segregation and cancer. Neighborhood-level variation in the burden of CRC and other noncommunicable diseases may be caused by social and economic disadvantage, which is both caused by segregation and has been shown to interact with segregation in harming health (2,8–10). A major obstacle to this research is that, typically, epidemiologists do not know where people with cancer lived before diagnosis. As a result, racial disparities in cancer burden have not been definitively linked to neighborhood

conditions at the time the disease began to develop. We also do not know if there is a critical time period or dose at which segregation and its effects are especially harmful.

B. Research Objective

To address this need, this research evaluates the relationship between neighborhood disadvantage experienced across the life course and CRC outcomes. This research addresses CRC and residential history in two settings: the Panel Study of Income Dynamics (PSID), and the UI Health tumor registry. The PSID is a longitudinal, nationally representative study that has collected data on individuals and households since 1968. The PSID includes rich information about participants' income, education, and reasons for moving; can be linked to Census data; and is linked to the National Death Index (NDI). Residential data in the PSID were collected prospectively and used to contact participants, avoiding concerns about data quality and bias that may affect residential histories collected from electronic health records (EHRs) and commercial data sources (11–14). Disease prevalence estimates in the PSID cohort are comparable to those obtained from national public health surveys (15,16). However, the PSID data have not previously been used to study determinants of cancer incidence and outcomes. People treated at UI Health are overwhelmingly residents of Chicago and greater Cook County, IL. This study uses a measure of neighborhood disadvantage that was developed in Chicago and that has been shown to both predict outcomes and explain racial disparities in other health conditions (9,10,17,18).

C. Hypothesis

I hypothesize that having been exposed to neighborhood disadvantage in the past will explain a significant portion of variation in CRC prognosis and mortality. I also expect to find that the effect of past neighborhood environment is partially mediated by the neighborhood environment at diagnosis, because many individuals will live in similar environments across the life course due to segregation. This study will provide information about how exposure to

neighborhood disadvantage at the time cancer is developing may influence present-day outcomes in a large nationally representative sample.

D. Study Aims

1. Aim 1

Evaluate the association between past neighborhood disadvantage and all cause mortality among people with cancer in the PSID cohort. Using PSID data will eliminate likely ascertainment bias from financial data sources, while providing rich supplementary information about residential histories including whether moves were voluntary.

2. Aim 2

Conduct a mediation analysis to determine whether neighborhood disadvantage at diagnosis mediates the relationship between past neighborhood disadvantage and all cause mortality among people with cancer in the PSID cohort. This analysis will provide an initial estimate of whether residences in the distant past can affect present-day cancer outcomes; and if so, how much of their effect is contained in the address at diagnosis.

3. Aim 3

Evaluate the association between neighborhood disadvantage in the past and CRC stage at diagnosis in a cohort of cases treated at UI Health. Conduct an exploratory mediation analysis to assess whether residential CD at diagnosis mediates this relationship.

E. Evidence Linking Segregation and Racial Cancer Inequities

Available evidence strongly suggests that unequal suffering and death due to CRC is preventable. In the US and globally, CRC burden varies both geographically and by economic development (3,4). Non-Hispanic Black people are unfairly burdened at every stage of the cancer continuum: they experience higher incidence than white people, are less likely to be screened or referred for screening, are diagnosed at later stages, are less likely to be referred for appropriate

treatment, and are more likely to die from their cancer (1,19–26). No genetic or biological factor can explain the severity of today's inequities, which are only about 40 years old. Once diagnosed with CRC, non-Hispanic Black and white patients respond similarly to appropriate treatment (1). In equal access health systems, racial disparities in screening, treatment, and survival are reduced or eliminated (27–31). This pattern has also been observed in use of screening and in other cancer sites (27,32). In addition, racial and ethnic disparities reflect the fact that CRC itself is highly preventable. Diet, body composition, smoking, and physical activity all affect individual risk of CRC (1). Screening colonoscopy also has the potential to prevent CRC via removal of precancerous polyps; more than 90% of CRCs are adenocarcinomas that arise via this pathway (1,33). The burden of CRC and its associated health behaviors vary geographically across the US and in Chicago (1,34–36). These patterns reflect racial segregation and the disproportionate cancer burden that is borne by non-Hispanic Black people as a result.

Racial segregation is a fundamental cause of health disparities because it causes individual- and area-level poverty, and physical and social isolation (37,38). Segregation results in isolation from job opportunities and institutional resources such as health care. Employers and service providers may avoid areas where many Black people live, effectively denying access to many people based on their race. Even when made without an explicit racial motivation, these decisions are informed by racist stereotypes to which segregation contributes. In the case of public institutions like schools, which must serve segregated areas, segregation enables separate but unequal and inferior treatment (37). Historic segregation has reduced Black families' access to wealth through homeownership and continues to contribute to disparities in net worth today. Ongoing processes of segregation widen the gap, causing present-day black homeowners to realize inferior returns to their purchase of a home (13,37). By causing socioeconomic disadvantage in individuals, segregation contributes to a circumstance that powerfully influences people's health behavior, environmental exposures, access to health care, and chance to live a full and healthy life.

Segregation also affects health behavior more directly, by damaging the environments in which people must live. In recent years, public health researchers have demonstrated the effects of neighborhood environments on factors that affect cancer risk including physical activity, access to healthy food and health care, and other health resources and behaviors (2,39). People living in highly segregated neighborhoods are exposed to more advertising for alcohol and tobacco, more crime, poorer infrastructure, and fewer healthy food options (2,40). As Williams and Collins point out, risky health behaviors such as tobacco use “are coping strategies that are frequently employed to obtain escape and relief from the personal suffering and deprivation that characterizes many disadvantaged environments” (37). Meanwhile, people in segregated neighborhoods face greater barriers to adopting healthy behaviors due to lack of access to resources. No wonder segregation is associated with racial disparities in both chronic and infectious diseases, self-rated health, and all-cause mortality (2).

One important measure of the neighborhood environment is concentrated disadvantage (CD), an index of multiple Census measures related to area socioeconomic status (SES) such as employment, education, and share of people living in poverty (9,10). People who live in more socioeconomically disadvantaged neighborhoods in Chicago have lower self-rated health, are more likely to be diagnosed with HIV, and have lower life expectancy (9,10,17,18,41). In highly segregated urban areas, like Chicago, African Americans are disproportionately exposed to neighborhood disadvantage. As a result, neighborhood concentrated disadvantage contributes to racial disparities in cancer burden and death. Deprivation, enabled by segregation and represented by concentrated disadvantage, is one way that racism kills.

While measures of area-level SES and environmental quality can capture this injustice in the aggregate, they are much more than ecological substitutes for individual deprivation. Segregation causes inequities in neighborhood quality that are independent of individual SES, and these inequities disproportionately affect Black people regardless of their individual resources. This is

because, in the US, segregation of Black people is severe and distinctive. Non-Hispanic Black people are more concentrated and isolated by racial segregation than other groups that live or have lived in ethnic enclaves (37). Racial segregation does not reflect the preferences of Black people, who express support for integrated neighborhoods and search for homes in neighborhoods that match their stated preferences (2,37,42). Nor are patterns of segregation simply an artifact of disparities in individual SES. Even when holding income, education, and wealth equal, non-Hispanic Black people are housed in poorer and less integrated neighborhoods than other racial and ethnic groups through a process called locational attainment (43,44).

Locational attainment refers to how individuals' characteristics, such as race, nationality, education, language, income, or wealth, result in their being able to live in neighborhoods with more resources. Locational attainment theories explain how neighborhood concentrated disadvantage can produce racial health disparities that are not fully accounted for by individual SES: race and ethnicity influence where people are able to live, even after controlling for their individual education, income, and wealth (43,44). Furthermore, the effect of individual SES on neighborhood attainment varies by individual race and ethnicity. Regardless of their individual or household resources, African Americans are disproportionately likely to be exposed to neighborhood disadvantage. Therefore, locational attainment research suggests that the series of neighborhood environments a person lives in are related, because each is predicted by unchangeable personal traits including race. This raises the possibility that, particularly for disparities epidemiologists, the neighborhoods people lived in when diagnosed with cancer may have a legitimate interpretation reflecting a lifetime of cumulative exposure to disadvantage. However, this possibility needs to be investigated in a cohort that does not systematically exclude the vulnerable populations we want to understand.

F. Barriers to Studying Fundamental Causes of Cancer Inequities

A significant barrier to this research has been lack of information about where people with cancer lived in the past. Knowing where a person has lived unlocks valuable information, not only about their past exposure to environmental carcinogens; but about the social, policy, and built environments in which they have lived. Early life exposures, and exposures and behaviors specific to a person's stage of development, are known to affect the risk of common cancers, most notably breast cancer (45). Lack of residential history data severely limits cancer cluster investigations by inducing misclassification bias and reducing power in studies that are already technically challenging and often under-powered (46,47). It may also lead to exposure misclassification in studies that attempt to model residential environmental exposures, although the effect on these investigations may be less pronounced depending on the exposure of interest and the setting (48).

Recently, multiple investigators have demonstrated the feasibility of finding cases' residential histories—where they lived and when they lived there—using either electronic health records (EHRs) or commercial public records data (12,49–51). These researchers have rightly pointed out that even incomplete residential history data represent a significant improvement over the typical approach epidemiologists must use: assuming that cases have always lived at their only known address (49,51). However, both data sources have limitations that may be particularly important to disparities researchers. EHR addresses may be limited to a single health care system, and updated only when patients interact with it in some way (11,12). Therefore, residential histories constructed using these data may be less complete and accurate in people who are less apt to receive regular care. They are also unlikely to be relevant in registry-based studies. Commercial credit reporting data is known to be less complete in African American and Hispanic adults, as well as in people with lower incomes or who live in low income neighborhoods (13). Racial disparities in credit scores perpetuate residential discrimination and contribute to the segregation and neighborhood disadvantage that disparities researchers aim to remedy (13,14,52).

As a result, users of the two best-known sources of residential history data are left back where we started: assuming that the residential environments we know about are representative of those we don't. This problem is especially important for health disparities researchers, who risk mismeasuring the environments of the most disadvantaged members of the population. However, the problem is serious for environmental epidemiologists as well. Segregation and environmental racism must also be taken into account when considering who is likely to be exposed to environmental carcinogens; those at greatest risk of exposure may also be at greatest risk of having too little residential history data to analyze. This situation renews, rather than answers, a long-standing question in the field: how are most people's present and past environments related, and how much information about past residential environments can be inferred from the current one? It also raises new theoretical and methodological questions for epidemiologists interested in segregation and neighborhood disadvantage. How should we conceptualize lifetime exposure to disadvantage—as a cumulative or time-varying dose? As salient at particular times in the human life course or along the cancer continuum?

G. Research Contribution

Working with data from the PSID provides disparities researchers with a unique opportunity to learn about the effect of SES on health throughout the life course. The PSID is designed to be nationally representative, and this extends to health information collected from participants. Prevalence estimates obtained from analyses of PSID data are comparable to those from recognized public health sources such as the National Health Interview Survey (NHIS) (15,16,53). While inconsistencies can be found in the self-reported cancer data in the PSID, these errors are random across sociodemographic groups and are expected to reduce power without inducing bias (53).

Because participants are surveyed every 1-2 years, their addresses are collected prospectively and validated by being used to conduct the survey. Some adult participants were also

members of participating households as children, meaning their residential histories may be able to be reconstructed for most of their lives. For many participants who have been diagnosed with cancer, data about their neighborhoods and families is available from the time they likely began to develop cancer. For almost any participant with cancer, data about their neighborhoods and families is available for part of the period of disease latency.

As a result, using location data from the PSID allows evaluation of the relationship between past neighborhood disadvantage and cancer outcomes, without assuming that subjects have never moved, and without the risk of selection or information bias presented by using commercial credit data sources and EHRs (13,50). Because neighborhood CD can be calculated for every year that a subject participated and had a geocodable address, it is also possible to use this dataset to conduct a sensitivity analysis to explore the best formulation of disadvantage across the life course. Having full residential histories also enables the mediation analysis in Aim 2. Knowing whether and how residential environments are related across the life course in people with cancer may also inform the interpretation of results from other epidemiological studies, the vast majority of which lack access to residential histories.

Aim 3 uses data from the UI Health tumor registry to address conceptual and methodological limitations of working with the PSID data. There are benefits to conducting research on the health effects of segregation and disadvantage in a single urban center. In their systematic review of the relationship between segregation and cancer disparities, Landrine, et al. recommended that epidemiologists conduct their investigations within single states because degree of segregation, and the specific policies used to accomplish it, vary by jurisdiction (2). It is of particular importance, then, that the measure of neighborhood disadvantage used in this study was also developed in Chicago and has been shown to predict severity and disparities of multiple health conditions in this metropolitan region (9,10,17,18). Diagnosis data will also be more complete and accurate in this cohort because it will come from the medical record, not from self-reports. Among

PSID respondents who reported that they had been diagnosed with cancer, the date of diagnosis was the most frequent inconsistent response (53).

II. METHODS

A. Analytic Approach

The selection and design of the research aims for this project were intended to support one another in answering the overarching questions of whether and how past exposure to neighborhood disadvantage influences cancer survival; and how to interpret much more readily available information about neighborhood environment at diagnosis in light of this.

Aim 1 addressed the effect of past neighborhood disadvantage directly by attempting to use all available information about past CD on survival after diagnosis with cancer in a nationally representative sample. In Aim 2, I planned to explore how the relationship between past CD and survival, if any, relates to the established relationship between cancer survival and neighborhood CD at diagnosis. Neighborhood CD at diagnosis may act as a proxy for past environments if people tend to live in similar neighborhoods over time; it may mediate the relationship between past environment and future survival because it is both the result of past life circumstances and a known determinant of cancer outcomes. If either of these situations holds, it could inform the interpretation of neighborhood CD at diagnosis in the majority of situations where full residential histories are not available.

After being granted access to the PSID restricted data sets, I learned that changes in the study's contract prevented them from releasing death certificate data to outside researchers (personal email from PSIDhelp@umich.edu, November 2021). The staff of the PSID collect detailed information to attempt to re-contact study attritors or to submit to NDI to confirm when participants have died. This information includes survey variables related to non-response, communications from surviving household members, and obituaries (54). The mortality file available for this research included the month and year of death from these sources, which were part of the matching information sent to NDI. Available documentation indicate that the PSID was

able to obtain generally high quality matches using this approach. However, because this data set did not include the anticipated information about cause of death, the outcome of interest for Aims 1 and 2 was changed to all-cause mortality.

Finally, Aim 3 addressed limitations of using a national and non-clinical dataset like the PSID for cancer epidemiology. Cancer data in the PSID are self-reported and do not include information about cancer stage at diagnosis, a critical prognostic indicator. In addition, there is evidence of measurement error in self-reported dates of cancer diagnosis, although these errors are randomly distributed across social and demographic groups (53). The data source for Aim 3, the UI Health tumor registry, includes dates of diagnosis, anatomic sites, and stages at diagnosis derived from the medical records of people diagnosed with cancer. The tumor registry study was not affected by challenges to NDI data access, allowing us to ascertain whether deaths were due to cancer. The restriction to a single geographic area inherent in this data source may also be a strength. Because patterns of local and regional segregation vary due to the different policies that contributed to them, some researchers have argued that research on the health effects of segregation should be conducted within single jurisdictions (2).

B. Data

1. Panel Study of Income Dynamics

Exposure, outcome, and covariate data for Aims 1 and 2 were obtained from the PSID and its restricted linked datasets: Census tract characteristics of participants' residences, and the mortality file for participants who have died.

All PSID respondents who self-reported having been diagnosed with cancer as an adult, and have analyzable data, as of the 2019 data release, were included. This question was asked in each survey wave since 1999, but questions about cancer site were not added until 2005. Therefore, this analysis used only reports of a previous cancer diagnosis made in the 2005 wave or later.

Participants were asked if they had ever been diagnosed with cancer, resulting in reports of cancers

diagnosed before 2005. Self-reported tumor site, approximate timing of diagnosis, and demographic information were taken from the survey wave in which each participant first reported a cancer diagnosis.

Relevant variables were drawn from a combination of the family and individual files in the PSID Public Data Index (55). The variables selected their file locations, and the years for which they are available are shown in Table I.

2. University of Illinois Cancer Center Tumor Registry

Case data for Aim 3 were obtained from the University of Illinois Cancer Center tumor registry. Included cases were those diagnosed with or treated for CRC at the Cancer Center or University of Illinois Hospital in Chicago, IL between 1995 and 2004. The majority of these cases lived in Cook County, IL at the time of diagnosis. Case information included full name, date of birth, race, ethnicity, sex, age at diagnosis, first contact date, site code, descriptive stage, full address at diagnosis, and current or last known address at the time the data were accessed.

3. LexisNexis Accurant

Residential histories for Aim 3 were obtained using the LexisNexis Accurant for Government database (LexisNexis Risk Solutions, Inc., Florida) (56). Searches were conducted using full name, date of birth, and residential address at the time of diagnosis using a matching procedure that has been described previously (51,57). Results included up to 15 previous addresses associated with each case, and the month and year that each address was first and last seen associated with the case's name.

TABLE I. VARIABLES FROM THE PSID PUBLIC DATA INDEX

Variable	File Location	Years Available
Ever diagnosed with cancer – head of household	Family Public Data Index 01>HEALTH STATUS	Biennial 2005 – 2019
Ever diagnosed with cancer – spouse	02>Physical Health 03>conditions 04>cancer, whether 05>head [OR 05>spouse]	Potential reach before 1999 due to question wording
Cancer type – head of household	Family Public Data Index 01>HEALTH STATUS	Biennial 2005 – 2017
Cancer type – spouse	02>Physical Health 03>conditions 04>cancer, whether 05>head [OR 05>spouse]: 06>type 07>1st mention	Potential reach before 2005 due to question wording
Age at interview	Individual Data Index 01>DEMOGRAPHIC 02>Age	1968 – 2017
Sex	Individual Data Index 01>DEMOGRAPHIC 02>Sex	1968 - 2017
Race	Family Public Data Index 01>DEMOGRAPHIC 02>Race and Ethnicity 03>race 04>head [OR 04>spouse] 05>1st mention	Head: 1968 – 2017 Spouse: 1985 – 2017
Ethnicity	Family Public Data Index 01>DEMOGRAPHIC 02>Race and Ethnicity 03>ethnicity 04>head [OR 04>spouse]	1997 - 2017
Hispanicity	Family Public Data Index 01>DEMOGRAPHIC 02>Race and Ethnicity 03>hispanicity 04>head [OR 04>spouse] 05>1st mention:	1985 – 1996 2005 – 2017
Completed education level: grades completed including college; encodes years of college but not degree completion	Family Public Data Index 01>EDUCATION 02>Grades Completed 03>including college 04>head [OR 04>spouse]	1975 – 1984 1991 – 2017
Completed education level: high school graduate, GED recipient, or neither	Family Public Data Index 01>EDUCATION 02>Grades Completed 03>excluding college 04>high school, no GED 05>graduate, whether 06>head [OR 06>spouse]	1985 - 2017

TABLE I. VARIABLES FROM THE PSID PUBLIC DATA INDEX (continued)

Variable	File Location	Years Available
Completed education level: highest college degree received	Family Public Data Index 01>EDUCATION 02>College 03>college degree, whether 04>head [OR 04>spouse]: 05>highest received	1985 - 2017
Total family income	Family Public Data Index 01>INCOME 02>Family Money 03>total family income	1968 - 2017
Total family wealth, excluding home equity	Family Public Data Index 01>WEALTH 02>Total Family Wealth 03>excluding home equity	1984, 1989, 1994, 1999 - 2017
Moved since spring of last year/last interview: reason for moving	Family Public Data Index 01>LOCATION AND MOBILITY 02>Mobility 03>moved since spring of previous year/last interview, whether: 04>reason 05>1st mention	1969 - 2017

4. Census Data

Components of the CD score at the Census tract level were downloaded from the LTDB for 1970, 1980, 1990, 2000, and 2010. Components, numerators, and denominators are shown in Table II. The LTDB provides estimated values from the Census and related programs in harmonized 2010 tract boundaries (58,59). In 2010, the LTDB provides actual Decennial Census counts and supplements these counts with the American Community Survey (ACS) 5-year estimates for 2008-2012.

