ABSTRACT

Title of Dissertation: THE LONG TAIL OF HISTORY: COMBINING THE 1940 CENSUS, REDLINING MAPS, AND HRS: METHODS FOR ANALYZING THE IMPACT OF REDLINING ON HEALTH, ECONOMIC, AND HEALTHCARE OUTCOMES IN OLDER ADULTS TODAY

Shuo Huang, Doctor of Philosophy, 2023

Dissertation directed by: Neil Jay Sehgal, PhD, MPH

Department of Health Policy Management

Background

As part of the New Deal in the 1930s, the Federal government used the Home Owners' Loan Corporation (HOLC) to draw real estate security maps that were color coded or redlined to discourage lending in majority Black neighborhoods. Redlined areas still have worse health and economic outcomes in the present day. Current literature is focused on present-day residents of redlined areas. Tracking exposures to redlining and conditions of redlining close to the 1930s with present-day health is an unexplored area.

Methods

We utilize geo-referenced 1930s HOLC maps to locate individuals and map demographic considerations. We use novel algorithmic solutions to geolocate unknown 1940 enumeration districts. Using a 1940 census-linked sample of the Health and Retirement Study to locate individuals in HOLC areas at the time, we conduct survival analysis on HOLC categories' effect on age at death as well as other analysis on health, economic, and healthcare utilization in the near present. We test for a potential mediator.

Results

Population density is not associated with either HOLC category or present day life expectancy, and is unlikely to be a mediator. In uncontrolled models, for HRS individuals in the 1940 census HOLC category is associated with greater hazards, worse odds of self-rated health, and worse economic outcomes. With controls, HOLC category is only associated with worse odds of self-rated health. HOLC category is not associated with health insurance or healthcare utilization in this sample.

Conclusion

Redlining is associated with health and economic outcomes which are attenuated when controlling for likely pathways between redlining and health. Future research should focus on whether individuals stay in redlined areas, and on identifying policy and initial state matrix that can describe what redlining may be a proxy for.

THE LONG TAIL OF HISTORY: COMBINING THE 1940 CENSUS, REDLINING MAPS, AND HRS: METHODS FOR ANALYZING THE IMPACT OF REDLINING ON HEALTH, ECONOMIC, AND HEALTHCARE OUTCOMES IN OLDER ADULTS TODAY

by

Shuo Huang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2023

Advisory Committee:

Neil Sehgal, PhD, MPH, Chair

Kellee White, PhD, MPH

Michel Boudreaux, PhD, MS

Dahai Yue, PhD, MS, MBBS

Dylan Roby, PhD, MPP

Typhanye Vielka Dyer, PhD, MPH

© Copyright by

Shuo Huang

2023

Dedication

For Dad.

For Mom.

For Andrea.

Acknowledgements

A dissertation is never the work of just one person. I could not have done this without the support of an entire community. I cannot possibly make a comprehensive list of everyone who deserves thanks, but I'd like to highlight a few individuals.

I'd like to thank Dr. Nina Harris for mentorship, critical feedback, and providing a rare space to discuss the interplay between identity and public policy.

My committee has helped shape my research and my academic trajectory. Dr. Kellee White introduced me to the HRS dataset, introduced me to collaborative research, and has mentored me over the years in the process of formulating health equity research questions and methods. Dr. Michel Boudreaux always provides sharp and critical feedback, trusted me with teaching students, and has encouraged my research in unorthodox health policy directions. Dr. Dahai Yue introduced me to 1940 census work, provided a sounding board to discuss restricted data issues, and always has sharp critiques for methods. Dr. Dylan Roby taught me to think about how policies and politics interact, to think about how academics can wield power in policy processes, and mentored me in teaching. Dr. Typhanye Vielka Dyer has provided guidance on the dissertation process, and I am greatly appreciative of the time spent as the Dean's Representative.

Dr. Jie Chen and Dr. Andrew Fenelon provided early feedback on methods and directions in this research. Dr. Kevin Roy and Dr. Eva DuGoff provided valuable mentorship for the dissertation process. Dr. Luisa Franzini has provided much support and mentorship over the years. Dr. Melvin Seale has provided academic, professional, and personal mentorship. Bianca Garcia has made the process of using departmental resources possible. Dr. Steve T. McGarvey and Dr. Omar Galarraga helped me publish my first paper and set me on the path towards health equity and health policy.

My fellow graduate students deserve a lot of credit, and I have learned a lot from working with them. In particular, I'd like to thank Jun, Ivy, Sandra, Sofi, Portia, Elle, and Dan for being great friends and colleagues, commiserating through courses, and providing multiple rounds of feedback throughout my time here. Thank you as well to Elleese, Deanna, Aita, Ebony, Shamis, and Liyang for your mentorship, feedback, and friendship. Kaitlynn, Michelle, Makeda, Heather, and Alice all provided much needed morale boosts at the end.

I would like to thank the staff at HRS, BNIA, the Mapping Inequality Project, and OpenCage who have made this valuable data available, and who have provided advice and technical assistance throughout.

This work would not have come into being without the insights of communities and individuals that have and continue to resist structural racism while imagining better worlds. While the literature and discourse from these communities might only be reflected in part in this dissertation's theoretical framework, they are very much reflected in the dissertation's topic of focus, framing, and in my use of multiple, layered, interdisciplinary data sources. In particular Tracye, NeeNee, Raven, and Marsha, thank you for your friendships and mentorships, for introducing me to new framings, and for fighting for a better world.

Dissertations are not possible without the support of friends and family. I'd like to thank Keithlee and Rebecca, who provided feedback, moral support, and morale support. Other friends who have provided morale (and material) support at varying times through the years include Anna, Justin, Merrick, Will, Dana, Keren, Charles, Ataa, El, and Ria, who all deserve thanks for their parts.

It's been a great pleasure working with Dr. Neil Sehgal as my chair. You've become much more than a mentor, and I'm proud to call you a friend. I've learned so much not just about how to be a scholar, but also how to mentor others, and how to chart a moral course from you.

My father Yulin and my mother Keling's support stretches back much further than anyone else's. It's one of my great regrets that I didn't finish in time for my father to see the finished product, but I know he would have been proud of me.

My partner, in crime and in life, Andrea is probably the one person most responsible for the completion of this dissertation. Thank you for introducing me to the digitized redlining maps, for keeping me grounded, for being patient with me, for letting me bounce ideas, for providing critical feedback, and for growing together as scholars and people.

Last but not least, I'd like to thank my cats, Luna, Barney, and LB, who remind me to eat, rest, and sleep.

Table of Contents

Abstract..........
Copyright..........
Dedication.......... ii
Acknowledgements.......... iii
Table of Contents.......... v
List of Tables.......... ix
List of Figures.......... x
List of Abbreviations.......... xi
1. Introduction.......... 1
1.1 Objective.......... 2
1.2 Structure of the manuscript.......... 2
1.3 Overview.......... 3
1.3.1 Background on historic redlining and present day health.......... 3
1.3.2 Measuring Health.......... 4
1.3.3 Methodological concerns of survivorship bias in samples of an elderly population.......... 5
1.3.4 Housing as an aspect of wealth.......... 5
1.3.5 Relationship between economic outcomes and health.......... 6
1.3.6 Aday-Andersen Model.......... 6
1.3.7 Does present day density explain redlining's association with health?.......... 7
1.3.8 Population density.......... 7
2. Chapter 1: Population Density as a Mediator.......... 9
2.1 Hypothesis:.......... 9
2.2 Methods.......... 9

2.2.1 Creating Measures.......... 9

2.2.1.1 Main Independent Variable.......... 9

2.2.1.2 Main Dependent Variable and Controls.......... 10

2.2.1.3 Potential Mediator Variable.......... 10

2.2.2 Analysis Methods..........11

2.2.2.1 Base Model.......... 11

2.2.2.2 Mediation Analysis.......... 11

2.3 Results.......... 11

2.3.1 Base Analysis.......... 11

2.3.2 Mediation Analysis..........13

2.4 Discussion..........14

3. Chapter 2: Health Outcomes.......... 16

3.1 Hypothesis.......... 16

3.2 Methods.......... 16

3.2.1 Measure creation..........17

3.2.1.1 Assigning HOLC redlining category to enumeration districts..........17

3.2.1.2 Construction of “virtual” enumeration districts..........18

3.2.1.3 Assigning HOLC Categories.......... 21

3.2.1.4 Creating Main Independent Variable..........22

3.2.1.5 Health and Retirement Study..........22

3.2.1.6 Creating the Main Dependent Variables.......... 22

3.2.1.7 Covariates and Controls.......... 23

3.2.2 Analysis Methods.......... 24

3.3 Results.......... 25

3.3.1 Tabulations.......... 25

3.3.2 Survival analysis....28

3.3.3. Self-rated health.... 30

3.4 Discussion....31

4. Chapter 3: Economic Outcomes and Healthcare Utilization.... 33

4.1 Hypothesis:.... 33

4.2 Methodology:.... 33

4.2.1 Data Sources....33

4.2.2 Creating Measures.... 34

4.2.2.1 Independent Variable - HOLC Category.... 34

4.2.2.2 Control Variables....34

4.2.2.3 Main Dependent Variables - Economic Outcomes....34

4.2.2.4 Main Dependent Variables - Healthcare Utilization....35

4.2.3 Analysis Methods.... 36

4.2.3.1 HOLC categories' association with race.... 36

4.2.3.2 Economic Outcomes....36

4.2.3.3 Healthcare Utilization Outcomes....36

4.3 Results....37

4.3.1 Tabulations.... 37

4.3.2 HOLC Category and Race....37

4.3.3 Economic Outcomes.... 38

4.3.4 Healthcare Utilization Outcomes.... 39

4.4 Discussion....40

5. Final Discussion.... 43

5.1 Limitations....43

5.2 Future Directions....44

5.3 Policy Implications 46
Appendix A: Figures 48
Appendix B: Previous paper 57
References 82

List of Tables

Table 1: Population density as a potential mediator for HOLC category's association with area life expectancy.. 12

Table 2: Real Enumeration District (ED) Overlap with Virtual Enumeration Districts. 25

Table 3: Summary Statistics.. 25

Table 4: Self-rated health by HOLC..28

Table 5: Survival Analysis..28

Table 6: Self rated health at entry..30

Table 7: Race by HOLC Composition...37

Table 8: Total wealth 5 categories.. 38

Table 9: Insured at study entry.. 39

Table 10: Likelihood of seeing a doctor in last 12 months at study entry................... 40

List of Figures

Figure 1: Flowchart of virtual enumeration district construction and assigning HOLC categories....... 18

Figure 2: Chicago real and virtual enumeration districts tagged by HOLC. HOLC backdrop....... 48

Figure 3: Street grid of a section of present day Philadelphia, PA, USA....... 49

Figure 4: Philadelphia geolocated 1940 addresses by enumeration district (untrimmed)....... 50

Figure 5: Philadelphia constructed virtual enumeration districts without trimming addresses....... 51

Figure 6: Philadelphia trimmed address virtual enumeration districts overlaid with untrimmed addresses....... 52

Figure 7: Philadelphia trimmed address virtual enumeration districts overlaid with real enumeration districts....... 53

Figure 8: Brooklyn geolocated 1940 addresses by enumeration district....... 54

Figure 9: Brooklyn 1930 HOLC map....... 55

Figure 10: Brooklyn virtual enumeration districts tagged with HOLC....... 56

List of Abbreviations

API: Application Programming Interface

BNIA: Baltimore Neighborhood Indicators Alliance

CSA: Community Statistical Area

ED: Enumeration District

GED: General Education Development

GRF: Geographic Reference File

HOLC: Home Owners’ Loan Corporation

HRS: Health and Retirement Study

HS: High School

IPUMS: Integrated Public Use Microdata Series

NHGIS: National Historical Geographic Information System

OLS: Ordinary Least Squares

Q1-Q4: Quartile 1 to Quartile 4

WGS84: World Geodetic System 1984

1. Introduction

As part of the New Deal in the 1930s, the Federal government used the Home Owners' Loan Corporation (HOLC) to draw real estate investment security maps (Nelson et al., 2017). Areas were color-coded or “redlined” and drawn to reflect predominant neighborhood race, discouraging lending in majority Black neighborhoods (Rothstein, 2017). Neighborhoods “redlined” (Nelson et al., 2017; Rothstein, 2017) in the 1930s are still associated with worse financial (Aaronson et al., 2017; Lynch et al., 2021) and health outcomes (Lynch et al., 2021; Krieger, Wright, et al., 2020; Krieger, Van Wye, et al., 2020; Nardone et al., 2020; Wilson et al., 2018; Alexander and Currie, 2017; Trangenstein et al., 2020; McClure et al., 2019; Richardson et al., 2020; Huang and Sehgal, 2022) today. Redlining was likely a proxy for a matrix of discriminatory antecedent conditions, policymaker attitudes, and decades-long policy vectors (Michney, 2022; Rothstein, 2017), though there is some debate about its potentially causal nature as well (Aaronson et al., 2017). This century long path dependence (Michaels and Rauch, 2016) masks the potentially differing roles of social policies throughout the 20th and 21st centuries acting on marginalized, primarily Black communities and individuals, including the impact of lifetime discrimination (White et al., 2012) and of displacement and gentrification (Stancil, 2019), and may impact healthcare utilization (Aday and Andersen, 1974; Williams and Jackson, 2005). Previous area estimates of health effects associated with redlining (Lynch et al., 2021; Krieger, Wright, et al., 2020; Krieger, Van Wye, et al., 2020; Nardone et al., 2020; Wilson et al., 2018; Alexander and Currie, 2017; Trangenstein et al., 2020; McClure et al., 2019; Richardson et al., 2020; Huang and Sehgal, 2022) may poorly capture the full impact of initial conditions from redlining on individuals (Huang and Sehgal, 2022). Clarifying the policy mechanisms that redlining may have acted through, and connecting health, economic, and utilization outcomes specifically to redlining status close to the 1930s, could allow us to examine the role of initial conditions and present day health.

Health outcomes are theorized to be due to ultimate social causes (Williams and Jackson, 2005) as well as access to and use of healthcare resources (Aday and Andersen, 1974), which are also sources of health disparities (Williams and Jackson, 2005). Redlining is associated with disparities in such social factors including continued worsened area estimates of financial outcomes today (Aaronson et al., 2017;

Lynch et al., 2021), and associations between worsened economic well-being with health outcomes have been demonstrated (McClure et al., 2019). However, the existing literature relies on estimates of economic well-being for current residents, and cannot capture how redlining exposure may have impacted economic well-being and healthcare utilization over the life course. Economic outcomes and healthcare utilization may be important factors at the individual level for health, including for older adults who have higher healthcare needs than younger cohorts.