Additional estimates from the ACS were used to represent individual years beyond 2010. ACS estimates already obtained from the LTDB were carried forward to 2011 and 2012. The ACS 5-year estimates for 2013-2017 were downloaded from the data.census.gov website and used to calculate CD scores for the remaining years of follow-up (60).

5. National Death Index

Both sources of case data were matched to the National Death Index. However, as described above, PSID restricted data obtained from the NDI were not available to researchers during the time this study was conducted.

Follow-up of the tumor registry cohort studied in Aim 3 was censored as of January 31, 2014. Vital status, cause of death, and date of death were obtained from NDI Plus (61). Survival time was calculated using either the date of death per NDI or the end of follow-up, whichever came first.

6. Data Preparation

i. Panel Study of Income Dynamics

Several CD score variables were created using different amounts of residential history. Point values of the CD scores were collected at 5, 10, 15, and 20 years before diagnosis. Because the PSID switched to collecting data every other year in 1997, CD score values collected in odd-numbered years were carried forward to a single even-numbered year before selection.

TABLE II. CENSUS AND AMERICAN COMMUNITY SURVEY VARIABLES USED TO CONSTRUCT THE CONCENTRATED DISADVANTAGE SCORE

CD Score Component	Numerator	Denominator
Educational status	Population over 25 with at least a 4-year college degree	Population over 25
Female-headed households	Households with children present and a female head	Households with children present
Poverty	Households with income <100% federal poverty level	Households
Unemployment	Population age 16 and older in the labor force and unemployed	Population age 16 and older in the labor force

Other missing CD scores were treated as truly missing. One average CD score variable was created, containing the mean observed CD score in the up to 20 years before diagnosis. No odd-year CD scores were carried forward to create this variable; instead, it is weighted to reflect the different frequency of data collection before and after 1997. Mean involuntary moves per year and mean annual income in the up to 20 years before diagnosis were created similarly. Income was adjusted for inflation using the CPI-U before averaging.

ii. Census Data

Concentrated disadvantage scores were calculated for each Census tract available in each year of potential exposure, 1970-2017 for the two studies combined. For intercensal years beyond 2010, the ACS 5-year estimates were used as described above. For intercensal years between 1970 and 2010, the CD score components were first estimated using linear interpolation. Linear interpolation was carried out using the EXPAND procedure of SAS 9.4 (62). Interpolated score components were added together to yield one CD score per tract per year.

C. Statistical Analysis

1. Panel Study of Income Dynamics

All statistical analysis of this data set was conducted in SAS 9.4 via remote access within the Virtual Data Enclave maintained by the University of Michigan. Because linked data from the NDI were not available to outside researchers during the period when this research took place, the outcome of interest was all-cause mortality. All-cause mortality and timing of death were ascertained using vital status and time of death as determined by PSID staff from survey non-response variables, returned surveys, information from surviving household members, and obituaries (54). The main explanatory variables of interest were lagged CD score at 5, 10, 15, or 20 years before being diagnosed with cancer, and race.

Descriptive statistics were calculated for the entire sample and stratified by vital status. Groups were compared by vital status using the Pearson chi square test for categorical variables, Mantel-Haenszel chi square for ordinal variables, and a two-sample t-test for continuous variables. The distribution and availability of the CD scores were explored using statistical and graphical methods.

Survival was modeled using Cox proportional hazards. Because this cohort was drawn from the PSID, a sample representative of the general population rather than a cancer cohort, age was used as the time scale for survival. This approach has been recommended by Korn, et al. for survival analysis in studies of healthy people where probability of death is expected to vary more as a function of age than of time in the study (63,64). This approach inherently adjusts for age. The cohort was divided into six birth year cohorts and the model was stratified by cohort to adjust for calendar effects that would otherwise influence probability of survival after a cancer diagnosis. For each cancer site model, the lagged income and CD variables closest to the empirical latency were used as predictors. For breast, prostate, and lung cancer, a lag of 20 years was used (65). For colon cancer, a lag of 10 years was used (66)

2. University of Illinois Hospital Tumor Registry

Data were analyzed in SAS 9.4. Chi-square tests were used to evaluate crude associations between covariates and stage at diagnosis or vital status.

The association between CD score and stage at diagnosis was modeled using multinomial logistic regression to allow inclusion of the unstaged category as an outcome. An additional logistic regression model excluded cases with missing stage data and evaluated the association between CD score and the odds of having advanced (AJCC stage III/IV) CRC at diagnosis. The association between CD score and CRC-specific survival was evaluated using the Cox proportional hazards model. Because this sample constituted a cancer cohort, the time scale for the survival model was days since cancer diagnosis. Two versions of each model were run, one using CD score at diagnosis and one using the CD score with a 10-year lag.

III. EXPOSURE TO NEIGHBORHOOD CONCENTRATED DISADVANTAGE ACROSS THE LIFE COURSE AND SURVIVAL AFTER CANCER DIAGNOSIS IN THE PANEL STUDY OF INCOME DYNAMICS COHORT

A. Introduction

Racial inequities in the risk of death from colorectal cancer (CRC) are preventable and caused primarily by racism. Present-day inequities emerged in the US within living memory: until the late 1970s, CRC-specific mortality was higher in whites. Today, non-Hispanic Black people are burdened by CRC incidence that is 20% greater, and CRC mortality more than 40% greater, than that experienced by non-Hispanic whites (1). Segregation and resulting neighborhood disadvantage are associated with racial inequities in multiple health behaviors and conditions, including cancer, which could be reduced through housing policies that prioritize desegregation, equity, and racial justice. However, this research is in its infancy. In 2016, a systematic review found just 17 studies on the association between segregation and cancer in the US context (2).

Foundational research is needed in this area to establish valid data sources and methods and to demonstrate that exposure to segregation precedes disparate cancer outcomes in time. A critical barrier to this research is that cancer registries do not collect the residential histories of cases. As a result, epidemiologists rarely know where people with cancer lived before being diagnosed. Because of this limitation, the conditions of segregated neighborhoods have not been linked to cancer outcomes in the years or decades before diagnosis when present-day cancers actually began to develop. The National Cancer Institute has identified residential history as an area of special interest, and encouraged researchers to use commercial credit reporting databases as a source of this information (67). However, these data are less available for people who are African American or Hispanic as compared to those who are white, and in people who have died, creating serious potential for bias (68). In addition, few residential history studies to date have attempted to

study segregation or explain racial cancer inequities. None have used a longitudinal cohort design, which could provide critical evidence that segregation causes cancer inequities.

To address these needs, this exploratory study evaluates the relationship between past neighborhood disadvantage and mortality among people diagnosed with cancer in the Panel Study of Income Dynamics (PSID). The PSID is a longitudinal, nationally representative survey that has followed individuals and households since 1968. Although primarily intended to study the social and economic conditions of families, the PSID also collects information on common health conditions and behavior. Disease prevalence estimates from the PSID are comparable to those from national public health surveys such as the National Health Interview Survey (NHIS) and the Surveillance, Epidemiology, and End Results registry (SEER) (15,16,53). While PSID data have been used to study the effects of a cancer diagnosis on economic and family life, to date they have not been used to study the causes of cancer or cancer outcomes.

This study was designed to provide an initial estimate of the relationship between past neighborhood disadvantage around the time of cancer initiation and mortality among people with cancer. We hypothesized that having experienced neighborhood disadvantage in the past would explain a significant portion of variation in cancer mortality.

B. Materials and Methods

1. Primary Data Source

The primary data source for this study is the PSID public data index and two of its restricted access datasets: the Geocode Match File containing the Census tracts where participant households have lived, and the Mortality File for participants who have died (55,69). The PSID is conducted every two years, meaning participant addresses are collected prospectively and validated by being used to conduct the survey. The PSID sample is self-replacing. When the children of study families split off to form their own households, they are eligible to continue to participate as heads of

households themselves. This means that for some participants, information about their residential history and family income is available from their childhood and potentially from birth.

Participant demographic information, income history, reason for moving, and self-reports of cancer come from the PSID public data index. In 1999, the PSID began collecting information about lifetime prevalence of cancer from heads of households and their spouses. In 2005, the PSID added follow-up questions that allowed participants to report the site(s) of their cancer. Participants were asked, “Has a doctor or other health professional ever told you that you had cancer or a malignant tumor?” Participants were asked the age at which they were diagnosed.

Additional covariate and outcome information was obtained from two of the PSID restricted datasets. Participant vital status was obtained by linking the records of included participants to the PSID Mortality File, which included the month and year of death and the age at death (54). The full mortality file also includes cause of death from record linkage with the National Death Index (NDI). However, after access to the mortality file was obtained, it was discovered that changes in the PSID’s contract with NDI would not allow death certificate date to be shared with external researchers. Participant residential histories were obtained from the 2010 Census PSID Geocode Match File. Data from this file included the 2010 Census tract of residence for each wave in which the participant’s household responded. The Address File, a confidential dataset, was geocoded by PSID staff using the latest Census TIGER/Line shapefiles for all states as of the 2021 release (70). Addresses were geocoded for all years through 2019 using SAS 9.4 PROC GEOCODE (71).

2. Sampling Frame

All included participants were spouses or heads of household in the PSID, for whom the most extensive data is available. They were included if they participated in 2005 or later, reported a history of cancer with a site other than skin cancer, and reported that they were at least 18 years old when the cancer was diagnosed. Participants were included in the survival analysis if they reported a history of breast, prostate, colon, or lung cancer; and had available residential history

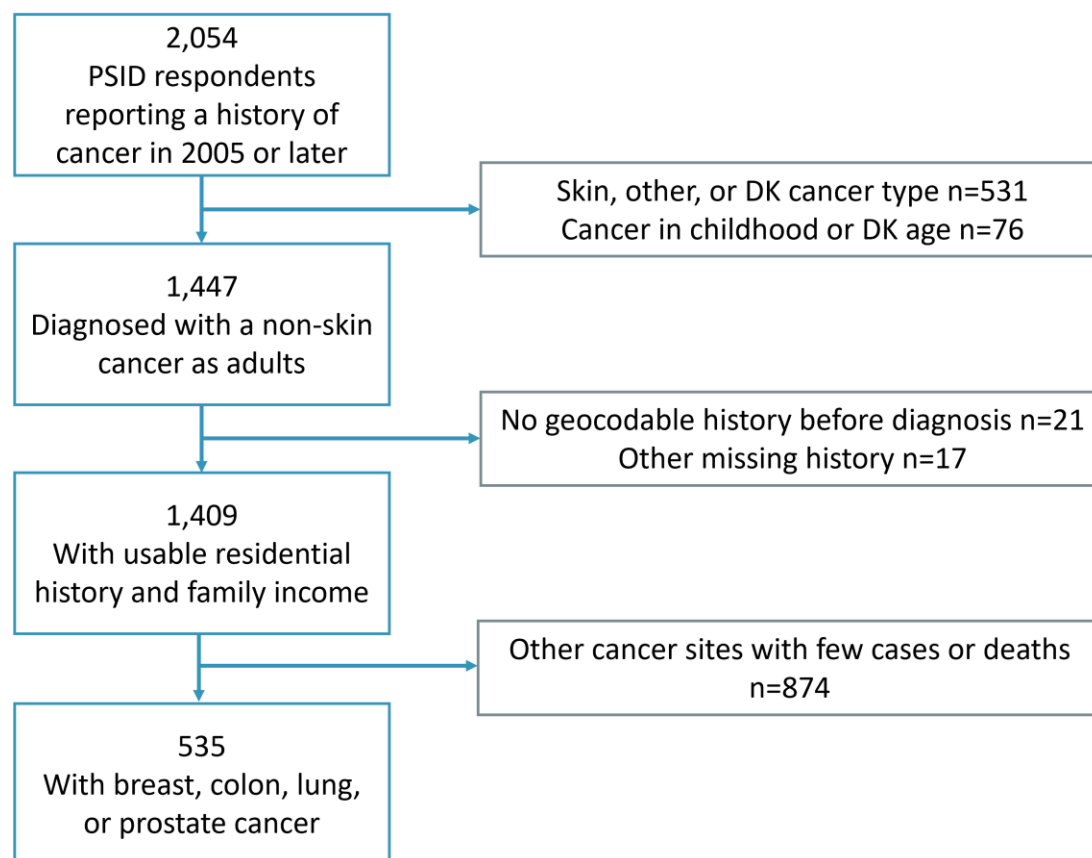


Figure 1. Selection criteria and sample history applied to the Panel Study of Income Dynamics data set

and family income data for the period of interest before diagnosis. These sites were selected because they were the most common reported cancer sites by PSID participants, and each had at least 20 observed deaths. (Cervical cancer was the third most common reported cancer with 125 cases but had only 17 observed deaths.) The selection criteria and sample history are shown in Figure 1.

3. Vital Status Ascertainment

Due to a change in accessibility of NDI data, PSID staff were unable to share linked death certificate data with external researchers for the study period. Based on the information about match quality in the Mortality File documentation, this study relied on the match information that PSID typically sends to NDI including the month and year of death and the age at death. For all deaths ascertained in the PSID cohort since 1980, information submitted to NDI was sufficient to obtain a “best” match in 82% of cases (54). In addition, an Attritor Tracking Project concluded in 2007—the next survey wave after cancer site information began to be collected—uncovered a large number of prior deaths that were able to be matched to NDI data with a high degree of confidence comparable to other survey years. Because cause of death data from the NDI match was not provided, this study used all cause mortality as the outcome of interest.

4. Residential Exposure Measurement

The primary measure of participants’ neighborhood environments over time was concentrated disadvantage (CD) as developed by Browning and Cagney (9,10). This is a composite measure of population characteristics including percent unemployment, percent of households in poverty, percent of households with a female head, and percent of the population with less than a college degree. The original measure, developed from a factor analysis of Census variables, included the percent of the population that is African American. Because this and related analyses were intended to study the contribution of neighborhood characteristics to racial health inequities, that component was dropped from the measure. In the original measure, all components other than the

tract racial composition had factor loadings of 0.85. The factor loading for the African American share of the population was 0.6. Therefore, dropping the African American population component yielded a score composed of equally weighted factors (10,17). Historical Census data were obtained from the Longitudinal Tract Database (LTDB) (58). The LTDB provides interpolated population estimates in 2010 tract boundaries from the 1970 Census onward, and tools for researchers to create additional estimates (59). Family incomes were adjusted to 2020 dollars using the Consumer Price Index for All Urban Consumers (CPI-U) (72).

5. Data Preparation

Demographic and cancer information were taken from the first survey year in which a participant reported a personal history of cancer. Participants were divided into six age cohorts based on observed years of birth: 1900-1919, 1920-1934, 1935-1949, 1950-1964, 1965-1979, and 1980-1995. Point CD measures were calculated for each participant at lags of 5, 10, 15, and 20 years before diagnosis. An average CD measure was calculated for each participant using all available residential history for up to 20 years before diagnosis. In 1997, the PSID moved from 1- to 2-year intervals between surveys. During this period, income and CD values from odd-numbered years were carried forward once to the subsequent even-numbered year. No values were carried forward for years when the survey was conducted, but respondents did not participate. The 20-year mean CD score variable was calculated without carrying forward odd-numbered year scores. Instead, this average was weighted to reflect the different frequency of data collection before and after 1997. Variables for mean and lagged-point household income were created using the same approach.

6. Statistical Analysis

All data analysis and preparation were conducted using SAS 9.4. Because of the limited available mortality data, the outcome for this analysis was all-cause survival in PSID participants who reported a personal history of cancer. The independent variables of interest were lagged CD and individual race.

The distribution of demographic characteristics and cancer sites was summarized using the FREQ procedure. Groups were also compared by vital status using the Pearson chi square test for categorical variables, Mantel-Haenszel chi square for ordinal variables, and a two-sample t-test for continuous variables. Differences in the distribution of the lagged point and 20-year mean CD score values were explored using the MEANS and UNIVARIATE procedures.

Kaplan-Meier survival curves were calculated for each site stratified by quintile of mean 20-year concentrated disadvantage score. Survival was modeled using the Cox proportional hazards model and the PHREG procedure. For this analysis, age was treated as the time scale and the model was stratified by birth cohort. This approach treats survival time before the first report of a history of cancer as left-truncated, and implicitly compares the risk of death among people who are the same age, rather than in people who have the same amount of follow-up time (63,64). Use of age as the time scale has been recommended by some researchers for analysis of cohort data in healthy people, like the PSID cohort from which this sample was drawn, when risk of death is expected to vary more by age than by time spent participating in the study (63). In this case, the use of age as the time scale also avoids measurement error that would be introduced by estimating exact dates from the PSID data, which only provided the year, month, and age at cancer diagnosis and at death. Finally, this method adjusts for both age and calendar effects that might influence either the outcome or risk factors (63,64).

Individual age cohorts and racial groups were only included in each model if they contained members who both had and had not died. This resulted in restricting the analysis to white and Black participants.

For each cancer site model, the lagged income and CD variables closest to the empirical latency were used as predictors. For breast, prostate, and lung cancer, a lag of 20 years was used (65). For colon cancer, a lag of 10 years was used (66).