1.1 Objective

To use historical redlining, community data, and linked 1940 census and Health and Retirement Survey (HRS) data to analyze the impact of redlining on outcomes in older adults today. We aim to assess 1) whether redlining is associated with long-tail health policy outcomes at the community level, and whether early exposure to redlining is associated with 2) health outcomes in older adults and 3) economic well-being and healthcare utilization in older adults.

1.2 Structure of the manuscript

We start with an overview of the literature to identify important areas of study, conceptual frameworks, and gaps. We adapt a conceptual framework on potential individual and neighborhood pathways from redlining to health (Swope et al., 2022) to the case of initial redlining conditions. We start by theorizing how we might measure the present day health of individuals residing in redlined areas in the 1940 census. We move on to theorizing about potential pathways to those present day health states of those individuals. And, finally, we consider a potential mediating factor at the neighborhood level that reflects alternative possibilities for what HOLC planners were concerned with in the 1930s.

In chapter one, we investigate the potential mediating factor as we respond to community stakeholder feedback to our paper (Huang and Sehgal, 2022). We include that paper as appendix B. Population density has been associated with certain negative health outcomes. Some community stakeholders suggested that the HOLC in 1930 was far more concerned with mapping out future potential density in Baltimore, rather than race (Personal Correspondence, 2022). As a result, Baltimore’s redlining map may reflect neighborhoods that densified over the 20th and early 21st centuries, rather than as a marker of systemic racism, and suggesting a mediating pathway from redlining to density to health. This chapter assesses whether population density might serve as such a mediator.

In chapters two and three, we begin to operationalize and test our theorized pathways between individuals living in 1930s redlined areas and their near-present health. We do so by utilizing the felicitous availability of several datasets. The Health and Retirement Study has geographic variables at the enumeration district level from the full count 1940 census (the first census taken immediately after redlining), of which 191 of the largest cities have tractable information that allows us to reconstruct enumeration districts (and a publicly available partially overlapping set of 40 cities which have enumeration districts publicly mapped out). This allows us to track a group of people from their exposure close to redlining to their outcomes as elderly adults some fifty to eighty years later. In chapter two, we examine how early exposure to redlining is associated with health outcomes, and how propensity to mobility may be associated. In chapter three, we begin to examine some ways that exposures to redlining may affect economic and health utilization outcomes. We finish with a final discussion of all our findings.

1.3 Overview

1.3.1 Background on historic redlining and present day health

Since the public release of georeferenced and digitized HOLC maps by the Mapping Inequality Project (Nelson et al., 2017), substantial scholarship has examined associations between 1930s redlining and present day health (Swope et al., 2022). Historic redlining has been associated both in specific cities and nationwide with present day social determinants of health, environmental outcomes, maternal and child outcomes, COVID-19 outcomes, and mortality and life expectancy (Richardson et al., 2020; Wilson et al., 2018; Alexander and Currie, 2017; Nardone et al., 2020; Lynch et al., 2021; Huggins, 2017; Linde et al., 2023; Tragenstein et al., 2020; Krieger, Wright, et al., 2020; Krieger, Van Wye, et al., 2020). While some of these studies have tracked individual outcomes (Krieger, Wright, et al., 2020; Krieger, Van Wye, et al., 2020), most were at an area level.

Given the diversity of methods, locales, disciplines, and data sources involved in this research, there have been some preliminary efforts to provide guiding frames to this work. Swope et al. (2022) provide a preliminary conceptual model based off of Krieger et al.'s work linking redlining (as a proxy for or as a parcel of a racially discriminatory landscape and policies) to present day health through effects on both individuals and intergenerational communities and neighborhoods. The generated pathways include disparate treatment and effects by race through effects on homeownership and wealth, through proximal factors and to immediate individual factors that impact an individual's health.

This conceptual model easily accommodates some of the already known and theorized mechanisms between redlining and present day health through residential segregation (White et al., 2012), political disenfranchisement and state suppression (Pappoe, 2016; McCoy, 2017; Naman et al., 2020; Tragenstein et al., 2020) that create environmental hazards (Bailey et al., 2017; Jacoby et al., 2018) as well as potential economic influences on health (Williams, 1999).

However, the proposed model does not yet differentiate between those who now live in redlined areas versus those who at one point lived in redlined areas (who may or may not continue to live in redlined areas), which may be one way we can begin to identify and classify neighborhood versus individual effects. To our knowledge, we did not find any other study linking exposure to redlining in the 1940 census to life expectancy or health outcomes in the present at the individual level.

1.3.2 Measuring Health

To adapt Swope et al.'s model (2022) to our case of initial exposures to redlining, we start at the most immediate association: that of health. Health as a construct has been defined in many ways (Callahan, 1973; Huber et al., 2011). In addition to how the construct is defined, measures of health are also very diverse (Bowling, 2005). For the purpose of healthcare decision-making and studies of health at the level of national policy-making (WHO, 2009; Smith et al., 2009), health utility–combining length of life with quality of life–may be the most commonly recognized health measures.

There are strong limitations of measuring health in this way. These measures may not suitably account for a context of structural barriers that socially construct disability–a particular critique of expert consensus measures like DALYs (Hurst, 2003)–and also that contribute to discrimination against disabled people–which time preference models like most QALYs calculations are likely conditioned upon (Sinclair, 2018). However, the basic premise of measuring both length of life and a more general definition of quality of life bounds common definitions of health in a measurable way.

1.3.3 Methodological concerns of survivorship bias in samples of an elderly population

Survival data that contain a lag between exposure and the start of the measurement period where only survivors can be identified–known as left truncated data–can suffer from survivorship bias, where subjects exposed to hazards appear to be more resilient to the hazard simply because they have not yet died from it (Cain et al., 2011). Specifically–in studying health disparities–datasets describing elderly people as the population of interest can find disparities attenuated due to strong survivorship bias in the population with less equitable health (Mayeda et al., 2018). This can lead to conclusions that appear to suggest that a population is more resilient to a hazard than a comparison population when in fact the opposite is true (Mayeda et al., 2018).

1.3.4 Housing as an aspect of wealth

Moving further to proximate potential mechanisms at the individual and neighborhood level between redlining and health, we theorize on how redlining might affect economic outcomes and healthcare utilization. Redlining likely serves as a proxy indicator for an existing state of housing segregation and for a range of continuing discriminatory practices later in the 20th century that reinforced this state (Rothstein, 2017; Hillier, 2003; Michney, 2022). Housing was (and continues to be) a primary source of wealth for many households throughout the 20th century (Aaronson et al., 2017; Appel and Nickerson, 2016). Since redlining is meant to demarcate areas of greater mortgage risk, redlining thus also demarcates reduced real estate values. As demonstrated by Aaronson et al. (2017), reductions in housing value (and concomitant displacement that may wipe out potential wealth gains) may weigh down the financial resources available to those who lived in redlined areas.

1.3.5 Relationship between economic outcomes and health

Income and economic resources are social determinants of health (Link and Phelan, 1995; Williams and Jackson, 2005; Williams, 1999). Those who have more economic resources are more able to acquire technologies and resources that enhance health (Link and Phelan, 1995); secure other social determinants of health such as safe shelter, safe food, and safe water; and avoid environmental hazards to health. Additionally, economic resources increase relative political power (and vice-versa) (Ferguson, 2013), which helps individuals secure allocation of health-enhancing resources and health hazards in a favorable manner (Link and Phelan, 1995).

One potential pathway for individuals to secure health-enhancing resources via economic resources is through economic resources' role in the ability for individuals and households to access and make decisions for medical care and other forms of healthcare. This includes both increased access to healthcare financing via health insurance (Nyman, 2004) as well as increased social access and direct access to medical care (Link and Phelan, 1995).

1.3.6 Aday-Andersen Model

We can utilize a portion of the Aday-Andersen Model (1974) to describe this theorized pathway from redlining to healthcare utilization via economic outcomes. The Aday-Andersen Model posits a series of predisposing and enabling factors that are antecedent to an individual's actual use of healthcare services. Enabling factors are those factors which create functional access to healthcare services should an individual choose to seek healthcare (the "able" part of "willing and able"). Income and economic resources are one set of those enabling factors (Aday and Andersen, 1974). For our purposes, we theorize that the policies that redlining is a proxy for reduce the wealth available to individuals and their households, which thus add a constraint to enabling factors for healthcare utilization. As a result, healthcare utilization would be reduced for those who lived in redlined areas in 1940 as compared to those who lived in green A areas. As a corollary, we would find reduced intermediary enabling factors such as health insurance in this pathway.

The prior literature is limited when it comes to redlining's association with health insurance and healthcare utilization. Nardone et al. (2020) and White et al. (2021) found that redlining in several large US cities is associated with increased chances of being uninsured at the neighborhood level. Poulson et al. (2021) found lower odds in the present day for undergoing lung cancer screenings in Black patients in redlined areas. We found no studies that examined healthcare utilization for residents of redlined neighborhoods from close to the time the redlining maps were drawn.

1.3.7 Does present day density explain redlining's association with health?

We move on to consider possible mediating mechanisms of action as part of our ongoing investigation into whether community and neighborhood effects may have long-tailed policy impacts by examining whether 1930s redlining is associated with a lingering disparate impact on health outcomes at the community statistical area (CSA) level in Baltimore today. Specifically, we investigate whether 1930s redlining operates largely through population density today as measured by crowding that would primarily explain our findings that redlining is associated with life expectancy in Baltimore CSAs (Huang and Sehgal, 2022).

On direct feedback from Baltimore City community members (Personal Correspondence, 2022) about the Association of Redlining and Health in Baltimore paper (Huang and Sehgal, 2022) who were interested in this question–particularly as a race-neutral explanation for our previous findings–we include population density by neighborhood into our model as a potential mediator.

1.3.8 Population density

Population density's potential role in etiology is two-fold. For pathogenic disease transmission, increasing density increases potential exposures (Sy et al., 2021). For non-communicable diseases, density has been theorized to decrease each individual's access to green space, could increase their exposure to environmental pollution and hazards, could degrade the quality of available housing, or could disrupt networks of social solidarity (Beenackers et al., 2018). Historically, population density was a cause of concern for both planners and public health advocates (Factor and Waldron, 1973; Cassel, 2009). On the other hand, increasing population density could improve access to amenities and physical activity

(Beenackers et al., 2018), healthcare facilities (Kjellstrom, 2007), and increase the pool of supportive resources in socially cohesive communities (Cassel, 2009). Alternatively, human densities may not approach levels where health concerns are of concern yet (Factor and Waldron, 1973).

Earlier studies also posited a relationship between population density and disease outcomes. Some studies have found relationships between population density and cause specific mortality (Sy et al., 2021) or with all-cause mortality in specific cities (Beenackers et al., 2018). Previous studies of population density as a causal factor in health have not generally found associations between density and health when factors like housing discrimination, wealth and class are controlled for in models for non-communicable disease (Cassel, 2009; Factor and Waldron, 1973; Freedman et al., 1975). Population density is associated with communicable disease transmission during the COVID-19 pandemic (Sy et al., 2021).

Some recent prior literature has attempted to reduce focus on race as a core aspect of redlining in favor of more “race neutral” explanations of HOLC map patterns (Michney, 2022). Conversely, Krimmel (2018) found redlining to be associated with a reduction in population density, as disinvestment and white flight emptied out city cores (Krimmel, 2018). Additionally, the historiography of redlining is clear when it comes to what factors redlining cartographers were concerned with when they assessed mortgage default risk. There is only one “green-lined” neighborhood in all of the US that was primarily Black (Michney, 2022). The language accompanying each marked area in HOLC maps is starkly racist (Nelson et al., 2017). Even if HOLC maps were not used to determine HOLC lending patterns, the pervasiveness of institutional anti-Black racial attitudes inform the drafting of the maps as well as both contemporary (through the parallel Federal Housing Administration) and later policy-setting around race and housing risk throughout the 20th century (Michney, 2022; Hillier, 2003; Krimmel, 2018; Rothstein, 2017).

To our knowledge no previous scholarship attempts to link redlining to health with population density as the mediator.

2. Chapter 1: Population Density as a Mediator

We begin our analysis with this last question, since if redlining is mediated through urban density, then our theorized pathways for how 1930s redlining might be associated with present-day health would need to focus on population density as a main determinant. We lay out a hypothesis, propose and run methods to test the hypothesis, report results, and return to the prior literature in the chapter discussion section.

2.1 Hypothesis:

Our hypotheses are as follows:

H0: Population density does not mediate the observed relationship between 1930s redlining and current Baltimore City life expectancy at the CSA level.

H1: Population density mediates the observed relationship between 1930s redlining and current Baltimore City life expectancy at the CSA level.

2.2 Methods

2.2.1 Creating Measures

2.2.1.1 Main Independent Variable

As per our 2022 paper (included as appendix B), we take HOLC redlining polygons (digitized shapes that correspond to drawn boundaries) that were digitized and georeferenced (meaning located and drawn digitally in a consistent way for example on a grid) by the Mapping Inequality Project (Nelson et al., 2017) and re-project them (meaning we convert the map to account for the curvature of the Earth) using WGS84 Pseudo-Mercator projection. We pull Baltimore CSA neighborhood polygons from the Baltimore Neighborhood Indicators Alliance (BNIA) dataset. Baltimore CSAs are “neighborhoods” made up of collections of census tracts that roughly correspond to commonly demarcated neighborhoods used by the city government to aggregate statistics in a way that is more legible to residents. We intersect Baltimore

CSA neighborhoods with HOLC redlining maps via a spatial join in QGIS and assign HOLC categories to CSA neighborhoods by predominance. Predominance is a common method in multiple disciplines to assign features and values between two geographic data sources that lack contiguous polygon boundaries where–for a given polygon of a cartographic feature in one data source–the value of the largest overlapping polygon from the intersecting data source is assigned (Lindgren, 2013). Each CSA's assigned HOLC category is our main independent variable.

We collapse blue and green neighborhoods together due to sample size. The sociological distance between A and B categories is likely to be smaller than the distance between B and C (Nelson et al., 2017, Rothstein, 2017). In our 2022 paper, when performing sensitivity analysis using just A or just B as the reference category, we found no meaningful differences from our base model (Huang and Sehgal, 2022).

2.2.1.2 Main Dependent Variable and Controls

From the BNIA dataset, we draw current CSA health outcomes using 2013 CSA life expectancy as measured in years. Our controls are 2013 median CSA household income (in 1000's of US dollars) and 2013 percentage of African American residents in a CSA. 2013 data was chosen because CSA boundaries reported in that year were consistent across multiple BNIA datasets that supplied health measures.