C. Results

1. Study Participant Characteristics

The final study sample included 535 PSID participants who reported a personal history of breast, colon, lung, or prostate cancer as an adult; and had usable residential history for the relevant period before the age at which they were diagnosed. Descriptive statistics for the sample are shown in Table III. Crude associations with survival were evaluated using 20-year mean income and CD score because these values were available for all participants with all cancer types.

Vital status was significantly associated with birth cohort, age at diagnosis, sex, 20-year mean household income, and cancer site. White participants were slightly more likely than Black or African American participants to have died, but the association was not significant. Mean CD score in the 20 years before diagnosis was not associated with vital status.

2. Distribution and Availability of CD Score

All CD score variables had similar central tendencies and were available in the great majority of participants (Table IV). The only CD score variable with complete availability for the entire sample was the 20-year mean. However, the 20-year mean CD score value varied less across participants, with a much smaller standard deviation and range than the lagged point values. It also represented different amounts of residential history in different participants, rather than the neighborhood environment around the time of cancer initiation. Therefore lagged point CD scores were used in the survival models to both maximize observed variation in the main explanatory variable, and on grounds of biological plausibility.

3. Survival Analysis

For each of the four analyzable cancer sites recorded among PSID participants, there was no difference in the survival function when stratified by quintile of historical concentrated disadvantage (Figure 3). There was also no adjusted relationship between lagged point CD score

TABLE III. CHARACTERISTICS OF PSID PARTICIPANTS WHO REPORTED A HISTORY OF BREAST, COLORECTAL, LUNG, OR PROSTATE CANCER, BY VITAL STATUS, 2005-2019

	Total (535)	Alive (366)	Died (169)	p
Birth Cohort				<0.0001
1900 - 1919	8 (1.5%)	1 (12.5%)	7 (4.1%)	
1920 - 1934	112 (20.9%)	33 (9.02%)	79 (46.8%)	
1935 - 1949	183 (34.2%)	136 (37.2%)	47 (27.8%)	
1950 - 1964	199 (27.2%)	166 (45.4%)	33 (19.5%)	
1965 - 1979	22 (6.2%)	30 (8.2%)	3 (1.8%)	
Age at Diagnosis (Mean, SD)				
18-35	8 (1.5%)	8 (2.19%)	0	<0.0001
36-55	185 (34.6%)	150 (41.0%)	35 (20.7%)	
56-75	283 (52.9%)	185 (50.6%)	98 (58.0%)	
≥76	59 (11.0%)	23 (6.3%)	36 (21.3%)	
Sex				0.03
Male	239 (44.7%)	152 (41.5%)	87 (51.5%)	
Female	296 (55.3%)	214 (58.5%)	82 (48.5%)	
Race				0.09
White	376 (70.3%)	249 (68.0%)	127 (75.2%)	
Black or African American	159 (29.7%)	117 (32.0%)	42 (24.9%)	
Mean Income in the up to 20 years before dx (median, IQR)				0.002
Lowest quintile	97 (18.1%)	57 (15.6%)	40 (23.7%)	
2nd quintile	92 (17.2%)	55 (15.0%)	37 (21.9%)	
3rd quintile	105 (19.6%)	71 (19.4%)	34 (20.1%)	
4th quintile	107 (20.0%)	75 (20.5%)	32 (18.9%)	
Highest quintile	134 (25.1%)	108 (29.5%)	26 (15.4%)	
Cancer Site				<0.0001
Breast	224 (41.9%)	176 (48.1%)	48 (28.4%)	
Colon	79 (14.8%)	41 (11.2%)	38 (22.5%)	
Lung	63 (11.8%)	27 (7.4%)	36 (21.3%)	
Prostate	169 (31.6%)	122 (33.3%)	47 (27.8%)	
Mean concentrated disadvantage in the up 20 years before dx (Mean, SD)				0.44
Least disadvantaged	111 (20.8%)	80 (21.9%)	31 (18.3%)	
2nd quintile	103 (19.3%)	68 (18.6%)	35 (20.7%)	
3rd quintile	109 (20.4%)	80 (21.9%)	29 (17.2%)	
4th quintile	101 (18.9%)	68 (18.6%)	33 (19.5%)	
Most disadvantaged	111 (20.8%)	70 (19.1%)	41 (24.3%)	

TABLE IV. DISTRIBUTIONS AND AVAILABILITY OF CD SCORES AMONG PSID PARTICIPANTS WHO REPORTED A HISTORY OF BREAST, COLORECTAL, LUNG, OR PROSTATE CANCER, 2005-2019

Versions of CD Scores	N	Mean	SD	Median	Min	Max	Range
Point values							
In year of diagnosis	521	1.12	0.38	1.10	0.17	2.36	2.19
5 years before diagnosis	530	1.13	0.38	1.11	0.25	2.47	2.22
10 years before diagnosis	527	1.16	0.37	1.12	0.36	2.53	2.17
15 years before diagnosis	524	1.19	0.38	1.13	0.37	2.64	2.26
20 years before diagnosis	529	1.26	0.40	1.18	0.39	2.73	2.35
Mean value							
20 years before diagnosis	535	1.17	0.12	1.17	0.86	1.55	0.70

before diagnosis and probability of all-cause survival. This was true regardless of the chosen lag time or the relationship of that time to either diagnosis or biological latency (Table V). Full model result tables for each cancer site are shown in the appendix. Only one model result, for the association between all-cause mortality and CD score 10 years before diagnosis in people with lung cancer, was statistically significant. However, this result was inconsistent with modeled HRs at other lag times, and there was no clear biological or theoretical importance to the 10-year lag.

In each cancer site, adjusted point value HRs were inconsistent and without pattern across time. For example, point value HRs in colorectal cancer were consistent with harmful effects of increasing disadvantage when measured at diagnosis or 20 years before diagnosis; but beneficial effects of increasing disadvantage for the years in between. The adjusted hazard ratio for the effect of CD score at each lag time is plotted in Figure 4 to demonstrate the lack of a trend or pattern to the relationship. Taken together, these results are consistent with either no effect of past CD score on risk of death in PSID participants with a history of cancer; or with insufficient power to detect an effect in this cohort and using this approach.

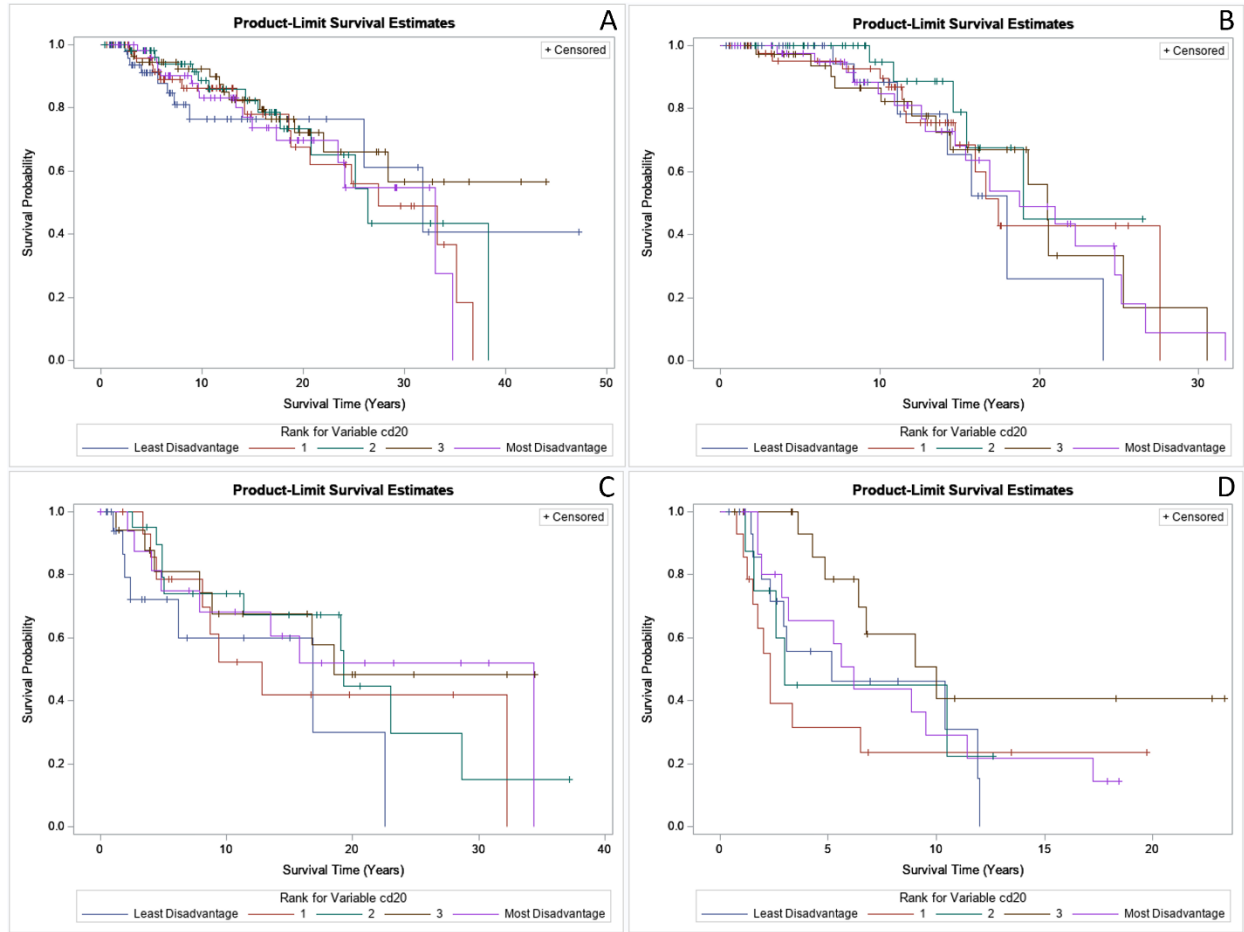


Figure 2. Kaplan-Meier survival estimates stratified by mean historical concentrated disadvantage among PSID participants with each type of cancer: breast (A), prostate (B), colorectal (C), or lung (D)

TABLE V. ADJUSTED HAZARD RATIOS FOR THE ASSOCIATION OF CENSUS TRACT-LEVEL CONCENTRATED DISADVANTAGE AT DIFFERENT POINTS IN TIME AND ALL-CAUSE MORTALITY AMONG PSID PARTICIPANTS WHO REPORTED A HISTORY OF BREAST, COLORECTAL, LUNG, OR PROSTATE CANCER, 2005-2019¹

Cancer Site	Concentrated Disadvantage									
	In year of diagnosis		5 years before diagnosis		10 years before diagnosis		15 years before diagnosis		20 years before diagnosis	
	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
Breast	1.31	[0.62, 2.74]	1.86	[0.86, 4.01]	1.18	[0.55, 2.55]	1.18	[0.61, 2.29]	0.79	[0.34, 1.82]
Colorectal	1.60	[0.55, 4.63]	0.35	[0.11, 1.07]	0.55	[0.16, 1.89]	0.75	[0.30, 1.88]	2.01	[0.66, 6.17]
Lung	0.99	[0.45, 2.15]	1.74	[0.60, 5.07]	0.25*	[0.07, 0.90]	0.68	[0.24, 1.97]	1.60	[0.61, 4.18]
Prostate	1.14	[0.49, 2.65]	0.97	[0.42, 2.24]	1.22	[0.57, 2.65]	1.81	[0.82, 3.97]	1.41	[0.63, 3.16]

¹Bolded cells indicate model results for the CD score lag nearest to biological latency

* p < 0.05

After adjustment for CD score and race in all sites, and sex in colorectal and lung cancer, lagged income was also not significantly associated with survival in any cancer group. Hazard ratios for the adjusted association between income and survival were consistent with no effect.

Race was not significantly associated with survival in any cancer, but the specific results varied somewhat by site. Hazard ratios were consistent with increased risk of death in white participants with breast cancer (HR 1.30, 95% CI 0.60, 2.82) and lung cancer (HR 1.91, 95% CI 0.75, 4.88). For participants with colorectal cancer (HR 1.50, 95% CI 0.66, 3.43) or prostate cancer (HR 1.16, 95% CI 0.53, 2.55), survival was non-significantly poorer in Black participants. Results were generally not consistent with racial inequities in cancer mortality in the general population, where Black people are consistently disadvantaged.

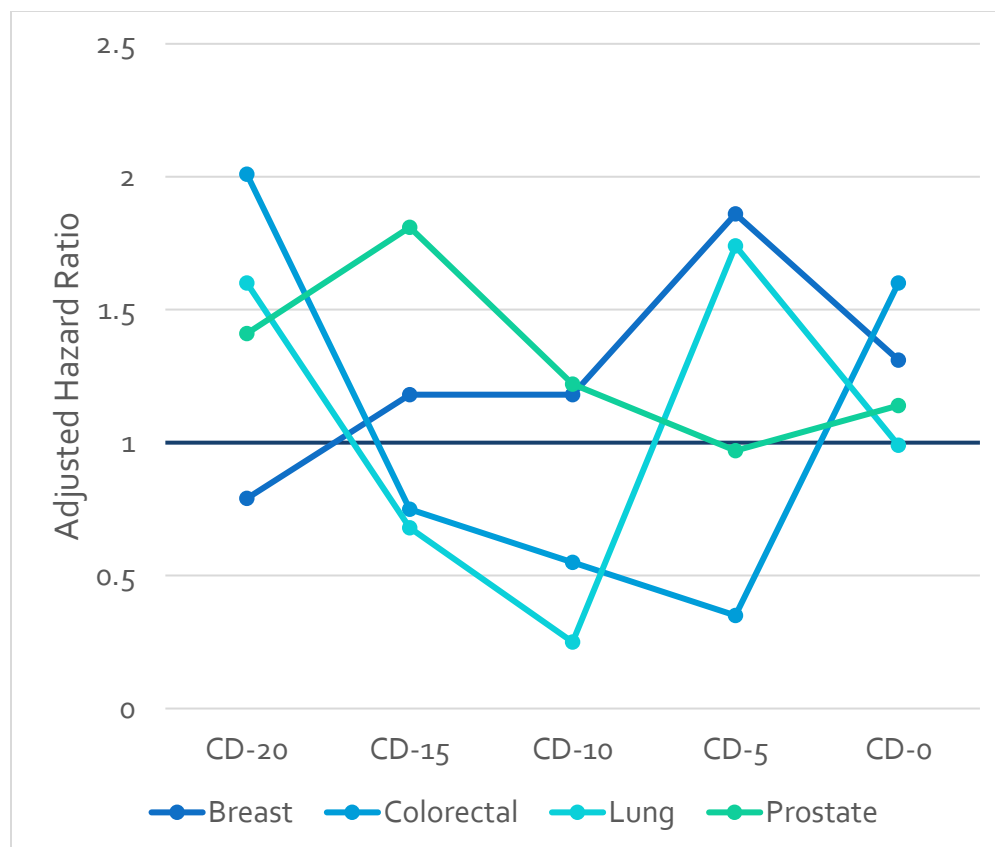


Figure 3. Relationship between CD score and survival at each point in time for PSID participants with a history of breast, colorectal, lung, or prostate cancer

D. Discussion

Concentrated disadvantage around the time of disease initiation does not predict all-cause survival in PSID participants with an adult history of breast, prostate, colorectal, or lung cancer. Both the design and procedures of the PSID, and the true mortality experience of PSID participants, likely contributed to this result. Lessons learned from this analysis may be informative to other epidemiologists seeking to leverage sociological surveys as a source of residential history data or to evaluate associations between social conditions and health outcomes.

Despite the large sample size and long duration of the PSID as a whole, this analysis was underpowered compared to initial expectations. Original sample persons were adult heads of

household or spouses in 1968, but the first cancer history questions were not asked until 1999, and cancer site information was not collected until 2005. As a result, many potential cancer cases were never ascertained in this cohort—particularly in participants who died due to any cause before 2005. Because the most detailed health information is collected from adult heads of household and spouses (termed sample persons), remote family history information was only relevant in this subgroup of PSID participants, rather than in all people who contributed data to the survey as household members. There were 26,084 individuals in the 2019 survey wave, the last conducted in the data set used for this analysis. However, only 9,694 participants were sample persons in that year (73). From 2005-2019, 1,449 sample persons met the inclusion criteria of reporting an adult history of any non-skin cancer site; of these, 1,410 had usable geographic data to construct the CD scores. These 1,410 individuals varied further by the cancer site they reported, resulting in a relatively small number of cases available for each survival model.

The primary endpoint in this analysis was relative survival after cancer diagnosis, a time-dependent outcome. Even within the restricted datasets, few precise dates relating to PSID participants are routinely available to researchers. For example, participants report the age at which they were diagnosed with cancer and their birth and death dates are given with a month and year only. This analysis used age as the time scale in models with stratification by birth cohort. This approach avoided introducing further measurement error by estimating exact dates. It also allowed control of calendar effects. However, additional measurement error was likely introduced by recall bias among PSID participants. While a previous analysis by Zajacova et al. determined that cancer recall errors were random within the PSID cohort, the loss of power was likely meaningful given the site-specific sample sizes available for this analysis (53).

Multiple previous studies have evaluated the quality of health data in the PSID and its comparability to national health surveys and registries (15,16,53). These analyses found that health indicators derived from the PSID were broadly comparable to those obtained from the NHIS and

SEER. Since the cancer questions were revised in 2005, this section of the PSID questionnaire has been substantially similar to the largest and most relevant comparison survey, the NHIS. Adult participants in the 2023 NHIS were asked, “Have you EVER been told by a doctor or other health professional that you had ... Cancer or malignancy of any kind?”; participants who answer “yes” can report the site of the cancer (74). The wording of this NHIS question similar in 1999, when a more limited cancer question was added to the PSID. Since 2005, the main difference between the cancer history data collected in the NHIS and PSID is in the number of cancer sites recorded. In 1999, the PSID recorded self-reports of non-skin cancers. In 2005, the updated questions collected structured data on 10 cancer sites. In 2013, the questions were updated again and recorded 13 cancer sites. The NHIS, a larger and health focused survey, has collected structured data on history of cancer in 29 sites since at least 1999 (75). The NHIS and PSID are alike in asking participants the age at which their cancer was diagnosed. The ability to evaluate the consistency of self-reported age at diagnosis may be regarded as a strength of the PSID’s longitudinal design. Researchers planning to use survey data to study cancer outcomes should consider that these questions elicit information on lifetime prevalence of cancers diagnosed in adults. The decision to use self-reported age at diagnosis to estimate cancer occurrence rests on assumptions of accuracy that may not be testable in cross-sectional studies.