2.2.1.3 Potential Mediator Variable

For each CSA, we draw the total 2013 population and 2013 number of residential properties (which can include properties with multiple units), and divide population into number of residences to calculate crowding as our measure of population density. We implement this method to measure population density at the suggestion of a community stakeholder (Personal Correspondence, 2022) and reflects 20th century concerns about population density and crowding (Factor and Waldron, 1973; Cassel, 2009). We perform mediation analysis with population density as the proposed mediator between redlining category and health outcomes.

2.2.2 Analysis Methods

2.2.2.1 Base Model

Before we can investigate population density's potential role as a mediator, we need to establish that our independent variable HOLC category is associated with our main dependent variable of Baltimore City CSA life expectancy. We run an ordinary least squares regression (OLS) of yellow C and red D categories on CSA life expectancy. We then run a model with median household income and percentage African American as controls.

2.2.2.2 Mediation Analysis

We run a three part mediation analysis using the Baron and Kenny method (1986) with population density as the potential mediator between the main independent variable of HOLC category and the main dependent variable of life expectancy. In the first part, we run a multivariable OLS to test for HOLC category's association with population density with controls. In the second part, we run a multivariable OLS to test the association of population density with life expectancy with controls. In the final part, we run a multivariable OLS model to test for HOLC category's association with life expectancy while controlling for population density and controls. All analyses are tested at the 95% confidence level. We must reject the null hypothesis that population density is not a mediator and accept the hypothesis that it is a mediator in the association between HOLC category and life expectancy if three conditions are met. 1) HOLC category is significantly associated with population density, 2) population density is significantly associated with life expectancy, and 3) the inclusion of population density in the final model appreciably attenuates the coefficient of HOLC category's association with life expectancy.

2.3 Results

2.3.1 Base Analysis

We classified 54 Baltimore CSAs with 1930s HOLC categories. There were 4 green A CSAs, 17 blue B, 19 yellow C, and 14 red D. Mean CSA population density is 3.52 people per residence with a standard deviation of 2.06 people. Minimum population density is 1.90 people and maximum is 12.74 people per residence.

We report results for base and mediation analyses in table 1. In our base analysis, HOLC category is statistically significantly associated at the 95% confidence level with a reduction in CSA level life expectancy in Baltimore City in 2013. CSA neighborhoods categorized as yellow C are associated with -5.36 fewer years of expected life (95% CI: -7.69 to -3.02) as compared to green A or blue B neighborhoods. CSA neighborhoods that were redlined (D) are associated with -4.01 fewer years of expected life (95% CI: -6.55 to -1.47) compared to green A or blue B neighborhoods. When controlling for median household income and percentage African American, the results are very similar and also statistically significant. Yellow C CSAs are associated with -4.93 fewer years of expected life (95% CI: -6.63 to -3.32) compared to A/B, and red D CSAs with -5.23 fewer years of expected life (95% CI: 6.98 to -3.49) compared to A/B. Both CSA median household income and percentage African American are statistically significant controls.

Table 1:
Population density as a potential mediator for HOLC category's association with area life expectancy

	Model 1: **Unadjusted**. Life expectancy - years	Model 2: **Controls**. Life expectancy - years	Mediation 1: **HOLC on population density**. People per residence	Mediation 2: **Population density on life expectancy**. Years	Mediation 3: **Full mediation model**. Life expectancy - years
HOLC n=54					
Blue/Green	ref	ref	ref		ref
Yellow	**-5.36 [-7.69, -3.02]**	**-4.93 [-6.63, -3.22]**	-.813 [-2.39, 0.76]]		**-5.12 [-6.81, -3.42]**
Red	**-4.01 [-6.55, -1.47]**	**-5.23 [-6.98, -3.49]**	-.937 [-2.55, 0.68]		**-5.46 [-7.20, -3.71]**
Controls					
Median HH Income - $1000		**0.06 [0.01, 0.11]**	-0.05 [-0.09, 0.00]]	**0.14 [0.09, 0.20]**]	0.05 [-0.01, 0.10]

African American - %	**-0.06 [-0.09, -0.03]**	-0.02 [-0.05, 0.01]	-0.01 [-0.05, 0.02]]	**-0.07 [-0.10, -0.04**]
Mediator				
Population density- people per residence			-0.05 [-0.47, 0.37]]	-0.24 [-0.54, 0.07]

2.3.2 Mediation Analysis

HOLC category is not statistically significantly associated with CSA population density at the 95% confidence level. Yellow C CSAs have -0.813 (95% CI: -2.39 to 0.76) fewer people per residence compared to A/B CSAs. Red D CSAs had -0.937 (95% CI: -2.55 to 0.68) fewer people per residence compared to A/B CSAs. Our controls of median household income and percent African American residents are not significant in the model.

CSA population density is not significantly associated with CSA life expectancy in Baltimore. For every additional person per residence, CSA life expectancy is associated with a reduction of -0.05 years of expected life (95% CI: -0.47 to 0.37). Our control of median household income is significant in the model and the control of percent African American residents is not.

When included in the full model as a potential mediator in the association between HOLC category and life expectancy, CSA population density is not statistically significant, and did not appreciably change the coefficients for either HOLC category or any of the other controls. Both C and D categorizations are still significantly associated with reduced CSA life expectancy. Compared to A/B CSAs, yellow C CSAs were associated with a -5.12 year reduction in life expectancy (95% CI: -6.81 to -3.42) and red D CSAs were associated with a -5.46 year reduction in life expectancy (95% CI: -7.20 to -3.71).

2.4 Discussion

We cannot reject the null hypothesis that population density is not a mediator in the association between 1930s HOLC category and 2013 CSA life expectancy in Baltimore City. Population density is not associated with HOLC category, is not associated with life expectancy, and does not change the magnitude of association that HOLC category has with life expectancy. The mean effect size between population density and life expectancy is effectively none.

While not statistically significant, our coefficients for yellowlined C and redlined D HOLC categories in the first part of our mediation analysis suggest that, if anything, C and D areas trend towards *lower* population density than A and B areas, not higher. Were the association significant, this would have been consistent with what Krimmel (2018) found. However, our analysis is not designed to investigate the association between redlining and population density.

Our null finding for population density as a mediating factor is consistent with the body of literature from other fields (Michney, 2022; Rothstein, 2017; Nelson et al., 2017) that show that at its core, HOLC's drafting of redlining maps was not a "race neutral" incidental exercise of structural racism. The policies that redlining is a proxy for were primed to harm Black communities and individuals because they were designed to harm Black communities, by policymakers who understood and supported the racial power hierarchy (Michney, 2022). Our finding thus also consistent with the implicit framing undergirding almost all other studies today of redlining and health: that racially discriminatory policies and/or initial conditions set the stage for worsened health disparities later, and our role as researchers is to investigate the temporal and spatial mechanisms of that path dependence in order to intervene today.

Given the preponderance of historical and documentary evidence (Michney, 2022; Nelson et al., 2017) in the historiography of redlining, neither a null nor a significant finding for population density as a mediating factor would lend support to or detract from the heterodox literature suggesting a "race neutral" mechanism. Structural racism operates on health through policies that can be facially "neutral" (Jones, 2000). Even today, city planning is rife with facially "neutral" policies designed to shape certain demographic outcomes, and “anti-density” is even used as a racist dog whistle by some corners of the planning world.

Taken in total, this does not suggest a direction for future research on population density as a mediator between redlining and life expectancy. Both from its lack of effect and from the prior literature in multiple fields, population density certainly is not a replacement construct for racial discrimination in redlining maps in Baltimore City. Of course, since this study is limited in its generalizability to Baltimore City, it is possible that population density does play a mediating role with life expectancy and health outcomes in studies that have looked at other cities, or nationwide.

However, population density is strongly associated with communicable disease transmission during the COVID-19 pandemic (Sy et al., 2021). Our outcomes data predates the pandemic. Richardson et al. (2020) found that redlining nationwide is also strongly associated with worsened pandemic outcomes. Thus, population density might serve as an important control variable for studies that examine linkages between historical policies like redlining and communicable disease in the present day.

3. Chapter 2: Health Outcomes

We begin chapter 2 with our hypothesis for the association between living in a redlined neighborhood in the 1940 census and health outcomes in the near-present. We then propose methods, including a novel method for generating currently non-georeferenced enumeration districts. We run our analyses, report the results, and finish with chapter 2's discussion section.

3.1 Hypothesis

Our hypotheses are as follows:

H0: Living in a redlined neighborhood in the 1940 census is not associated with health outcomes in older adults today.

H1: Living in a redlined neighborhood in the 1940 census is associated with health outcomes in older adults today.

3.2 Methods

To create our analytic dataset, we combine data from multiple historical, geographic, and longitudinal sources. We use the restricted HRS 1940 Census complete count geographic dataset (Warren et al., 2020) with enumeration district information along with the RAND longitudinal version of the HRS (Bulgliari et al., 2023) for data on controls and dependent variables. For address locations to construct enumeration districts, we use the Geographic Reference File (Logan et al., 2020) and OpenCage database. For a subset of already defined enumeration districts, we use the city enumeration shapefiles of the Urban Transition Project (Logan and Zhang, 2018). For HOLC categories, we use HOLC maps as digitized and georeferenced by the Mapping Inequality Project (Nelson et al., 2017). Finally, we use the National Historical Geographic Information System's (NHGIS) dataset of 1940 county boundaries (McMaster et al., 2003).

3.2.1 Measure creation

3.2.1.1 Assigning HOLC redlining category to enumeration districts

In order to investigate any association between living in a redlined neighborhood in the 1940 census and health outcomes in older adults in the near present, we need to be able to assign HOLC categories to where those older adults lived in 1940. While the HRS 1940 Census complete count geographic dataset contains the identifier for each participant's 1940 ED (which would theoretically give us the participant's geographic location), HRS does not have the 1940 ED's geocoded with latitude, longitude, and/or shapefile polygons. Outside of HRS, the exact boundaries and geolocations of those 1940 ED's have also not been systematically georeferenced (meaning that maps that do exist are not projected and plotted with latitude/longitude or other geographic measures). Many were hand-drawn using whatever maps were available to census enumerators, including non-standard maps (Morse, 2007; Logan and Zhang, 2018). Thus, in order to assign HOLC categories to the 1940 ED's found in HRS, we need to be able to locate the 1940 ED's and draw their boundaries.

Enumeration districts were compact geographic areas that one enumerator could cover during the period of the census (Morse, 2007; Logan and Zhang, 2018). Thus the raw data from the 1940 full count census is organized by ED (Morse, 2007). Households in an ED were visited in a preset walking pattern. In urban areas, ED's were particularly dense, and might only cover one or two blocks, or even a single building. ED's were located within counties, and many cities had their own collection of ED's. ED's are not the same as census tracts, though census tracts were generally aggregated from ED's. ED's were intended to be merely an administrative tool not intended for public use; thus, unlike census tracts, ED's could shift drastically both in identifiers and geographic scope (e.g., size, shape, location) between each census (Morse, 2007; Logan and Zhang, 2018). The administrative nature of ED's also meant that ED boundaries were sometimes informally drawn on maps in non-standardized ways and preservation, digitization, and geocoding of ED's was not prioritized (Logan and Zhang, 2018).

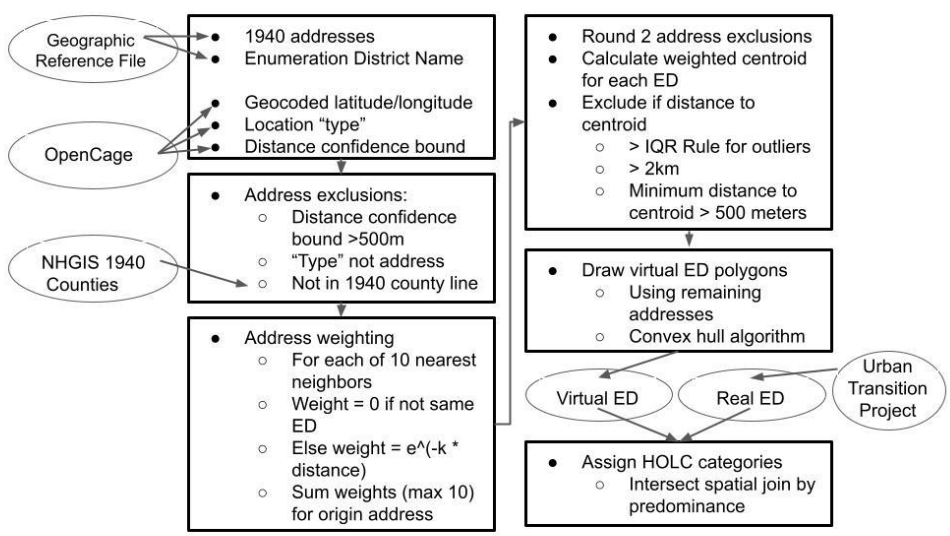
To start, we need a dataset of geocoded 1940 enumeration district geometries. We start with the Urban Transition Project's publicly available dataset of enumeration districts for 40 large US cities in 1940 as described in Logan and Zhang (2018). To summarize, each georeferenced enumeration district polygon is created based on three sources of data: digitized but not georeferenced 1940 enumeration district maps from the National Archives, 1940 enumeration district descriptions from the National Archives, and a set of recreated 1940 street names and locations based on multiple historic data sources. An address walk algorithm then locates specific addresses along recreated roads. The process is resource-intensive and time-consuming, though the results are likely highly precise.

Critically, however, the Urban Transition Project dataset's 40 cities do not contain enumeration districts for the 5 boroughs of New York City, nor the incorporated near suburbs of most metropolises (where a lot of particularly HOLC green A and blue B graded neighborhoods lay). To solve this problem, we supplemented these identified "real" enumeration districts with "virtual" constructed enumeration districts.

3.2.1.2 Construction of "virtual" enumeration districts

The general method by which we construct 1940 virtual enumeration districts is to brute force approximate the shape of the enumeration district via bulk address geocoding. We describe the method below and provide a flowchart (figure 1).

Figure 1: Flowchart of virtual enumeration district construction and assigning HOLC categories



We use the publicly available Geographic Reference File (Logan et al., 2020), which is a dataset of 191 largest US cities by population in 1940, also made publicly available by the team behind the Urban

Transition Project. The GRF is constructed from the IPUMS USA version of the 1940 fullcount census. For each city, the GRF contains the 1940 address for every enumerated individual from the 1940 census, along with the up-to-4 numeric digit plus 1 possible alphabetical digit suffix enumeration district (sometimes with and sometimes without the dash prefix).

Since we are aiming for the geometry of the enumeration district (rather than each individual's specific geographic location within that ED), we assume that the numbering and naming scheme of 1940 addresses has not shifted appreciably between 1940 and 2023. We further assume that the geolocation of the same address has not shifted appreciably on the surface of the Earth's sphere between 1940 and 2023 due to the reconstruction of streets and changes in city layout.