However, some relevant patterns in the PSID sample have not been fully explained. Even after weighting, PSID sample persons are significantly more likely than NHIS participants to report that they have limited ability to work (18-20% of PSID participants vs. 11-12% of NHIS participants) (15). In an analysis of the 1999-2005 survey waves, Zajacova, et al. found that all-site cancer prevalence was gradually increasing in the PSID sample even after weighting (53). The reason for this pattern remains unclear. However, it is consistent with the finding in this study that Black PSID participants with cancer were not at greater risk of death compared to their white counterparts. Because of the requirement that residential histories be available, people who met

the inclusion criteria for this study were long-term participants by definition. This raises the possibility of a healthy participant bias that could also have influenced the results. If PSID participants are living longer with chronic illnesses or activity limitations compared to the general population, and racial survival disparities are reduced or absent, this could explain the gradually increasing prevalence of cancer in this cohort.

Finally, the approach taken in this study—focused on a single measure of neighborhood conditions, but agnostic to place—may not be the best suited to understand the hypothesized relationship in which exposure to segregation might influence survival with cancer. Segregation has been described as a fundamental cause of health inequities through its influence on access to resources (37). This relationship is supported by evidence that access to relevant resources, especially health care, is associated with cancer outcomes and partially accounts for outcome disparities (5,76,77). Evidence also supports a relationship between concentrated disadvantage and cancer disparities (17,18). However, geographic patterns of residential segregation, and the specific policies used to enforce it, vary by jurisdiction, with different practical effects on access to care (2,78–80). Researchers interested in capturing the health effects of racial segregation may need to limit their studies to specific geographic areas at the time points of interest and accept the resulting limitations on the types of residential histories represented in their studies. The interpretation and relevance of concentrated disadvantage is also likely to vary over both space and time. For example, the negative associations, if any, of living in a female-headed household were very different in Chicago in the 1990s than in a national study conducted in 2022 (10). Even for components with more enduring relevance, the meaning of a specific unemployment rate or rate of educational attainment depends on the economy and norms of the region. To be interpreted as a consequence of segregation, neighborhood concentrated disadvantage must be able to be evaluated relative to nearby neighborhoods, not to the entire country. Segregation and its consequences may also occur in metropolitan regions, crossing jurisdictional boundaries. They should be thought of as

scale dependent, and measured in geographic units that are relevant to both local patterns of disparity and potential policy solutions (78).

Lessons learned from this study can support other health research using the PSID, particularly of conditions and behaviors that are more common, have shorter latency, or are more commonly measured by self-report. Other chronic conditions self-reported by PSID sample persons include arthritis, asthma, hypertension, diabetes, heart disease, history of heart attack or stroke, and chronic lung disease. Most of these were asked in every odd-year survey wave since 1999. Sample persons have also answered questions about activities of daily living since 1999, smoking in 1986 and repeatedly since 1999, and rated their general health status since 1984. Additional health information was collected regarding the health status and behavior of children, adolescents, and young adults in the Child Development Supplement and Transition to Adulthood Supplement. Despite this study's negative findings, health inequities have been observed among PSID participants. For example, Johnson, et al. documented racial disparities in self-reported health status and found that these disparities were partly explained by neighborhood conditions experienced in young adulthood (81). Strengths of this study included its choice of a health outcome with a plausible link to residential history, control of calendar effects through restriction to a single age cohort, and choice of exposure measurements tied to the life course rather than a self-reported disease time point.

The PSID remains an exceptional source of longitudinal data on social and economic conditions measured at the individual and household level. The health effects of state level policies could be studied, where appropriate, using data on current state of residence, state of birth, and the state in which sample persons grew up, available in the public data index.

E. Conclusion

Concentrated disadvantage around the time of disease initiation did not predict all-cause survival in PSID participants with an adult history of breast, prostate, colorectal, or lung cancer in

our study. This finding differs from a body of research literature linking unequal socioeconomic conditions in general, and concentrated disadvantage specifically, to inequities in cancer outcomes. Findings were likely influenced both by limitations of the PSID data set that were relevant to this study design, and an apparent lack of racial cancer survival disparities among PSID sample persons. Social epidemiologists considering using PSID data to study health inequities should focus on common conditions and behaviors, socioeconomic factors that have the same interpretation across the entire study area and evaluate whether the specific outcomes or disparities observed in the PSID sample reflect those observed in the general population.

IV. PLANNED MEDIATION ANALYSIS IN THE PSID COHORT IS NOT JUSTIFIED BY PRIOR RESULTS

The original Aim 2 for this study was to “Conduct a mediation analysis to determine whether neighborhood disadvantage at diagnosis mediates the relationship between past neighborhood disadvantage and all cause mortality among people with cancer in the PSID cohort. This analysis will provide an initial estimate of whether residences in the distant past can affect present-day cancer outcomes; and if so, how much of their effect is contained in the address at diagnosis.”

However, as Chapter III describes, there was no meaningful relationship between neighborhood CD score and all-cause survival following a cancer diagnosis in the PSID cohort. This was true regardless of whether neighborhood CD score was measured in the past or in the same year the cancer was diagnosed. Therefore, no mediation analysis was justified or conducted.

V. NEIGHBORHOOD CONCENTRATED DISADVANTAGE PREDICTS SURVIVAL BUT NOT STAGE AT DIAGNOSIS IN A COLORECTAL CANCER COHORT

A. Introduction

Racial inequities in colorectal cancer in the United States result in disease-specific mortality rates that are 37% greater in non-Hispanic Black people than in non-Hispanic whites (82). No genetic or biological factor can convincingly explain this disparity, which operates across the cancer continuum and means that Black people experience higher CRC incidence, are less likely to be screened, are diagnosed at younger ages and more advanced stages compared to white people, and are less likely to receive appropriate treatment in a timely manner (19–26). While risk of CRC remained similar between Black and white people through the mid-1980s, mortality disparities emerged in the late 1970s even as screening tests for CRC became available and treatment options improved (82–84). Yet differences in receptivity to treatment are also not to blame: Black and white patients with CRC have similar responses to appropriate cancer-directed treatment (1,85,86). In equal access health care systems, disparities in receipt of this treatment are reduced or eliminated, and disparities in outcomes are correspondingly reduced (27–31).

Social epidemiologists have argued that segregation, and the conditions of segregated neighborhoods, are a fundamental cause of health inequities (37). The negative health effects of segregation have persisted across time, regardless of the local policies used to segregate or the people being targeted for exclusion, and are consistent whether the health outcome of interest concerns chronic disease or COVID-19 (87–90). Neighborhood conditions influence both access to health care and other resources, and exposure to environmental hazards and discrimination. Measures of the social and economic conditions of segregated neighborhoods, such as concentrated disadvantage (CD), are associated with racial disparities in cancer survival and other chronic diseases when measured at diagnosis (9,17,18). The biosocial process by which neighborhood

conditions undermine the health of non-Hispanic Black people has been termed “weathering”, and as the name implies, it is a cumulative process (91). However, because information about past residences is not normally available from cancer registries, cancer epidemiologists usually do not know if a person’s address at diagnosis is typical of the environments in which they have lived over time. This limits our understanding of how reducing segregation and improving neighborhood conditions might benefit health, and what other policies might be necessary to eliminate racial cancer inequities.

To address lack of information about historical exposures, cancer epidemiologists are increasingly turning to commercial public records databases to reconstruct the residential histories of people diagnosed with cancer (49,92–94). These studies have largely focused on environmental exposures rather than social or economic neighborhood conditions. The objective of this study is to evaluate the association between past neighborhood CD, and cancer prognosis and outcomes, in a cohort diagnosed with CRC in a predominantly urban health system.

B. Methods

1. Setting

This study was conducted in the UI Health System, located in Chicago, IL. The UI Health System includes a 465-bed tertiary hospital, outpatient clinics, a Cancer Center, and a Federally Qualified Health Center network. Cook County, IL, which contains Chicago, is the primary service area for UI Health and the Cancer Center.

2. Case Data, Ascertainment and Inclusion Criteria

This study was reviewed and approved by the Institutional Review Board of University of Illinois Chicago. Following approval, the tumor registry was used to identify all incident colorectal cancers diagnosed within the UI Health System between January 1, 1995 and December 31, 2004. Included cases were diagnosed with primary cancers of the colon or rectum (International Classification of Diseases for Oncology, third edition [ICD-O-3], codes C18-C20).

Case information includes full name, date of birth, race, ethnicity, sex, age at diagnosis, first contact date, anatomic site, descriptive stage, full address at diagnosis, and current or last known full address at the time the data were accessed. Race and ethnicity were categorized as non-Hispanic white (NHW), non-Hispanic Black (NHB), Hispanic, and Other based on the small number of cases in patients of other races or ethnicities.

The selection of the analytic cohort is illustrated in Figure 1. There were 780 cases identified during the study period. Of these, 169 were excluded due to age. Selected cases were at least 40 years old, and less than or equal to 75 years old, at diagnosis. The lower age limit was selected to ensure that all included cases would have at least 10 years of adult residential history before diagnosis. The upper age limit was selected to ensure that all included cases would have several years of potential follow-up after diagnosis to support survival analysis.

An additional 158 cases were excluded because they could not be matched in the LexisNexis Accurint for Government database, described below. The selection of this sample, and the limitations associated with use of commercial residential histories, have also been described previously in Freeman, et al. (68). Finally, 9 cases were matched by LexisNexis but no residential histories could be derived from the returned addresses.

3. Tumor Characteristics

Stage at diagnosis was based on the American Joint Committee on Cancer (AJCC) 8th edition summary stages for cancers of the colon and rectum (95). Stages included 0, I, II, III, IV, or missing. Three anatomic sites were abstracted from medical records: colon, rectosigmoid junction, and rectum. Vital status as of June 30, 2018, was ascertained by the tumor registry through linkage to the Social Security Death Master File (96).

4. Residential History Data Collection and Preparation

Residential histories were obtained using the LexisNexis Accurint for government database (LexisNexis Risk Solutions, Inc., Florida). Searches were conducted using full name, date of birth,

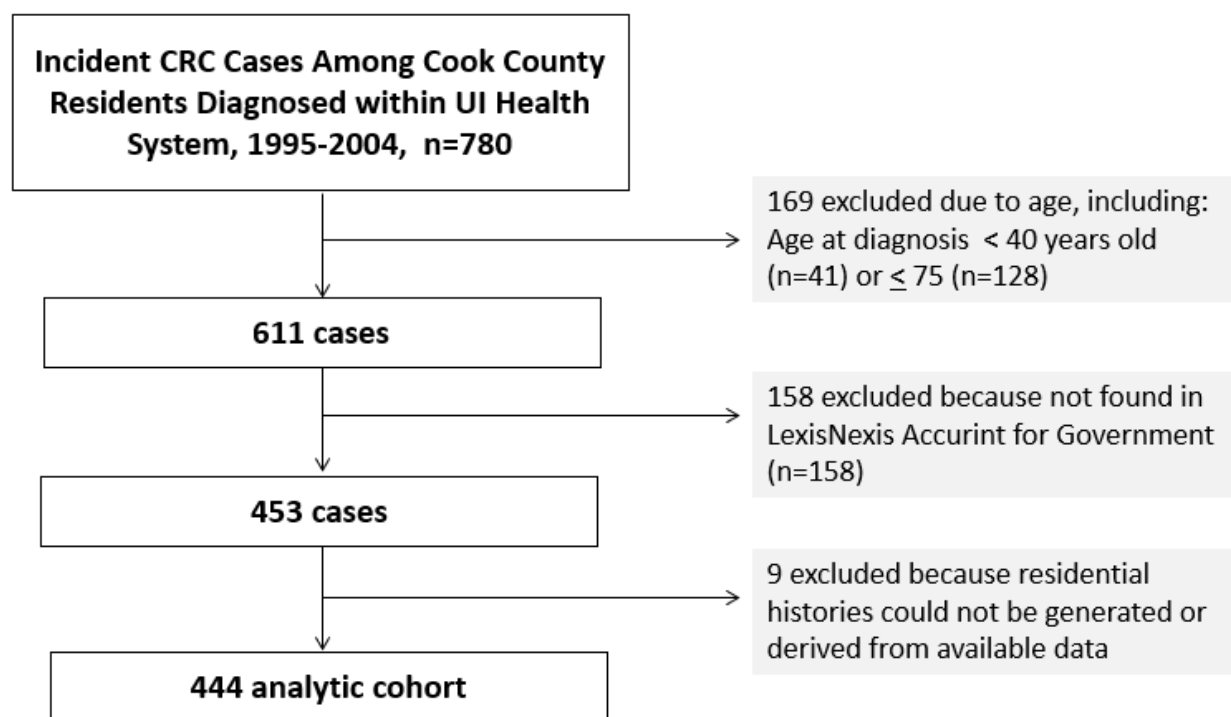


Figure 4. Selection of the analytic cohort

and residential address at the time of diagnosis using a matching procedure that has been described previously (51,57). Results included up to 15 previous addresses associated with each case, and the month and year that each address was first and last observed in association with the case's name.

Commercial residential history data from LexisNexis and residential addresses from the tumor registry were combined, geocoded, and matched to their 2010 Census tract in R 4.1 by using the censusxy package to access the Census batch geocoder (97).

Geocoded addresses were processed using ResHistGen, a publicly available SAS macro developed by Westat (51,98). The program deduplicates address data and creates a single continuous residential history for each case by reconciling the start and end dates associated with individual addresses.

Most addresses that could not be geocoded were able to be adjudicated by manual review and web searching to look for possible spelling or data entry errors. The corrected addresses were

then resubmitted to the Census batch geocoder. However, because the residential histories included older addresses by definition, not all addresses were able to be geocoded or processed by ResHistGen. These included post office boxes, obsolete address formats such as rural routes that could not be matched to an updated address, foreign and some military addresses, and addresses that no longer exist. The ResHistGen deduplication process also does not support analysis of residential histories in which a person leaves an address and returns repeatedly. In most cases, these errors reduced the number of addresses available for analysis but did not prevent creation of a partial residential history. In the 9 cases described above, all available addresses were unable to be geocoded and also unable to be adjudicated by human reviewers.

Concentrated disadvantage (CD) data were assembled from the Longitudinal Tract Database (LTDB), a public database of Census and American Community Survey (ACS) estimates within harmonized tract boundaries covering 1970-2019 (58). In 1970-2010, population data were estimated using the decennial census and linear interpolation for intercensal years. In 2010 and beyond, population data come from the ACS 5-year estimates for 2008-2012 and 2013-2017. Because the purpose of this analysis was to evaluate the effect of neighborhood disadvantage on racial health inequities, the CD score was modified from the original formula published by Brown and Cagney (10) to exclude the share of the tract population that was African American. In the original measure, the components related to poverty, female-headed households, employment, and education each had factor loadings of 0.85, while the factor loading for percent African American was 0.6. Therefore, dropping the component for tract racial composition yielded the sum of equally weighted tract-level poverty rate, share of households with a female head, unemployment rate, and share of the population age 25 or older with no college degree (10,17). After the individual components were estimated, a separate CD score was calculated for each year/tract combination and joined to the residential history file. Two scores were selected for analysis for each case: CD

score at residence in the year of diagnosis, and CD score at then-current residence 10 years before diagnosis.

Cases were matched to the National Death Index (NDI) with follow-up through December 31, 2014. Based on the date and vital status at last contact, 53 cases were known to still be alive at the end of follow-up. For the remaining 391 cases vital status and cause of death, if applicable, were ascertained using the NDI-Plus service. Searches were based on first name, last name, date of birth, sex, and race. Matches were adjudicated based on tumor registry records, other health system records, and public records including obituaries. Of these cases, six were known to have died based on hospital records but were not found by the NDI search. In these cases, attribution of cause of death was based on cancer stage at diagnosis and the time from diagnosis to death. Survival time was calculated in days between the tumor registry date of diagnosis and the date of death, if any, in the NDI. The outcome of interest was CRC-specific death.

5. Statistical Analysis

Data were analyzed in SAS 9.4. Chi-square tests were used to evaluate crude associations between covariates and stage at diagnosis or vital status.

The association between CD score and stage at diagnosis was modeled using multinomial logistic regression to allow inclusion of the unstaged category as an outcome. Because stage data was missing for a large proportion of cases, this association was also modeled using binary logistic regression. In the binary logistic regression model, all cases with missing stage data were excluded. The model predicted the probability of having advanced (stage III/IV) cancer at diagnosis. For these models, stages 0 and I were combined.

Kaplan-Meier survival curves were calculated stratified by stage at diagnosis. The results were used to determine how to model the effect of stage in further survival analysis. The association between CD score and CRC-specific survival was evaluated using the Cox proportional

hazards model. For this model, 19 stage 0 cases were excluded because no deaths were observed in this group.

Sex, race/ethnicity, age at diagnosis, anatomic site, and year of diagnosis were included in both models *a priori*. Age at diagnosis was divided by ten to yield modeled parameter estimates per 10-year change. Stage at diagnosis was included in the survival model as a nominal variable because of the large number of unstaged cancers. Female sex, non-Hispanic white race/ethnicity, and rectal cancer site were treated as reference categories. Two main versions of each model were run, one using CD score at diagnosis and one using the CD score with a 10-year lag. Additional survival models were run using CD score lags of 5, 15, or 20 years to assess whether there was a trend in association between survival and CD score at a given time point.

C. Results

There were 444 cases of colorectal cancer in the UI Health tumor registry diagnosed from 1995-2004 (Table VI). Mean age at diagnosis was 59.9 (SD 9.1). Most cancers were located in the colon. A large proportion of cancers had missing stage information (Figure 2).

There were no crude differences in cancer stage when stratified by concentrated disadvantage at diagnosis (Table VII). There were significant differences in CD at diagnosis by race and ethnicity, with non-Hispanic Black patients much more likely to be exposed to higher than median neighborhood CD. The majority of Hispanic patients were also living in neighborhoods with higher than median CD when their cancers were diagnosed. Those living in more disadvantaged neighborhoods were also more likely to be diagnosed with cancers of the colon, although the effect was relatively weak and barely significant.

As shown in Table VIII, this crude relationship reflects the known increased risk of proximal colon cancers in non-Hispanic Black people, which was also observed in this study. However, there were no meaningful differences in stage at diagnosis by race or ethnicity.

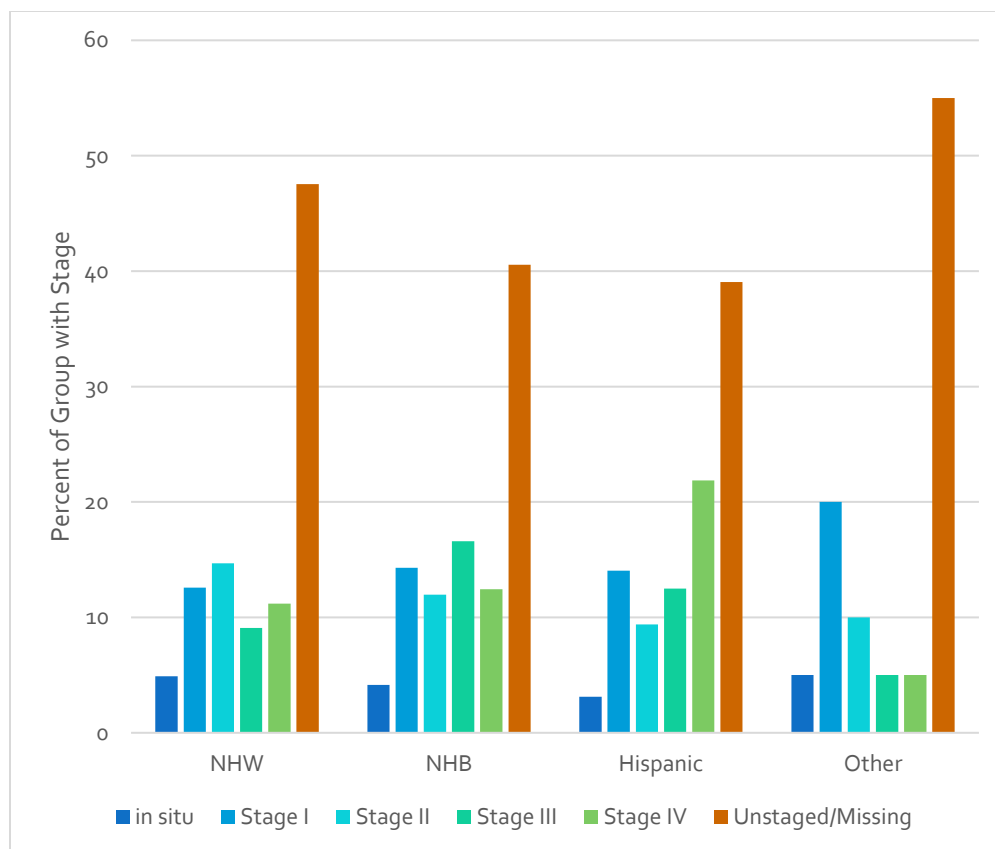


Figure 5. Distribution of cancer stage by race among UI Health patients diagnosed with colorectal cancer, 1995-2004

TABLE VI. CHARACTERISTICS OF UI HEALTH PATIENTS DIAGNOSED WITH COLORECTAL CANCER, 1995-2004, BY VITAL STATUS

	Censored	CRC Death	Total	p
N (%)	260 (58.6)	184 (41.4)	444	
CD Score at Diagnosis [mean (SD)]	1.30 (0.52)	1.37 (0.47)	1.33 (0.50)	0.15
CD Score at Diagnosis -10 years [mean (SD)]	1.33 (0.52)	1.40 (0.48)	1.36 (0.50)	0.19
Age [mean (SD)]	60.1 (9.1)	59.5 (9.1)	59.9 (9.1)	0.51
Sex				0.35
Female	117 (45.0)	91 (49.5)	208 (46.9)	
Male	143 (55.0)	93 (50.5)	236 (53.2)	
Race/Ethnicity				0.9
Non-Hispanic White	85 (32.7)	58 (31.5)	143 (32.2)	
Non-Hispanic Black	124 (47.7)	93 (50.5)	217 (48.9)	
Hispanic	38 (14.6)	26 (14.1)	64 (14.4)	
Other	13 (5.0)	7 (3.8)	20 (4.5)	
Stage at Diagnosis (%)				<0.0001
Stage 0/I	76 (29.2)	5 (2.7)	81 (18.2)	
Stage II	36 (13.9)	19 (10.3)	55 (12.4)	
Stage III	31 (11.9)	27 (14.7)	58 (13.1)	
Stage IV	12 (4.6)	46 (25.0)	58 (13.1)	
Unstaged	105 (40.4)	87 (47.3)	192 (43.2)	
Anatomic Site (%)				0.82
Colon	181 (69.6)	129 (70.1)	310 (69.8)	
Rectosigmoid junction	26 (10.0)	21 (11.4)	47 (10.6)	
Rectum	53 (20.4)	34 (18.5)	87 (19.6)	
Year of Diagnosis (%)				0.24
1995	23 (8.9)	12 (6.5)	35 (7.9)	
1996	26 (10.0)	12 (6.5)	38 (8.6)	
1997	22 (8.5)	16 (8.7)	38 (8.6)	
1998	27 (10.4)	17 (9.2)	44 (9.9)	
1999	30 (11.5)	31 (16.9)	61 (13.7)	
2000	25 (9.6)	25 (13.6)	50 (11.3)	
2001	42 (16.2)	19 (10.3)	61 (13.7)	
2002	21 (8.1)	21 (11.4)	42 (9.5)	
2003	22 (8.5)	20 (10.9)	42 (9.5)	
2004	22 (8.5)	11 (6.0)	33 (7.4)	

TABLE VII. CHARACTERISTICS OF UI HEALTH PATIENTS DIAGNOSED WITH COLORECTAL CANCER, 1995-2004, BY CONCENTRATED DISADVANTAGE AT DIAGNOSIS

	CD < median	CD > median	p
Age [mean (SD)]	59.8 (8.9)	60.0 (9.3)	0.8
Sex			0.295
Female	98 (44.1)	110 (49.5)	
Male	124 (55.9)	112 (50.5)	
Race/Ethnicity			<0.0001
Non-Hispanic White	128 (57.7)	15 (6.8)	
Non-Hispanic Black	49 (22.1)	168 (75.7)	
Hispanic	26 (11.7)	38 (17.1)	
Other	19 (8.6)	1 (0.5)	
Stage at Diagnosis			0.963
Stage 0/I	43 (19.4)	38 (17.1)	
Stage II	27 (12.2)	28 (12.6)	
Stage III	29 (13.1)	29 (13.1)	
Stage IV	27 (12.2)	31 (14.0)	
Unstaged	96 (43.2)	96 (43.2)	
Anatomic Site (%)			0.048
Colon	144 (64.9)	166 (74.8)	
Rectosigmoid junction	30 (13.5)	17 (7.7)	
Rectum	38 (21.6)	39 (17.6)	
Year of Diagnosis			0.339
1995	13 (5.9)	22 (9.9)	
1996	17 (7.7)	21 (9.5)	
1997	15 (6.8)	23 (10.4)	
1998	27 (12.2)	17 (7.7)	
1999	27 (12.2)	34 (15.3)	
2000	26 (11.7)	24 (10.8)	
2001	32 (14.4)	29 (13.1)	
2002	28 (12.6)	14 (6.3)	
2003	21(9.5)	21 (9.5)	
2004	16 (7.2)	17 (7.7)	

TABLE VIII. ASSOCIATION BETWEEN TUMOR CHARACTERISTICS AND RACE/ETHNICITY AMONG PEOPLE DIAGNOSED WITH COLORECTAL CANCER AT UI HEALTH, 1995-2004

	NHW	NHB	Hispanic	Other	p
N	143	217	64	20	
Stage at Diagnosis (%)					0.319
Stage 0/I	25 (17.5)	40 (18.4)	11 (17.2)	5 (25.0)	
Stage II	21 (14.7)	26 (12.0)	6 (9.4)	2 (10.0)	
Stage III	13 (9.1)	36 (16.6)	8 (12.5)	1 (5.0)	
Stage IV	16 (11.2)	27 (12.4)	14 (21.9)	1 (5.0)	
Unstaged	68 (47.6)	88 (40.6)	25 (39.1)	11 (55.0)	
Anatomic Site (%)					0.035
Proximal Colon	86 (60.1)	168 (77.4)	44 (68.8)	12 (60.0)	
Distal Colon	20 (14.0)	18 (8.3)	6 (9.4)	3 (15.0)	
Rectum	37 (25.9)	31 (14.3)	14 (21.9)	5 (25.0)	

There was no inherent order to the observed cancer stages because of the large share of cancers that were unstaged. Probability of having a stage II, III, IV, or unstaged cancer as opposed to stage 0/I was modeled using multinomial regression. Only anatomic site, male sex, and year of diagnosis were significantly associated with stage. The significant association with year of diagnosis largely reflected the fact that more recent cases were less likely to be unstaged. Results were similar for both CD score at diagnosis and 10-year lagged CD score, neither of which predicted CRC stage. Model results for the 10-year lagged CD score are shown in Table IX. Model results using CD score at diagnosis are shown in Table X.

Stage data were available in 363 cases used for the binary logistic regression models. Results were similar to those from the multinomial model. Neither CD score at diagnosis nor CD score 10 years before diagnosis was associated with having a more advanced cancer stage. Race/ethnicity remained unassociated with stage after adjustment.

Kaplan-Meier survival curves were calculated stratified by stage at diagnosis to determine whether cases with missing stage data could be included in the survival models (Figure 3). The

TABLE IX. MULTINOMIAL MODEL 1: PREDICTORS OF STAGE AT DIAGNOSIS WITH LAGGED CONCENTRATED DISADVANTAGE AMONG PEOPLE DIAGNOSED WITH COLORECTAL CANCER AT UI HEALTH, 1995-2004

	Stage II	Stage III	Stage IV	Unstaged
Age (per 10-year change)	0.02	0.0249	-0.2278	0.1569
Male	-0.0995	-.4002*	-0.081	-0.1424
Race/Ethnicity				
<i>Non-Hispanic Black</i>	-0.144	0.5074	-0.0384	-0.3682
<i>Hispanic</i>	-0.0834	0.3881	0.7066	-0.1118
<i>Other Race/Ethnicity</i>	-0.1798	-0.9065	-0.8308	0.2876
Anatomic Site				
<i>Colon</i>	0.2599	0.1893	0.2684	0.3984
<i>Rectosigmoid junction</i>	0.6853	0.4912	0.8131	-0.0242
Year of Diagnosis	-0.1332	-0.0711	-0.1288	-.2184**
CD Score at Diagnosis -10 years	0.1583	-0.197	0.1489	0.1799

*: p<0.05; **: p<0.01

TABLE X. MULTINOMIAL MODEL 2: PREDICTORS OF STAGE AT DIAGNOSIS WITH CONTEMPORARY CONCENTRATED DISADVANTAGE AMONG PEOPLE DIAGNOSED WITH COLORECTAL CANCER AT UI HEALTH, 1995-2004

	Stage II	Stage III	Stage IV	Unstaged
Age	0.0233	0.0193	-0.2038	-0.1555
Male	-0.1031	-0.403*	-0.1093	-0.142
Race/Ethnicity				
<i>Non-Hispanic Black</i>	-0.0936	0.7216	0.0541	-0.4848
<i>Hispanic</i>	-0.0720	0.4777	0.7962	-0.1567
<i>Other Race/Ethnicity</i>	-0.2167	-1.0992	-0.9402	0.3859
Anatomic Site				
<i>Colon</i>	0.2592	0.1926	0.2837	0.4026
<i>Rectosigmoid junction</i>	0.6820	0.4687	0.8060*	-0.0270
Year of Diagnosis	-0.1322	-0.0739	-0.1408	-0.2155**
CD Score at Diagnosis	0.0512	-0.7435	-0.1307	0.4304

*: p<0.05; **: p<0.01

survival experience of people with missing stage data was nearly identical to that of people with stage III CRC. Based on these results, cases with missing stage data were included in the survival models. Nineteen cases with stage 0 disease were excluded because there were no CRC deaths observed in this group. Stage at diagnosis was treated as a categorical variable in survival models.

Results of the main survival models are shown in Table XII. As expected, stage and year of diagnosis significantly predicted CRC-specific survival. Once accounting for stage, anatomic site was not associated with CRC-specific survival. There were no meaningful differences in survival by race/ethnicity, sex, or age at diagnosis.

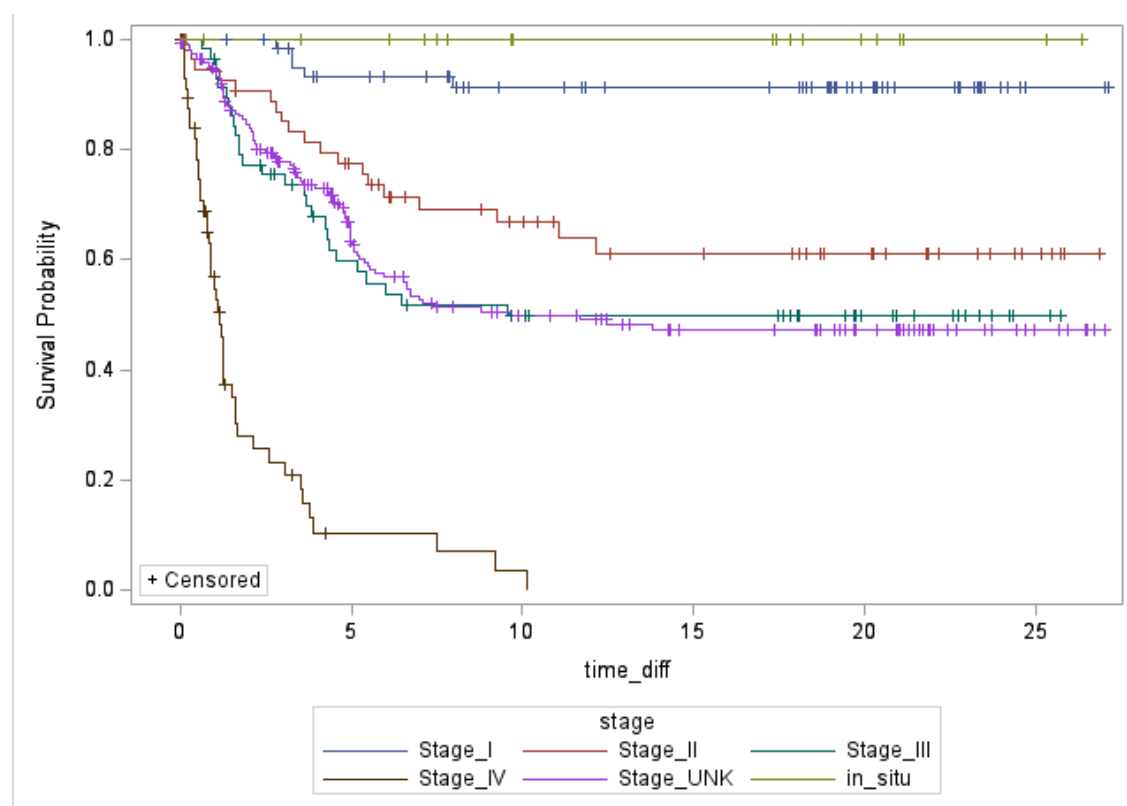


Figure 6. Kaplan-Meier survival estimates stratified by stage at diagnosis among UI Health patients diagnosed with colorectal cancer, 1995-2004

TABLE XI. LOGISTIC REGRESSION MODELS 1-2: PREDICTORS OF ADVANCED STAGE AT DIAGNOSIS AMONG PEOPLE DIAGNOSED WITH COLORECTAL CANCER AT UI HEALTH, 1995-2004, BY CHOICE OF CONCENTRATED DISADVANTAGE SCORE

		Model 1		Model 2	
		<i>OR</i>	<i>95% CI</i>	<i>OR</i>	<i>95% CI</i>
Age (per 10-year change)		0.85	0.67, 1.09	1.17	0.92, 1.49
Sex					
	Female	Ref		Ref	
	Male	0.81	0.52, 1.25	1.25	0.81, 1.94
Race/Ethnicity					
	Non-Hispanic White	Ref		Ref	
	Non-Hispanic Black	0.87	0.47, 1.63	1.14	0.62, 2.12
	Hispanic	1.12	0.54, 2.33	0.88	0.42, 1.82
	Other	1.00	0.35, 2.88	1.00	0.35, 2.88
Anatomic Site					
	Rectum	Ref		Ref	
	Rectosigmoid junction	0.96	0.44, 2.09	1.04	0.48, 2.26
	Colon	1.43	0.83, 2.27	0.70	0.40, 1.21
Year of Diagnosis		0.91	0.83, 0.99	1.10	1.01, 1.20
CD Score at Diagnosis -10 years		1.15	0.65, 2.03		
CD Score at Diagnosis				0.87	0.50, 1.54

Concentrated disadvantage score was a significant predictor of survival when measured at either main time point, but the association was somewhat stronger when measured at diagnosis. An increase of 1 in CD score measured at diagnosis 1—about two standard deviations in this sample—was associated with a 1.79 times increased hazard of CRC death (95% CI 1.21, 2.66). The same increase in CD score measured 10 years before diagnosis was associated with a 1.58 time increased hazard of CRC death (95% CI 1.08, 2.32). The trend in association between survival and CD scores measured at different times is shown in Figure 3. The adjusted effect of CD score was the same whether it was measured at 5 or 10 years before diagnosis. When measured 15 years before diagnosis, the effect of CD score was weaker and became non-significant. While all CD scores measured 10 or fewer years before diagnosis were associated with CRC survival, no information was added by developing a lagged CD score using commercial residential histories.