We take advantage of these assumptions to geocode 2023 latitude and longitude to 1940 addresses via OpenCage's georeferencing database. The OpenCage address matching algorithm is based in part on OpenStreetMaps. The exact matching and ranking algorithm is not publicly documented but it likely relies on partial string matching and relevance scoring.

While these assumptions may be easily violated for any given address (Logan and Zhang, 2018), in focusing on clusters of addresses and in aggregate, we should be able to estimate the location of enumeration districts and construct virtual approximations of those enumeration districts' polygons. We can also apply exclusion criteria to both trim individual addresses and to exclude malformed virtual ED's using what is known about how ED's were created.

We take each first occurrence of an address in the GRF with its attached ED marker. We use a geocoding API (a type of communication and programming protocol to query data) via Python (packages: opencage (OpenCage GmbH), asyncio (cpython/Lib/asyncio)) to query the OpenCage georeferencing database (bounded to US addresses). From the query, we geocode the latitude and longitude of a 2023 address, along with the open source tagged type of location returned and a distance confidence bound for each 1940 address. This provides us with a dataset of enumeration districts, each with its own set of multiple latitude and longitude pairs.

We retain geocoded locations only with a distance confidence bound of less than 500 meters. We also exclude geocodes if the target location is a road or neighborhood (rather than an address or address-like features). We exclude 2023 addresses if they lie outside the bounds of 1940 county lines from the National Historical Geographic Information System (NHGIS). For a handful of addresses that returned no result, we corrected malformed addresses by looking at the 1940 fullcount census pages (Morse, 2007) and queried OpenCage's database via its website, or discarded them if there were no matches.

Since we don't already know the location of 1940 enumeration districts, and since–due to unquantifiable erroneous geocoding and non-erroneous potential location changes–we don't know prima facie which geocoded 2023 addresses are correct, we use an iterative process to identify a density of points to draw each ED. We weight each address by a calculated likelihood to be within the 1940 enumeration district based on its clustering with other addresses. For each address, we calculate a 10 nearest neighbors distance matrix via ball tree (Python 3.11 packages: sklearn (Pedregosa et al., 2011), numpy (Harris et al., 2020), scipy (Virtanen et al., 2020), geopandas (Jordahl, 2014), pandas (McKinney et al., 2010)). We chose 10 nearest neighbors due to computational limitations. We upweight each address both by the proportion of 10 nearest neighbors that are from the same enumeration district and an exponential decay function on distance (weights are higher the closer a same ED neighbor is). The decay function is:

e^(-k * distance)

where k is a decay constant. Since the distance matrix was calculated at the kilometer scale, we set the decay constant to 5 to rapidly degrade weights as distance rises. In this manner, we generate a weight from 0 to 10 for each address point.

Using QGIS, we plot the weighted centroid of each putative enumeration district and calculate a new distance matrix from that centroid to each address point in the ED. We then exclude address points whose distances-to-centroid lie outside the Interquartile Range Rule for outliers (which is bounded at 1.5 x the interquartile range from the 3rd quartile). Since we are using urban enumeration districts, we also exclude any address points more than 2km from the centroid. Finally, to deal with centroids which may have ended up at the midpoint between false clusters, we exclude entire enumeration districts where the minimum distance to the centroid exceeds 500 m (i.e. no addresses found within 1 km diameter of the centroid). These are relatively conservative trims of address points.

A convex hull algorithm draws a polygon bounding a set of points, akin to stretching a rubber band over the outside of a set of pins. Using our final set of trimmed address points projected with a WGS84 Pseudo-Mercator projection, we use an alpha shapes convex hull algorithm via QGIS to draw the minimum bounding polygon around the address points. These polygons are our constructed virtual enumeration districts (appendix A: figures 2-10).

Potential issues include the fact that real enumeration districts can have an element of concavity (e.g. agglomerating a group of city blocks into the shape of a “L”). Additionally, these virtual enumeration districts are not mutually exclusive, and do sometimes cross each other. Future work could include adapting a concave hull algorithm and/or Thiessen/Voronoi shapes in a computationally feasible manner. Virtual enumeration districts in general are smaller than real enumeration districts. This may impact how HOLC category is discretized for the enumeration district (though a comparison of real to virtual enumeration districts in our analytic dataset finds very few differences in HOLC category assignment).

For each virtual enumeration district, we calculate a geometric centroid (Stata/MP 16.1 geodist package). For geographic locations with both virtual and real enumeration districts, we spatially join the virtual ED’s geometric centroid to the real ED polygon and compare their listed ED names both to aid in pairing and as a final error check.

3.2.1.3 Assigning HOLC Categories

We use an intersect spatial join in QGIS of 1930 HOLC shapefiles to our set of real and virtual enumeration districts. In cases of multiple overlapping HOLC categories, we assign the category by predominance. We do not have HOLC categories for Washington D.C. either because a redlining map was not created or the map was lost (Hillier, 2005).

We use a binary weight for assigning HOLC categorization to an enumeration district by whether the ED is “real” or “virtual”. For locations with both a real and a virtual ED polygon, our weight defaults to the HOLC category of the real ED polygon.

3.2.1.4 Creating Main Independent Variable

To operationalize our main independent variable, for each HRS participant in the 1940 full count sample, we assign HOLC category via their matched 1940 ED. We exclude participants where we do not have an ED with a joined HOLC category.

3.2.1.5 Health and Retirement Study

The HRS is described in detail elsewhere (Fisher and Ryan, 2018). HRS recruits US adults over the age of 50 and oversamples Black participants and participants living in Florida. The study follows the same participants in waves every two years until they become deceased or are otherwise lost to followup, and is still ongoing. Participants enter the study in birth period cohorts. We use the RAND longitudinal version of the dataset, described in detail elsewhere (Bugliari et al., 2023) that matches participants consistently across waves from 1992 (wave 1) to 2018 (wave 14), includes most core variables for each wave, and imputes and calculates economic variables. The HRS RAND longitudinal file includes variables for self-rated health and age at death.

The restricted 1940 fullcount geographic enumeration district level dataset has n=9602 HRS participants who were probabilistically matched to entries in the 1940 full count census version housed by IPUMS USA as described by Warren et al. (2020). Highly probable matched individuals and households were coded with their 1940 enumeration district identifier. We merge these records in turn to the HRS RAND longitudinal file.

3.2.1.6 Creating the Main Dependent Variables

We operationalize health as both survival time and self-rated health. We use age in months at death as our main outcome variable. If a participant was alive at the end of wave 14 in 2018, we use their age in months at the end of the interview as a right censor indicator.

Health is commonly operationalized not only in length but also quality of life. In this study, we operationalize quality of life with self-rated health at the first interview. The period leading up to death has been associated with drops in self-rated health (Mossey and Shapiro, 1982). We choose self-rated health at the first interview for a consistent measure. In accordance with common practice and the literature on how participants answer self-rated health (Manor et al., 2000), we dichotomize the measure by grouping Excellent with Very Good; and Good, Poor, and Fair.

3.2.1.7 Covariates and Controls

We operationalize propensity to migration using migration experience. The 1940 census includes a question on whether or not an individual lives in the same house as in 1935. We collapse all responses other than living in the same house and unknown into a category to reflect moved.

For race, we use RAND's crosswave coding of race which uses three categories of white, Black, and all other races regardless of hispanic ethnic assignment.

We use RAND's gender assignment variable, which has two categories of male and female.

For marital status, we collapse categories into 5 groupings to reflect potential social support systems or stressors from within the household that might impact health and other outcomes. These groupings are married or partnered; divorced, separated, or an absent spouse; widowed; and never married. We pick marital status at the date of death or last interview if the respondent is still alive.