TABLE XII. SURVIVAL MODELS 1-2: PREDICTORS OF DISEASE-SPECIFIC DEATH AMONG PEOPLE DIAGNOSED WITH COLORECTAL CANCER AT UI HEALTH, 1995-2004, BY CHOICE OF CONCENTRATED DISADVANTAGE SCORE

		Model 1		Model 2	
		<i>HR</i>	<i>95% CI</i>	<i>HR</i>	<i>95% CI</i>
Age (per 10-year change)		0.92	0.78, 1.08	0.94	0.80, 1.09
Sex					
	Female	Ref		Ref	
	Male	1.00	0.74, 1.35	0.99	0.73, 1.33
Race/Ethnicity					
	Non-Hispanic White	Ref		Ref	
	Non-Hispanic Black	0.93	0.60, 1.45	0.84	0.54, 1.33
	Hispanic	0.77	0.47, 1.27	0.77	0.47, 1.26
	Other	0.92	0.41, 2.05	0.95	0.42, 2.11
Stage at Diagnosis					
	Stage I	Ref		Ref	
	Stage II	6.90	2.56, 18.59	6.98	2.59, 18.80
	Stage III	10.41	3.99, 27.16	11.29	4.32, 29.53
	Stage IV	77.74	30.14, 200.54	80.79	31.33, 208.34
	Unstaged	11.85	4.78, 29.36	11.62	4.69, 28.79
Anatomic Site					
	Rectum	Ref		Ref	
	Rectosigmoid Junction	1.12	0.64, 1.96	1.12	0.64, 1.96
	Colon	0.88	0.59, 1.30	0.89	0.60, 1.32
Year of Diagnosis		1.09	1.03, 1.16	1.10	1.03, 1.16
CD Score at Diagnosis -10 years		1.58	1.08, 2.32		
CD Score at Diagnosis				1.79	1.21, 2.66

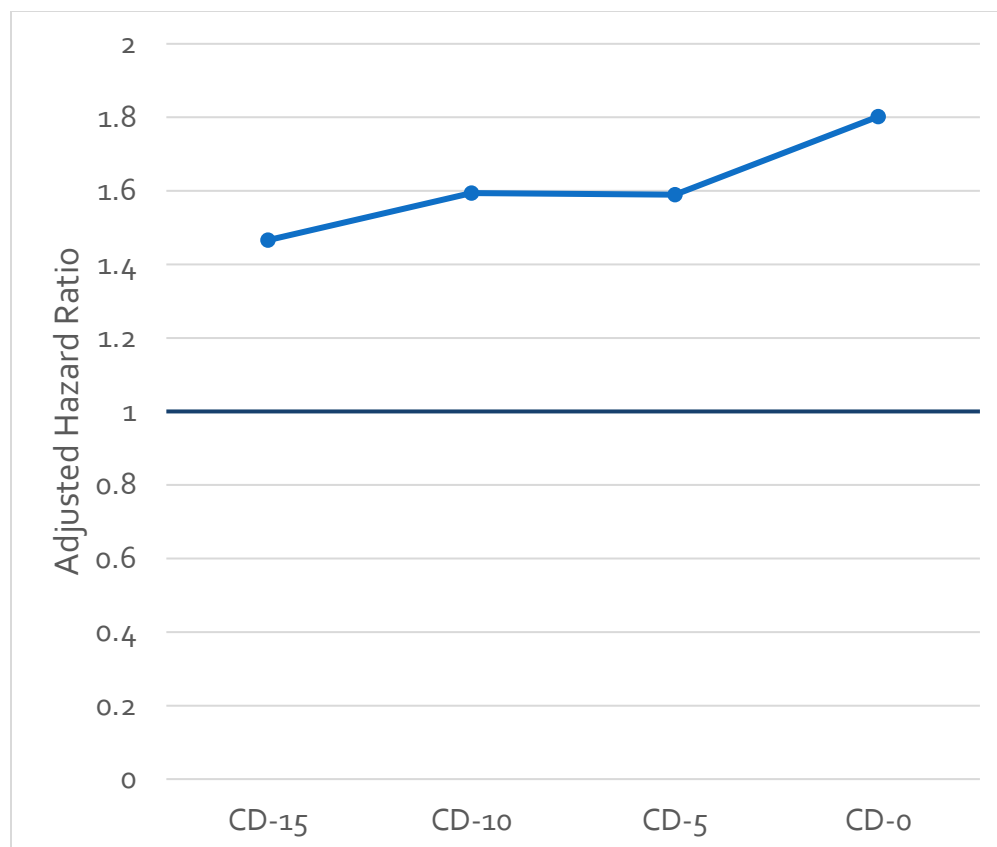


Figure 7. Trend in association between CD scores at different lag times and hazard of CRC-specific death in UI Health patients diagnosed with colorectal cancer, 1995-2004

D. Discussion

In this cohort of adults diagnosed with CRC between 1995 and 2004, CD score predicted disease-specific survival and the effect was stronger when CD was measured at diagnosis. No measure of CD score was associated with disease stage. No additional information was added to either analysis by considering CD exposure at 10 years before diagnosis.

Since lagged CD score did not predict disease stage, its association with survival is most likely due to its strong correlation with CD score at diagnosis, rather than because it represents a meaningful measure of historical exposure. In post hoc analysis, the correlation between contemporary and 10-year lagged CD score was 0.91 ($p < 0.001$).

Because neither historical nor contemporary CD was associated with stage at diagnosis, stage likely does not mediate the relationship between neighborhood CD—measured at any time—and survival outcomes. This finding was not sensitive to the inclusion or exclusion of cases with missing stage data. These results suggest that neighborhood conditions influence survival probability through a mechanism independent of anatomic spread, and that this effect occurs primarily around or after the time of diagnosis. The most obvious mechanism by which neighborhood conditions could influence survival but not stage at diagnosis is through the timely and consistent receipt of appropriate cancer-directed treatment. There is an extensive literature documenting that Black and white people with cancer experience similar survival outcomes when they are provided equivalent cancer-directed treatment (85,99). They also experience the same or similar outcomes when their treatment is provided in equal-access health systems, not only in CRC but in other cancer sites as well (27–32,86).

These results are consistent with prior work demonstrating that commercial residential history data, while “reasonably accurate and complete” (51) compared to self-reports, has the potential to introduce selection and information bias into epidemiologic studies (68). Commercial residential histories are assembled through public records including large federal databases, state

and local records, and credit reporting data (56). As a result, their completeness and accuracy may be affected by whether subjects have access to credit, own a home, whether they are registered to vote, or other characteristics that are known to vary by race, age, and sex. Credit histories, which rely on related data sources, are known to be less complete and accurate in Black and Hispanic people and in people who have low incomes or live in neighborhoods with lower average incomes (52). When we evaluated the completeness and accuracy of residential histories among people diagnosed with several cancers at UI Health between 2005 and 2016, we found that failure to match in the LexisNexis database was more common for people of any race or ethnicity other than non-Hispanic white, as well as for people who had died (68). Data loss related to this bias is cumulative and difficult to fully assess. Subjects may have some missing or unmatched addresses, resulting in study inclusion with less complete and accurate residential history data, or may have no matched addresses, potentially resulting in complete exclusion. When the study design requires data on historical exposures, as this one did, use of commercial address data may result in systematic exclusion of non-white, deceased, or lower-income subjects. This may explain why there was no difference in risk of death by race observed in this sample, an unexpected finding given the context. Available surveillance data show that CRC mortality was significantly higher in Black people as compared to whites since at least 2000 in Chicago, and since at least 1979 in Illinois (100–102).

An additional pitfall of this approach involved the use of historical disease data. The high percentage of cases with missing stage data, and the association between missing stage and year of diagnosis, strongly indicated a systemic problem with abstraction of disease data during the study period. Researchers interested in using residential histories should be aware that historical sources of case and exposure data may not meet currently accepted standards and anticipate that the underlying information may be difficult to recover.

These results have important implications for cancer disparities research in light of this limitation. The hypothesis that past neighborhood conditions contribute to cancer inequities has

strong theoretical support, in addition to the long latency periods associated with cancer and the evidence that present-day neighborhood conditions are associated with health outcomes. However, commercial public records databases may not be a viable source of information about these conditions. This study adds evidence that the dangers of selection and information bias associated with commercial residential histories are not only theoretical: studies that restrict their samples based on the availability of commercial data may reach misleading conclusions about the existence of serious racial health inequities. In this exploratory study, nearly as many cases were excluded based on lack of commercial public records as were excluded based on age at diagnosis. At the relevant stage of selection, the exclusion of 158 non-matched cases represented 25.9% of the sample (see Figure 1). Among analyzable cases, past neighborhood CD was not related to cancer prognosis and was a weaker predictor of survival than contemporary neighborhood CD. The association also weakened and became non-significant when CD score was measured more than 10 years before diagnosis, the typical latency period for CRC. The strong correlation between neighborhood CD at diagnosis and at a 10-year lag supports the interpretation that, at least among people able to be matched in the LexisNexis database, no information was added by using residential histories to measure neighborhood socioeconomic conditions in the past.

E. Conclusion

In this sample of Cook County, IL residents diagnosed with CRC between 1995 and 2004, neighborhood CD was associated with disease-specific survival but not with stage at diagnosis. Results are consistent with prior research demonstrating that neighborhood conditions at the time of cancer diagnosis predict disease-specific survival. However, no information was added by considering CD exposure in the past. Restricting study samples based on the availability of commercial residential history data may introduce selection bias into epidemiologic studies, distorting conclusions related to racial and socioeconomic inequities.

VI. CONCLUSION

A. Summary of Findings

The objective of this exploratory study was to evaluate the relationship between neighborhood disadvantage, experienced across the life course, and cancer outcomes. This study aimed to incorporate alternatives to commercial data into cancer disparities research using residential histories (Aim 1); explore the relationship between past and contemporary neighborhood environments in influencing cancer outcomes (Aim 2); and evaluate the association between past neighborhood conditions and cancer prognosis (Aim 3). The source of case and residential history data—the PSID for Aims 1-2 and a hospital tumor registry augmented by commercial public records data for Aim 3—were selected to balance one another’s strengths and weaknesses, potentially allowing a more holistic picture of neighborhood conditions and cancer inequities to emerge. In planning this study, we selected or created uniform sources of exposure and outcome data to be used for each aim: a tract-level CD score measure using harmonized Census estimates from the LTDB, and cancer-specific mortality from NDI. The condition of interest for Aim 3 was restricted to CRC. In the PSID sample used for Aims 1-2, additional cancer sites (prostate, breast, and lung) were studied.

Neighborhood CD score measured in the past did not provide meaningful information about survival after a cancer diagnosis in either sample. In the PSID sample, there was no association between CD score and all-cause survival, and this was true regardless of the time at which CD score was measured. In the tumor registry sample, in line with prior research, CD score did predict CRC-specific survival. However, the association was strongest when CD score was measured around the time of diagnosis. Because past and contemporary CD scores were strongly correlated, and because past CD score did not predict stage at diagnosis, this analysis provided no evidence that past exposure to neighborhood disadvantage exerts an independent effect on cancer outcomes. Results

and lessons learned from the research process are relevant to social epidemiologists with an interest in residential histories or in using social surveys to investigate causes of health inequities.

B. Comparison of Data and Results

1. Lack of Observed Racial Disparities in Either Data Set

In both study settings, there was no observed racial disparity in crude odds of death, raising the possibility of bias due to selection or attrition. Among PSID participants who reported a history of cancer between 2005 and 2019, white sample persons were non-significantly overrepresented among sample persons known to have died ($p = 0.09$). Among cases from the University of Illinois Hospital tumor registry with available residential history data, there were no meaningful or statistically significant differences in the racial distribution of deaths or censoring ($p = 0.9$).

These descriptive findings are not consistent with relevant national or local trends. Nationally, Black people are burdened by excessive incidence and mortality in CRC and prostate cancer in both sexes (103,104); excessive mortality in breast cancer among women (45); and excessive incidence and mortality from lung and bronchus cancer in men (104). Data from the Illinois State Cancer Registry show that CRC mortality was significantly higher in Black people as compared to whites since at least 2000 in Chicago, and since at least 1979 in Illinois (100–102).

2. Role of Residential Histories in Producing Selection or Attrition Bias

Multiple prior studies have demonstrated that commercial residential history data are less complete and accurate in Black and Hispanic people (50,68). Consumer credit reporting data, which relies on many of the same public records and is compiled by many of the same companies, has been shown to be less complete and accurate in Black and Hispanic people and to vary in quality by individual or neighborhood SES (13,52). Data availability also vary by age, vital status, and year (68). When study inclusion is conditioned on matching in a commercial public records database, selection bias is a potential result. The results presented in Chapter IV indicate that the potential for selection bias is not merely theoretical. Data collection approaches that rely on the availability of

commercial residential history data can produce final study samples that are not representative of important health inequities prevailing in the source population.

The question of bias due to attrition in the PSID is more complex. Only PSID sample persons, and not their other household members, were included in this study because these are the only individuals who routinely provide information about their own health history in 2005 and beyond. When the original study was designed, designation as a head of household or spouse was based on both marital status and sex. While the following rules have shifted over time, the basic definition of a PSID sample person remains: a person living in a PSID family in 1968, any adult child of a sample person, or their spouse. Using data from the 1968-1988 survey waves, Lillard and Panis found that risk of study attrition was related to race, sex, marital status, and marital transitions, and that marital status was related to mortality risk. However, they also found that study attrition contributed minimal bias to their analysis of mortality (105).

However, more recent analyses of the health data in the PSID have found evidence of non-representativeness that could be explained by systematic differences in either study attrition or risk of death. Health data from the PSID can be used to derive prevalence estimates that are comparable to those obtained from the NHIS (15,16). Using data from the 1999-2005 survey waves, Zajacova, et al. found that PSID data can also be used to calculate cancer prevalence estimates comparable to those obtained from SEER—however, cancer prevalence in the PSID was gradually increasing, with no obvious explanation (53). Sample persons in the PSID are also much more likely to report that they have limited ability to work compared to NHIS respondents (15). If PSID participants are living longer with chronic illnesses or activity limitations compared to the general population, and racial survival disparities are truly reduced or absent in non-attriters, this could explain the gradually increasing prevalence of cancer in this cohort.

C. Role of Missing Case Data

1. Panel Study of Income Dynamics

i. Death Data

The PSID was designed to collect information about the social and economic conditions of families, and many of the health-related questions were not added until the 1990s. As a result, health information relevant to epidemiologists may be missing, incomplete, or collected in ways that health researchers do not expect. The most significant form of missing data in this study was the lack of information from death certificates.

According to multiple communications with PSID staff, linked data from NDI was unable to be shared with outside researchers from approximately 2017 to 2023 (106,107). Publicly available information on the PSID website did not make clear that the mortality file available to researchers would include only data collected by the University of Michigan. Public documentation for the then-current mortality file described variables only available through linkage to NDI (108). The reduced mortality file included fields 1-24 of the 104 mentioned in the documentation, corresponding to identifiers and death information collected by PSID staff; and respondent information sent to NDI (108). Although the approval process to access restricted data requires that researchers submit a detailed research plan for review, the intention not to share death certificate data was not disclosed until after the contract was executed and payment transmitted in spring of 2021 (109). Researcher access to the full mortality file was restored on September 13, 2023 (107).

As a result of this omission, the analysis plan and relevant aims were updated to use all-cause mortality in people with a history of cancer as the endpoint.

ii. Exact Dates and Self-Reported Diagnosis Timing

Researchers planning to use the full mortality file should be aware that the PSID does not typically release full date information associated with participants, even in the restricted data files. Most dates, including date of birth and date of death, if applicable, include the month and year only.

Sample persons who report a current health condition, including cancer, report the age at which the condition was first diagnosed, or the event first occurred.

These data can still be used for time to event analysis if either months of follow-up or age is used as the time scale. There are advantages to using age as the time scale when using data drawn from the PSID. Because it is designed to be representative of the US population, the PSID includes households and sample persons of varied ages. When age is treated as the time scale in survival analysis, subjects enter the risk set at the age of cancer diagnosis, rather than on a date associated with their study participation. This method is adapted from Korn, et al., who argue that, for outcomes related to cancer, risk is expected to change more as a function of age than of length of study participation (63). This approach also adjusts for age while allowing a non-parametric age effect. An additional modification used in this study, and recommended by Canchola, et al., stratifies the model by birth cohort (64). This accounts for expected calendar effects on risk of death in this long-running panel study.

Because PSID participants are asked if they have ever been diagnosed with cancer (among other conditions), in any given year they may report multiple conditions or repeat a previously reported condition. Using repeat reports from the 1999-2005 survey waves, Zajacova, et al. found evidence of measurement error in reported age at cancer diagnosis (53). Inconsistent reports of diagnosis age were random across social and demographic groups but have the potential to reduce analytic power. Self-report errors may be even more significant in less severe health conditions. Researchers who plan to use PSID data related to the timing of health events should use caution, thoroughly explore publicly available data to ascertain quality and potential analytic power, and plan related analyses that do not depend on the accuracy of self-reported timing of diagnosis.

iii. Disease Information

From 2005 onward, PSID participants who report a history of cancer have been asked if they are “in treatment”, “in remission”, “or has it been cured?” This variable was not used for

analysis because, unlike the parent question, which asked about cancer history, this question referred to present-day disease status. Each year that it was asked, fewer than 1% of respondents reported currently receiving cancer treatment. Given the significant error found in reports of cancer timing, it was also unclear whether participants could give accurate self-reports of whether their cancer was in remission vs. cured. Not all chronic conditions surveyed in the PSID include any follow-up questions about treatment or condition severity.

Sample persons have answered questions about their ability to participate in activities of daily living since 1999 and have rated their general health status since 1984. For some conditions, researchers may consider these variables as a source of supplementary information about mortality risk.

2. University of Illinois Hospital Tumor Registry

In this sample of 444 cases diagnosed with CRC at University of Illinois Hospital between 1995 and 2004, a very high percentage were recorded as having unstaged disease. Stage is a critical indicator of cancer-specific mortality risk, and missingness on this variable presented an analytic challenge.

The survival experience of these cases was very similar to that of cases with stage III disease. Approaches used to analyze the data included grouping the unstaged diseases with one or more of the advanced stages, exclusion, and including the unstaged cases while treating stage as categorical in modeling.

No approach identified a meaningful association between CD score and stage at diagnosis. When treated as a categorical variable, stage significantly predicted disease-specific survival and the adjusted effect of having unstaged disease remained very similar to the adjusted effect of having stage III disease.

Incidence of unstaged cancer has decreased over time, but is more common in older people and in people who are any race or ethnicity other than non-Hispanic white (110,111). Previous

analyses in SEER have found that lack of stage information may be related to not receiving treatment, resulting in missing information about tumor size, lymph node involvement, or metastasis, especially before PET imaging became a common procedure in cancer staging (110,111).

When unstaged cases are retained in population-based CRC studies, it is common for their risk of disease-specific mortality to fall between the risks associated with regional and distant disease (76). In that regard, the stage variable in this sample performed as expected. However, there were no significant racial differences in the risk of having either late-stage disease or unstaged disease. This is not reflective of well-documented racial cancer inequities, in which non-Hispanic Black people are at increased risk of both late stage cancer and unstaged cancer compared to non-Hispanic whites (104).

D. Role of Exposure Specification

This study focused on a single measure of neighborhood conditions, concentrated disadvantage. Negative results may reflect a true lack of association between past concentrated disadvantage and cancer outcomes, or a limitation of other aspects of the analytic approach, rather than a limitation of residential history data itself. When it is available, residential history data represents a rich and complex form of information. Other researchers have used these data to study residential trajectories over time, residential instability, or spatiotemporal cancer risk for hypothesis generation. These approaches may be more informative for some research questions.

Using data from the 2003-2017 waves of the PSID and restricting the study sample to low income households, Kang employed sequence analysis to identify four distinct types of housing trajectories among families who had experienced housing instability (112). This study identified risk factors for involuntary moves and extended periods of housing instability among PSID participants, including marital/partner transitions, job loss, household age, and chronic health conditions among household members. This study also demonstrates the diversity of residential

histories that may be observed among PSID participants, as well as the many potential approaches to summarizing or measuring these trajectories. When available residential history data are believed to be adequate for the study approach, residential mobility may be an important factor in itself; however, researchers should use caution since moves may be undercounted in some groups (11).

Hurley, et al. (2005) used data from the California Teachers Study to describe study participants' lifetime residential mobility and its potential association with breast cancer risk (113). They found that residential stability was associated with living in higher-SES neighborhoods and with older age. Their results suggested that residence at diagnosis may reflect some aspects of women's residential histories, especially aspects related to living in urban areas. In a separate study, Hurley, et al. (2017) evaluated whether commercial residential histories obtained from LexisNexis agreed with histories provided by California Teachers Study participants (50). While they added previously unavailable residence information during the reproductive years for many participants, they also found that these data were less complete in Black women and younger women. Their approach may be useful to researchers who already have access to partial residential histories and wish to enhance them, rather than restricting their study to participants with commercial residential history available.

Finally, researchers have used a combination of cancer registries and commercial residential history data to identify potential clusters of cancer cases at specific locations and points in time (49,92,114–117). These studies typically have a case control design and do not specify a particular exposure, other than living in a particular area at a given time. This approach is particularly relevant to identifying potential environmental carcinogens and the investigation of cancer clusters, an important motivation for the use of residential history data. These approaches may be able to be adapted by researchers interested in social or political determinants of health, for

example, to study whether disease risks have changed over time in areas affected by segregation or policy change.

E. Future Directions

1. Using PSID for Social Epidemiology

Researchers interested in using PSID data should carefully consider the role of place in their studies. While segregation may operate at multiple scales—for example, in both neighborhoods and across an entire metropolitan region—the specific policies and practices that produce residential segregation may vary by jurisdiction. Multiple researchers have recommended studying segregation and its consequences in smaller areas, both to ascertain the specific patterns of segregation that are present, and to tailor research questions to potential policy solutions (2,78). Therefore, the approach taken in this study—focused on a single measure of neighborhood conditions, but agnostic to place—may not be the best suited to understand the hypothesized relationship in which exposure to segregation might influence survival with cancer.

Geographic PSID data include tract-level residential history information for every year that a given individual was represented in the survey. This information is available in the restricted geocode match file used for this study (71). Researchers who plan to use restricted data may also access the state(s) in which participants were born and died according to their death certificate in the updated mortality file. For research questions that will use data on the full national sample of participants, location data available from the public data index may actually be more useful and appropriate. The PSID Family Public Data Index includes the current state and region of residence and the self-reported state and region where participants were born and grew up. These variables provide potential links to explore the effects of state policies on social and economic well-being, health, and ultimately, mortality.

2. Weighing the Probability of Data Loss and Selection Bias When Using Residential Histories to Study Marginalized Populations

Using both commercial and survey data, this study found no evidence that CD scores derived from residential histories were associated with cancer outcomes. However, it confirmed previous findings that CD score around the time of diagnosis does predict disease-specific survival in people with cancer. It also provided evidence that conditioning study inclusion on the availability of extensive residential history data may result in an unacceptable degree of loss of other data: the diverse cancer and survival experiences of people harmed by racial health inequities. In multiple settings, the requirement that a person be observed repeatedly for years or decades resulted in study samples that were not representative of cancer inequities prevailing in the general population. Collecting or estimating increasingly detailed information about a relatively small group of individuals did not provide new information about a systemic inequity that has been well documented.

3. Acting on Evidence That Neighborhood Disadvantage at Diagnosis Contributes to Preventable Cancer Deaths Due to Systemic Racism

This study confirmed that neighborhood conditions around the time of cancer diagnosis do predict disease-specific survival, even if no information is added by looking back in time. As described above, the observed lack of association between past neighborhood CD and cancer prognosis is most likely a result of the multiple limitations of both the available data and the analytic approach.

Segregation has been described as a fundamental cause of health inequities through its influence on access to resources (37). This relationship is supported by evidence that access to relevant resources, especially health care, is associated with cancer outcomes and partially accounts for outcome disparities (5,76,77). Evidence also supports a relationship between concentrated disadvantage and cancer disparities (17,18).

A diverse body of evidence points to health care inequities as a driver of deadly disparities in cancer and other chronic diseases, even when neighborhood conditions contribute. These disparities occur across the cancer continuum, yet are reduced or absent when Black and white patients are treated in equal-access health systems and provided with the same recommended treatments (28–30,32,86). Neighborhood conditions may influence survival after diagnosis because they relate to multiple dimensions of potential access to health care, including spatial access and unmeasured elements of SES. They may also influence a person's general health status and comorbidities beyond their cancer diagnosis or relate to the social support available to cope with a severe health condition.

The findings of this study do not contradict the extensive evidence in support of its primary motivation: racial inequities in cancer outcomes are caused by racism. They can and must be prevented.

CITED LITERATURE

1. American Cancer Society. Colorectal Cancer Facts & Figures 2017-2019 [Internet]. Atlanta, GA: American Cancer Society; 2017. Available from: <https://www.cancer.org/research/cancer-facts-statistics.html>
2. Landrine H, Corral I, Lee JGL, Efrid JT, Hall MB, Bess JJ. Residential Segregation and Racial Cancer Disparities: A Systematic Review. *J Racial Ethn Health Disparities*. 2016 Dec 30;1-11.
3. Keum N, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat Rev Gastroenterol Hepatol*. 2019 Aug 27;1-20.
4. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut*. 2017 Apr 1;66(4):683-91.
5. Wan N, Zhan FB, Lu Y, Tiefenbacher JP. Access to healthcare and disparities in colorectal cancer survival in Texas. *Health Place*. 2012 Mar;18(2):321-9.
6. Wang F, Luo L, McLafferty S. Healthcare access, socioeconomic factors and late-stage cancer diagnosis: an exploratory spatial analysis and public policy implication. *Int J Public Policy*. 2010;5(2-3):237-58.
7. McLafferty S, Wang F. Rural Reversal? Rural-Urban Disparities in Late-stage Cancer Risk in Illinois. *Cancer*. 2009 Jun 15;115(12):2755-64.
8. LaVeist T, Pollack K, Thorpe R, Fesahazion R, Gaskin D. Place, Not Race: Disparities Dissipate In Southwest Baltimore When Blacks And Whites Live Under Similar Conditions. *Health Aff (Millwood)*. 2011 Oct 1;30(10):1880-7.
9. Cagney KA, Browning CR. Exploring neighborhood-level variation in asthma and other respiratory diseases: the contribution of neighborhood social context. *J Gen Intern Med*. 2004 Mar;19(3):229-36.
10. Browning CR, Cagney KA. Neighborhood structural disadvantage, collective efficacy, and self-rated physical health in an urban setting. *J Health Soc Behav*. 2002 Dec;43(4):383-99.
11. Hughes AE, Tiro JA, Balasubramanian BA, Skinner CS, Pruitt SL. Social disadvantage, healthcare utilization, and colorectal cancer screening: Leveraging longitudinal patient address and electronic health records data. *Cancer Epidemiol Prev Biomark*. 2018 Jan 1;cebp.0446.2018.
12. Hughes AE, Pruitt SL. The Utility of EMR Address Histories for Assessing Neighborhood Exposures. *Ann Epidemiol*. 2017 Jan;27(1):20-6.
13. Brevoort KP, Grimm P, Kambara M. Credit Invisibles. Consumer Financial Protection Bureau Office of Research; 2015 May p. 37. (Data Point).
14. Ludwig S. Credit scores in America perpetuate racial injustice. Here's how. *The Guardian* [Internet]. 2015 Oct 13 [cited 2020 Jan 24]; Available from: <https://www.theguardian.com/commentisfree/2015/oct/13/your-credit-score-is-racist-heres-why>

15. Insolera NE, Freedman VA. Comparing Health Estimates in the PSID and NHIS, 2001-2015. Ann Arbor, MI: Panel Study of Income Dynamics, University of Michigan Institute for Social Research; 2017 May p. 13. (Technical Series). Report No.: 17-01.
16. Andreski P, McGonagle K, Schoeni B. An Analysis of the Quality of the Health Data in the Panel Study of Income Dynamics. 2009 Sep 15;18.
17. Freeman VL, Ricardo AC, Campbell RT, Barrett RE, Warnecke RB. Association of census tract-level socioeconomic status with disparities in prostate cancer-specific survival. *Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol*. 2011 Oct;20(10):2150-9.
18. Peterson CE, Rauscher GH, Johnson TP, Kirschner CV, Freels S, Barrett RE, et al. The Effect of Neighborhood Disadvantage on the Racial Disparity in Ovarian Cancer-Specific Survival in a Large Hospital-Based Study in Cook County, Illinois. *Front Public Health*. 2015;3.
19. Berry J, Bumpers K, Ogunlade V, Glover R, Davis S, Counts-Spriggs M, et al. Examining racial disparities in colorectal cancer care. *J Psychosoc Oncol*. 2009;27(1):59-83.
20. Bromley EG, May FP, Federer L, Spiegel BMR, van Oijen MGH. Explaining persistent under-use of colonoscopic cancer screening in African Americans: a systematic review. *Prev Med*. 2015 Feb;71:40-8.
21. Clegg L, Li F, Hankey B, Chu K, Edwards B. Cancer survival among us whites and minorities: A seer (surveillance, epidemiology, and end results) program population-based study. *Arch Intern Med*. 2002 Sep 23;162(17):1985-93.
22. Cooper GS, Yuan Z, Landefeld CS, Rimm AA. Surgery for colorectal cancer: Race-related differences in rates and survival among Medicare beneficiaries. *Am J Public Health*. 1996 Apr;86(4):582-6.
23. Demissie K, Oluwole OO, Balasubramanian BA, Osinubi OO, August D, Rhoads GG. Racial differences in the treatment of colorectal cancer: a comparison of surgical and radiation therapy between Whites and Blacks. *Ann Epidemiol*. 2004 Mar;14(3):215-21.
24. Govindarajan R, Shah RV, Erkman LG, Hutchins LF. Racial differences in the outcome of patients with colorectal carcinoma. *Cancer*. 2003 Jan 15;97(2):493-8.
25. Jinjuvadia R, Jinjuvadia K, Liangpunsakul S. Racial Disparities in Gastrointestinal Cancers-Related Mortality in the US Population. *Dig Dis Sci*. 2013 Jan;58(1):236-43.
26. Schwartz KL, Crossley-May H, Vigneau FD, Brown K, Banerjee M. Race, socioeconomic status and stage at diagnosis for five common malignancies. *Cancer Causes Control*. 2003 Oct 1;14(8):761-6.
27. Dolan NC, Ferreira MR, Fitzgibbon ML, Davis TC, Rademaker AW, Liu D, et al. Colorectal Cancer Screening Among African-American and White Male Veterans. *Am J Prev Med*. 2005 Jun;28(5):479-82.

28. Dominitz JA, Samsa GP, Landsman P, Provenzale D. Race, treatment, and survival among colorectal carcinoma patients in an equal-access medical system. *Cancer*. 1998 Jun 15;82(12):2312–20.
29. Hassan MO, Arthurs Z, Sohn VY, Steele SR. Race does not impact colorectal cancer treatment or outcomes with equal access. *Am J Surg*. 2009 Apr;197(4):485–90.
30. Gill AA, Enewold L, Zahm SH, Shriver CD, Stojadinovic A, McGlynn KA, et al. Colon cancer treatment: Are there racial disparities in an equal-access healthcare system? *Dis Colon Rectum*. 2014 Sep;57(9):1059–65.
31. Rabeneck L, Soucek J, El-Serag HB. Survival of colorectal cancer patients hospitalized in the Veterans Affairs Health Care System. *Am J Gastroenterol*. 2003 May;98(5):1186–92.
32. Lee S, Reha JL, Tzeng CWD, Massarweh NN, Chang GJ, Hetz SP, et al. Race Does Not Impact Pancreatic Cancer Treatment and Survival in an Equal Access Federal Health Care System. *Ann Surg Oncol*. 2013 Dec;20(13):4073–9.
33. Pan J, Xin L, Ma YF, Hu LH, Li ZS. Colonoscopy Reduces Colorectal Cancer Incidence and Mortality in Patients With Non-Malignant Findings: A Meta-Analysis. *Am J Gastroenterol*. 2016 Mar;111(3):355–65.
34. Chicago Department of Public Health. Chicago Health Atlas. 2018 [cited 2019 Nov 26]. Colorectal Cancer Screening. Available from: <https://www.chicagohealthatlas.org/indicators/colorectal-cancer-screening>
35. Chicago Department of Public Health. Chicago Health Atlas. 2019 [cited 2019 Nov 26]. Colorectal Cancer Incidence. Available from: <https://www.chicagohealthatlas.org/indicators/colorectal-cancer-incidence>
36. Chicago Department of Public Health. Chicago Health Atlas. 2019 [cited 2019 Nov 26]. Colorectal Cancer Deaths. Available from: <https://www.chicagohealthatlas.org/indicators/colorectal-cancer-deaths>
37. Williams DR, Collins C. Racial Residential Segregation: A Fundamental Cause of Racial Disparities in Health. *Public Health Rep*. 2001;116:13.
38. Link BG, Phelan J. Social conditions as fundamental causes of disease. *J Health Soc Behav*. 1995;Spec No:80–94.
39. Diez-Roux AV. Neighborhoods and health: where are we and where do we go from here? *Rev Epidemiol Sante Publique*. 2007 Feb;55(1):13–21.
40. Kolak M, Bradley M, Block DR, Pool L, Garg G, Toman CK, et al. Urban foodscape trends: Disparities in healthy food access in Chicago, 2007–2014. *Health Place*. 2018 Jul 1;52:231–9.
41. Chicago Area HIV Integrated Services Council. Chicago-Area Unified HIV Plan for HIV Prevention, Care, Housing and Essential Services 2014–2016 [Internet]. Chicago, IL; 2014 [cited 2020 Feb 4]. Available from: https://www.chicago.gov/content/dam/city/depts/cdph/HIV_STI/Chicago_Area_HIV_Unified_Plan.pdf

42. Havekes E, Bader M, Krysan M. Realizing Racial and Ethnic Neighborhood Preferences? Exploring the Mismatches Between What People Want, Where They Search, and Where They Live. *Popul Res Policy Rev*. 2016 Feb 1;35(1):101–26.
43. Logan JR, Alba RD, McNulty T, Fisher B. Making a Place in the Metropolis: Locational Attainment in Cities and Suburbs. *Demography*. 1996 Nov;33(4):443.
44. Woldoff RA. Wealth, Human Capital and Family across Racial/Ethnic Groups: Integrating Models of Wealth and Locational Attainment. *Urban Stud*. 2008 Mar;45(3):527–51.
45. American Cancer Society. Breast Cancer Facts & Figures 2019-2020 [Internet]. Atlanta, GA: American Cancer Society; 2019. Available from: <https://www.cancer.org/research/cancer-facts-statistics.html>
46. Goodman M, Naiman JS, Goodman D, LaKind JS. Cancer clusters in the USA: what do the last twenty years of state and federal investigations tell us? *Crit Rev Toxicol*. 2012 Jul;42(6):474–90.
47. Goodman M, LaKind JS, Fagliano JA, Lash TL, Wiemels JL, Winn DM, et al. Cancer Cluster Investigations: Review of the Past and Proposals for the Future. *Int J Environ Res Public Health*. 2014 Feb;11(2):1479–99.
48. Oudin A, Forsberg B, Strömgren M, Beelen R, Modig L. Impact of Residential Mobility on Exposure Assessment in Longitudinal Air Pollution Studies: A Sensitivity Analysis within the ESCAPE Project. *Sci World J* [Internet]. 2012 Nov 28 [cited 2019 Aug 22];2012. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3515908/>
49. Jacquez GM, Slotnick MJ, Meliker JR, AvRuskin G, Copeland G, Nriagu J. Accuracy of Commercially Available Residential Histories for Epidemiologic Studies. *Am J Epidemiol*. 2011 Jan 15;173(2):236–43.
50. Hurley S, Hertz A, Nelson DO, Layefsky M, Von Behren J, Bernstein L, et al. Tracing a Path to the Past: Exploring the Use of Commercial Credit Reporting Data to Construct Residential Histories for Epidemiologic Studies of Environmental Exposures. *Am J Epidemiol*. 2017 Feb 1;185(3):238–46.
51. Stinchcomb DG, Roeser A. NCI/SEER Residential History Project Technical Report [Internet]. Rockville, MD: Westat, Inc.; 2016. Available from: <https://www.westat.com/tools-for-using-commercial-sources-of-residential-histories-for-cancer-research/>
52. Past Imperfect: How Credit Scores and Other Analytics “Bake In” and Perpetuate Past Discrimination [Internet]. Racial Justice & Equal Economic Opportunity Project, National Consumer Law Center; 2016 May [cited 2020 Jan 24]. Available from: https://www.nclc.org/images/pdf/credit_discrimination/Past_Imperfect050616.pdf
53. Zajacova A, Dowd JB, Schoeni RF, Wallace RB. Consistency and precision of cancer reporting in a multiwave national panel survey. *Popul Health Metr*. 2010 Dec;8(1):1–11.
54. Panel Study of Income Dynamics: 1968-2019 Mortality File Documentation [Internet]. Institute for Social Research, University of Michigan; 2021 Mar [cited 2023 Feb 4]. Report No.: Release 1. Available from: https://simba.isr.umich.edu/restricted/docs/Mortality/Mortality19_Introduction.pdf

55. Survey Research Center, Institute for Social Research, University of Michigan. Panel Study of Income Dynamics, public use dataset. Ann Arbor, MI; 2019.
56. LexisNexis Risk Solutions [Internet]. [cited 2023 Sep 2]. Accurint® for Government. Available from: <https://risk.lexisnexis.com/products/accurint-for-government>
57. Freeman VL, Boylan EE, Tilahun NY, Basu S, Kwan MP. Sources of selection and information biases when using commercial database-derived residential histories for cancer research. *Ann Epidemiol*. 2020 Nov;51:35-40.e1.
58. Logan JR, Xu Z, Stults BJ. Longitudinal Tract Database (LTDB) [Internet]. 2014. Available from: <https://s4.ad.brown.edu/projects/diversity/Researcher/LTDB.htm>
59. Logan JR, Xu Z, Stults B. Interpolating U.S. Decennial Census Tract Data from as Early as 1970 to 2010: A Longitudinal Tract Database. *Prof Geogr J Assoc Am Geogr*. 2014 Jul 1;66(3):412–20.
60. US Department of Commerce. Census Bureau Data [Internet]. [cited 2023 Sep 11]. Available from: <https://data.census.gov/>
61. National Center for Health Statistics. National Death Index user's guide. Hyattsville, MD; 2013.
62. 24560 - How can I linearly interpolate between the values in my data? [Internet]. [cited 2023 Aug 21]. Available from: <http://support.sas.com/kb/24/560.html>
63. Korn EL, Graubard BI, Midthune D. Time-to-Event Analysis of Longitudinal Follow-up of a Survey: Choice of the Time-scale. *Am J Epidemiol*. 1997 Jan 1;145(1):72–80.
64. Canchola A, Stewart S, Center NCC, Bernstein L. Cox Regression Using Different Time Scales.
65. Nadler DL, Zurbenko IG. Estimating Cancer Latency Times Using a Weibull Model. *Adv Epidemiol*. 2014 Aug 31;2014:1–8.
66. Lee JK, Jensen CD, Levin TR, Zauber AG, Schottinger JE, Quinn VP, et al. Long-term Risk of Colorectal Cancer and Related Deaths After a Colonoscopy With Normal Findings. *JAMA Intern Med*. 2019 Feb;179(2):153–60.
67. PA-17-298: Integration of Individual Residential Histories into Cancer Research (R01) [Internet]. [cited 2019 Nov 19]. Available from: <https://grants.nih.gov/grants/guide/pa-files/PA-17-298.html>
68. Freeman VL, Boylan EE, Tilahun NY, Basu S, Kwan MP. Sources of selection and information biases when using commercial database-derived residential histories for cancer research. *Ann Epidemiol*. 2020 Nov 1;51:35-40.e1.
69. Survey Research Center, Institute for Social Research, University of Michigan. Panel Study of Income Dynamics, restricted use data. Ann Arbor, MI; 2019.
70. US Census Bureau. TIGER/Line Shapefiles [Internet]. 2021. Available from: <https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html>

71. Documentation for the 2010 Census PSID Geocode Match File. Ann Arbor, MI: Institute for Social Research, University of Michigan; 2021 Dec.
72. Division of Consumer Prices and Price Indexes, US Bureau of Labor Statistics. Consumer Price Index Research Series Using Current Methods (R-CPI-U-RS), 1977 - 2020 [Internet]. 2021 [cited 2021 Aug 27]. Available from: <https://www.bls.gov/cpi/research-series/r-cpi-u-rs-home.htm>
73. PSID Main Interview User Manual: Release 2021 [Internet]. Institute for Social Research, University of Michigan; 2021 Mar [cited 2023 Jan 21]. Available from: <https://psidonline.isr.umich.edu/data/Documentation/UserGuide2019.pdf>
74. National Center for Health Statistics. National Health Interview Survey (NHIS) Questionnaire [Internet]. 2023 [cited 2023 Sep 30]. Available from: https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Survey_Questionnaires/NHIS/2023/EnglishQuest-508.pdf
75. National Center for Health Statistics. National Health Interview Survey (NHIS) Questionnaire [Internet]. 1999 [cited 2023 Sep 30]. Available from: https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Survey_Questionnaires/NHIS/1999/QSAMADLT.pdf
76. Freeman VL, Naylor KB, Boylan EE, Booth BJ, Pugach O, Barrett RE, et al. Spatial access to primary care providers and colorectal cancer-specific survival in Cook County, Illinois. *Cancer Med.* 2020 May;9(9):3211–23.
77. Wang F, McLafferty S, Escamilla V, Luo L. Late-Stage Breast Cancer Diagnosis and Health Care Access in Illinois. *Prof Geogr J Assoc Am Geogr.* 2008 Feb;60(1):54–69.
78. Riley AR. Neighborhood Disadvantage, Residential Segregation, and Beyond—Lessons for Studying Structural Racism and Health. *J Racial Ethn Health Disparities.* 2018 Apr 1;5(2):357–65.
79. Wan N, Zhan FB, Zou B, Chow E. A relative spatial access assessment approach for analyzing potential spatial access to colorectal cancer services in Texas. *Appl Geogr.* 2012 Mar 1;32(2):291–9.
80. McLafferty S, Wang F, Luo L, Butler J. Rural – urban inequalities in late-stage breast cancer: spatial and social dimensions of risk and access. *Environ Plan B Plan Des.* 2011;38(4):724–40.
81. Johnson RC, Schoeni RF, Rogowski JA. Health disparities in mid-to-late life: the role of earlier life family and neighborhood socioeconomic conditions. *Soc Sci Med* 1982. 2012 Feb;74(4):625–36.
82. American Cancer Society. Colorectal Cancer Facts & Figures 2020-2022 [Internet]. Atlanta, GA: American Cancer Society; 2020. Available from: <https://www.cancer.org/research/cancer-facts-statistics.html>
83. Czito BG, Willett CG. Thirty years of rectal cancer research: a brief history. *Oncol Williston Park N.* 2008 Nov 15;22(12):1441–2, 1444.

84. American Society of Clinical Oncology. ASCO. [cited 2022 Aug 6]. Colorectal Cancer Progress Timeline. Available from: <https://www.asco.org/research-guidelines/cancer-progress-timeline/colorectal-cancer>
85. Haller DG, Catalano PJ, Macdonald JS, O'Rourke MA, Frontiera MS, Jackson DV, et al. Phase III study of fluorouracil, leucovorin, and levamisole in high-risk stage II and III colon cancer: final report of Intergroup 0089. *J Clin Oncol Off J Am Soc Clin Oncol*. 2005 Dec 1;23(34):8671–8.
86. Eaglehouse YL, Georg MW, Shriver CD, Zhu K. Racial Comparisons in Timeliness of Colon Cancer Treatment in an Equal-Access Health System. *JNCI J Natl Cancer Inst*. 2020 Apr 1;112(4):410–7.
87. Freeman VL, Naylor KB, Boylan EE, Booth BJ, Pugach O, Barrett RE, et al. Spatial access to primary care providers and colorectal cancer-specific survival in Cook County, Illinois. *Cancer Med*. 2020 Mar 4;
88. LaVeist TA, Gaskin D, Trujillo AJ. Segregated Spaces, Risky Places: The Effects of Racial Segregation on Health Inequities [Internet]. Washington, DC: Joint Center for Political and Economic Studies; 2011 Sep [cited 2020 Mar 5]. Available from: <https://www.racialequitytools.org/resourcefiles/SegregatedSpaces.pdf>
89. Nadia Hassanein, Grace Hauck, Jayme Fraser, Aleszu Bajak. “Just not equal at all”: Vaccine rollout in Chicago a microcosm of racial disparities nationwide. *USA Today* [Internet]. 2021 Feb 12 [cited 2021 Apr 30]; Available from: <https://www.usatoday.com/in-depth/news/health/2021/02/12/data-analysis-chicago-vaccine-rollout-reflects-us-racial-disparities/4418978001/>
90. Perry AM, Harshbarger D, Romer C. Mapping racial inequity amid COVID-19 underscores policy discriminations against Black Americans [Internet]. The Brookings Institution; 2020 [cited 2022 Jan 24]. Available from: <https://www.brookings.edu/blog/the-avenue/2020/04/16/mapping-racial-inequity-amid-the-spread-of-covid-19/>
91. Keene DE, Geronimus AT. “Weathering” HOPE VI: The Importance of Evaluating the Population Health Impact of Public Housing Demolition and Displacement. *J Urban Health*. 2011 Jun;88(3):417–35.
92. Jacquez GM, Meliker JR, AvRuskin GA, Goovaerts P, Kaufmann A, Wilson ML, et al. Case-control geographic clustering for residential histories accounting for risk factors and covariates. *Int J Health Geogr*. 2006 Aug 3;5(1):32.
93. Jacquez GM, Meliker J, Kaufmann A. In search of induction and latency periods: Space-time interaction accounting for residential mobility, risk factors and covariates. *Int J Health Geogr*. 2007 Aug 23;6(1):35.
94. Simonsen N, Scribner R, Su LJ, Williams D, Luckett B, Yang T, et al. Environmental Exposure to Emissions from Petrochemical Sites and Lung Cancer: The Lower Mississippi Interagency Cancer Study. *J Environ Public Health* [Internet]. 2010 [cited 2018 Oct 3];2010. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2838364/>
95. American Joint Committee on Cancer. Chapter 20 - Colon and Rectum. In: *AJCC Cancer Staging Manual*. 8th ed. New York, NY: Springer; 2017.

96. US Department of Commerce. National Technical Information Service [Internet]. [cited 2023 Sep 2]. Available from: <https://dmf.ntis.gov/>
97. Prener C, Fox B. censusxy: Access the U.S. Census Bureau's Geocoding A.P.I. System [Internet]. 2021 [cited 2021 Jul 7]. Available from: <https://CRAN.R-project.org/package=censusxy>
98. Surveillance Research Program, National Cancer Institute. ResHistGen Residential History Generation Programs. 2016.
99. McCollum AD, Catalano PJ, Haller DG, Mayer RJ, Macdonald JS, Benson AB, et al. Outcomes and toxicity in african-american and caucasian patients in a randomized adjuvant chemotherapy trial for colon cancer. *J Natl Cancer Inst*. 2002 Aug 7;94(15):1160–7.
100. Colorectal cancer mortality rate - Chicago Health Atlas [Internet]. [cited 2023 Sep 2]. Available from: <https://chicagohealthatlas.org/indicators/VRCRR?topic=colorectal-cancer-mortality-rate>
101. Garner K. Colorectal Cancer in Illinois: An Overview of Key Statistics. Springfield, IL: Illinois Department of Public Health; 2011 Jun. (Epidemiologic Report Series). Report No.: 11:04.
102. Dolecek TA, Shen T. Cancers of the Colon and Rectum: Evidence of Disparities between Blacks and Whites in Illinois. Springfield, IL: Illinois Department of Public Health; 2000 Dec. (Epidemiologic Report Series). Report No.: 00:8.
103. American Cancer Society. Colorectal Cancer Facts & Figures 2020-2022 [Internet]. Atlanta, GA: American Cancer Society; 2020 p. 48. Available from: <https://www.cancer.org/research/cancer-facts-statistics.html>
104. American Cancer Society. Cancer Facts & Figures 2023 [Internet]. Atlanta, GA: American Cancer Society; 2023. Available from: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2023/2023-cancer-facts-and-figures.pdf>
105. Lillard LA, Panis CWA. Panel Attrition from the Panel Study of Income Dynamics: Household Income, Marital Status, and Mortality. *J Hum Resour*. 1998;33(2):437–57.
106. PSID Help Desk. Re: Question about PSID mortality file. 2023.
107. University of Michigan Institute for Social Research. PSID News. [cited 2023 Sep 15]. Panel Study of Income Dynamics News - September 2023. Available from: <https://psidonline.isr.umich.edu/Guide/News.aspx>
108. Survey Research Center, Institute for Social Research. Panel Study of Income Dynamics: 1968-2017 Mortality File Documentation [Internet]. Ann Arbor, MI: The University of Michigan; 2019 Feb [cited 2019 Dec 10]. Report No.: Release 1. Available from: https://simba.isr.umich.edu/restricted/docs/Mortality/Mortality17_Introduction.pdf
109. University of Michigan Institute for Social Research. PSID Process & Requirements for Obtaining Restricted Data [Internet]. [cited 2023 Sep 15]. Available from: <https://simba.isr.umich.edu/restricted/ProcessReq.aspx>

110. Herget K, Stroup A, Smith K, Wen M, Sweeney C. Unstaged cancer: Long-term decline in incidence by site and by demographic and socioeconomic characteristics. *Cancer Causes Control*. 2017 Apr 1;28(4):341–9.
111. Merrill RM, Sloan A, Anderson AE, Ryker K. Unstaged cancer in the United States: a population-based study. *BMC Cancer*. 2011 Sep 21;11(1):402.
112. Kang S. Severe and persistent housing instability: examining low-income households' residential mobility trajectories in the United States. *Hous Stud*. 2021;1–26.
113. Hurley SE, Reynolds P, Goldberg DE, Hertz A, Anton-Culver H, Bernstein L, et al. Residential mobility in the California Teachers Study: implications for geographic differences in disease rates. *Soc Sci Med*. 2005 Apr 1;60(7):1547–55.
114. Jacquez G, Barlow J, Rommel R, Kaufmann A, Rienti M, AvRuskin G, et al. Residential Mobility and Breast Cancer in Marin County, California, USA. *Int J Environ Res Public Health*. 2013 Dec 23;11(1):271–95.
115. Jacquez GM, Shi C, Meliker JR. Local Bladder Cancer Clusters in Southeastern Michigan Accounting for Risk Factors, Covariates and Residential Mobility. *PLOS ONE*. 2015 Apr 9;10(4):e0124516.
116. Nordsborg RB, Meliker JR, Ersbøll AK, Jacquez GM, Raaschou-Nielsen O. Space-Time Clustering of Non-Hodgkin Lymphoma Using Residential Histories in a Danish Case-Control Study. *PLOS ONE*. 2013 Apr 1;8(4):e60800.
117. Sloan CD, Nordsborg RB, Jacquez GM, Raaschou-Nielsen O, Meliker JR. Space-Time Analysis of Testicular Cancer Clusters Using Residential Histories: A Case-Control Study in Denmark. *PLoS ONE* [Internet]. 2015 Mar 10 [cited 2018 Oct 3];10(3). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4355495/>

VITA

NAME:	Emma Elizabeth Boylan
EDUCATION:	M.S. Epidemiology, University of Illinois at Chicago, 2016 B.A. History, University of Illinois at Urbana-Champaign, 2009
RESEARCH AND WORK EXPERIENCE:	<p>Office of Chronic Disease Prevention and Health Promotion, Bureau of Community Health, Chicago Department of Public Health, 2022-2023</p> <p>School of Public Health, University of Illinois at Chicago: Policy, Practice, and Prevention Research Center, 2019-2022</p> <p>Division of Epidemiology and Biostatistics, University of Illinois at Chicago School of Public Health: Chicago Prostate and Colorectal Cancer Survival Study, 2015-2016</p> <p>Institute for Health Research and Policy, University of Illinois at Chicago: Illinois Prevention Research Center, 2015-2019</p> <p>Institute for Health Research and Policy, University of Illinois at Chicago: We Choose Health Community Transformation Grant Evaluation, 2014</p> <p>Department of Medical Imaging, Ann & Robert H. Lurie Children's Hospital, Chicago, IL: 2009-2014</p> <p>The Beckman Institute, University of Illinois at Urbana-Champaign: The Health Literacy Project, 2008-2009</p> <p>College of Education, University of Illinois at Urbana-Champaign: The Adult Learning Lab, 2007-2008</p>
AWARDS:	<p>Honorable Mention, Faculty-Student Collaboration, University of Illinois at Chicago School of Public Health Research Article Awards, 2021: for Freeman, et al.</p> <p>Caffey Award for Best Educational Exhibit, International Pediatric Radiology, 2011: for Kappil, et al.</p>
PUBLICATIONS:	<p>Freeman VL, Boylan EE, Tilahun NY, Basu S, Kwan M-P. Sources of selection and information biases when using commercial database-derived residential histories for cancer research. <i>Annals of Epidemiology</i>. 2020 Nov 1;51:35-40.e1.</p> <p>Freeman VL, Naylor KB, Boylan EE, Booth BJ, Pugach O, Barrett RE, et al. Spatial access to primary care providers and colorectal cancer-</p>

specific survival in Cook County, Illinois. *Cancer Med.* 2020 May;9(9):3211–23.

Freeman VL, Boylan EE, Pugach O, Mclafferty SL, Tossas-Milligan KY, Watson KS, et al. A geographic information system-based method for estimating cancer rates in non-census defined geographical areas. *Cancer Causes & Control.* 2017 Aug 20;1–10.

Rigsby CK, Popescu AR, Nelson P, Orr RJ, Boylan EE, Schoeneman S, et al. Safety of Blood Pool Contrast Agent Administration in Children and Young Adults. *AJR Am J Roentgenol.* 2015 Nov;205(5):1114–20.

Rigsby CK, Hilpipre N, McNeal GR, Zhang G, Boylan EE, Popescu AR, et al. Analysis of an automated background correction method for cardiovascular MR phase contrast imaging in children and young adults. *Pediatr Radiol.* 2014 Mar;44(3):265–73.

Popescu A, Thakrar D, Kim ST, Boylan EE, de Freitas RA, Rigsby C. Right Heart Lesions. In: Coley BD, editor. *Caffey's Pediatric Diagnostic Imaging, 2-Volume Set, 12e.* 12th ed. Saunders; 2013.

Deng J, Rigsby CK, Schoeneman S, Boylan E. A semiautomatic postprocessing of liver R2* measurement for assessment of liver iron overload. *Magn Reson Imaging.* 2012 Jul;30(6):799–806.

Boylan E, Wyers M, Jaffar R. A rare case of thymoma in a 15-month-old girl. *Pediatr Radiol.* 2011 Nov;41(11):1469–71.

PRESENTATIONS:

Rigsby C, Popescu A, Boylan E, Schoeneman S, Orr J, de Freitas RA. Safety of a blood pool contrast agent in children and young adults. Society for Pediatric Radiology; 2014; Washington, DC.

Popescu A, Rigsby C, Boylan E, de Freitas RA, Bi X. Assessment of two different MRA techniques: Standard dynamic gradient echo with extracellular contrast agent versus gradient echo with blood pool contrast agent. Society for Pediatric Radiology; 2014; Washington, DC.

Deng J, Burrowes D, Goldman S, Boylan E. Advanced diffusion-weighted MRI analysis for brain tumor tissue characterization in pediatric patients. American Society of Neuroradiology; 2014; Montreal, Quebec, Canada.

Popescu A, Rigsby CK, Boylan EE, Bi X, de Freitas RA. Comparison of three coronary artery imaging techniques in children. Radiological Society of North America; 2012; Chicago, IL.

Hilpipre N, Rigsby C, Boylan E, Zhang G, Greiser A, McNeal GR, et al. Analysis of a simple automated background correction method for cardiac MR phase contrast imaging. Society for Pediatric Radiology; 2012; San Francisco, CA.

Aw J, Boylan E, Schoeneman S, Alden T, Burrowes D. Applications for fast brain MR in children with intrachranial arachnoid cysts. Society for Pediatric Radiology; 2012; San Francisco, CA.

POSTERS:

Boylan E, Pugach O, Sharp L, Freeman VL. Impact of Driving and Public Transit Travel Burden on Colorectal Cancer Prognosis. Poster presented at: American Public Health Association; 2017 Nov 6; Atlanta, GA.

Tossas-Milligan KY, Boylan E, Pugach O, Watson KS, Freeman V, Winn RA. From Active Surveillance to Activism with Surveillance: Cancer Rates for Politically-Designated Areas. Poster presented at: UI Cancer Center Research Forum; 2016 Oct 18; Chicago, IL.

Deng J, Burrowes D, Boylan E. Multi-compartment diffusion analysis for differentiation of malignant and benign brain tumors in pediatric patients. International Society for Magnetic Resonance in Medicine; 2013; Salt Lake City, UT.

Kappil M, Saker M, Boylan E, Rigsby C. Button battery ingestion in children: What the radiologist should know. International Pediatric Radiology; 2011; London.

Aw J, Maddalozzo J, Schoeneman S, Boylan E, Kanbur V, Burrowes D. Imaging of the Posterior Hyoid Space. American Society of Head and Neck Radiology; 2011; San Diego, CA.