Using reported birthplace regions, we also create a dichotomous variable for foreign birth to operationalize immigration status.

Education has a strong exogenous effect on life outcomes. To reflect the potential resources and limitations imposed on respondents due to different levels of educational attainment, we dichotomize educational attainment into those who graduated high school and those who did not (or who received their GED at a later date).

We collapse net household wealth into 5 ordered categories, with a category including zero and negative net wealth, then wealth quartiles. Please see chapter 3 for an expanded discussion of the wealth variable.

3.2.2 Analysis Methods

One way we could conceptualize analyzing the relationship of HOLC category exposure in 1940 with health outcomes today is as a time-to-event/survival prospective study with left truncation. In this case, the left truncation occurs because of the lag time between the 1940 census and the start of the HRS. Length of life is a common measure for health, and thus we can use time until death as our operationalized measure of health. However, because of mortality selection and survivorship bias (Suissa, 2008)–particularly given the large racial disparities in life expectancy throughout the 20th century–our analysis faces threats to external validity from average mortality selection effect and internal validity from variation in selection. Given the racial discrimination underlying redlining coupled with those racial disparities in life expectancy, it is plausible that many individuals from redlined areas may have died before the met the age-limited inclusion criteria for HRS, and thus those from redlined areas who survived long enough to enroll in HRS may have been especially healthy. If we look only at the average life expectancy of those individuals we observe in HRS, those living in redlined areas would look as if they live longer on average because our sample systematically misses all of those who had much shorter lifespans (Mayeda et al., 2018). As a result, our methods need to account for left truncation. Since HRS is ongoing and a proportion of our sample is still alive, our study is also subject to right-censoring.

We can use parametric normal distribution maximum likelihood estimation regression model to account for both left truncation and right censoring (Cain et al., 2011). Our variable entry time in the HRS (due to staggering both by cohort and by age within each cohort) might suggest a non-parametric product limit estimator (Cain et al., 2011). However, the near certain likelihood that left-truncated "observations" died earlier than the HRS inclusion criteria of age 50 would give us a biased estimator (Cain et al., 2011). Instead, we use a parametric model since "adult" or non-neonatal age of death is close to normally distributed (Tuljapurkar, 2010). Accounting for left truncation by modeling the shape of the distribution of "adult" death–which would include those who died before eligibility for HRS inclusion–should prevent attenuation of disparities like those found in Mayeda et al. (2018).

For our first model, we estimate hazards by HOLC category. In our second model, we include a set of controls. In our third model, we introduce migration experience to the fully controlled model.

For self-rated health, we perform a logistic regression of HOLC category on two levels of self-rated health. We run a second model with a full set of controls including propensity to migration.

For all models, we test statistical significance at the 95% confidence level and report 95% confidence intervals.

3.3 Results

Our sample covered 2973 enumeration districts, as reported in table 2. Of these, we constructed 2658 virtual ED’s, while 1727 had real ED’s, with 1412 ED’s having both virtual and real ED’s. 98% of the enumeration districts were either covered only by one type of district (real or virtual) or showed no difference in HOLC categorization when a counterpart existed. 2% of real enumeration districts had a different HOLC category from their virtual enumeration district counterpart, suggesting that the vast majority of virtual ED’s were correctly coded where both virtual and real ED’s were available.

Table 2:
Real Enumeration District (ED) Overlap with Virtual Enumeration Districts

Overlap Status	Frequency	Percent
No Difference	1,353	45.5
Difference	59	2.0
No Overlap	1,561	52.5

3.3.1 Tabulations

Tabulations are reported in table 3. 2.8% of our sample lived in green A enumeration districts, 14.3% in blue B, 41.6% in yellow C, and 41.3% in red D.

Table 3:
Summary Statistics

Variable	Frequency/units	Mean (SD)/Percent
Total HRS RAND sample	42233	
1940 Census Geo HRS sample	9602	
Analytical sample	2974	
Year of birth	year	1925. 6 (10.2)

Age at first interview (in years)	years	68.3 (10.8)
HOLC categories		
1.A (green)	83	2.79
2.B (blue)	425	14.29
3.C (yellow)	1,237	41.59
4.D (red)	1,228	41.29
Age at death in months by HOLC category		
1.A (green)	months	1025 (130)
2.B (blue)	months	1018 (107)
3.C (yellow)	months	1007 (111)
4.D (red)	months	995 (113)
Total	months	1004 (112)
Died during study period		
0.Right Censored	890	29.93
1.Died	2,084	70.07
Race		
1.white/caucasian	2,672	89.85
2.Black/african american	269	9.05
3.All other	33	1.11
Gender		
1.male	1,364	45.86
2.female	1,610	54.14
Marital Status		
0.Never Married	119	4
1.Married/Partnered	1,154	38.8
2.Divorced/Separated/Absent	368	12.37
3.Widowed	1,333	44.82
Foreign-born		
0.US-born	2,884	97.04
1.Foreign-born	88	2.96
1st interview total wealth ($)	$	269185.4 (508324)
1st interview house net value ($)	$	453101 (1402610)
Wealth in 5 categories		
0.Negative or 0	148	4.98
1.Q1	519	17.45
2.Q2	682	22.93
3.Q3	862	28.98
4.Q4	763	25.66
House Net Value 5 categories		
0.Negative or 0	626	21.05

1.Q1	425	14.29
2.Q2	528	17.75
3.Q3	744	25.02
4.Q4	651	21.89
Education		
0.GED or Less than HS grad	790	26.56
1.HS grad or more	2,184	73.44
Moved last 5 years 1940 census		
0.Same House	818	27.51
1.Has Moved	1,610	54.14
3.Unknown	546	18.36
Insurance 1st Interview		
0.Medicare	1,681	56.77
1.Medicaid	31	1.05
2.Dual Medicare and Medicaid	103	3.48
4.Private	1,041	35.16
6.Uninsured	105	3.55
Insurance last Interview		
0.Medicare	2,437	82.3
1.Medicaid	29	0.98
2.Dual Medicare and Medicaid	293	9.9
4.Private	164	5.54
6.Uninsured	38	1.28
Insured 1st Interview		
0.Uninsured	105	3.55
1.Insured	2,856	96.45
Insured Last Interview		
0.Uninsured	38	1.28
1.Insured	2,923	98.72
Doctor in last year 1st Interview		
0.No	368	12.38
1.Yes	2,604	87.62
Doctor in last year last Interview		
0.No	160	5.38
1.Yes	2,812	94.62
Self rated Health 1st Interview		
0.Excellent or Very Good	1,359	45.71
1.Good, Fair, or Poor	1,614	54.29

For those participants who became deceased, mean age at death was 1,004 months (approximately 83.7 years). 45.7% rated their own health at the first interview as Excellent or Very Good, and 54.3% as Good, Fair, or Poor. The ratio between those who rated their health as Excellent or Very Good to those who rated their health as Good, Fair, or Poor reduces as we move from green A enumeration districts towards redlined D enumeration districts (table 4).

Table 4:
Self-rated health by HOLC

HOLC	Excellent/Very Good	Good/Fair/Poor	Ratio
A	51	28	1.82
B	218	183	1.19
C	581	584	0.99
D	508	644	0.79
Total	1358	1,439	0.94

3.3.2 Survival analysis

Survival analyses are reported in table 5. Our base hazards model finds that redlined D enumeration districts are statistically significantly associated at the 95% confidence level with increased hazards when compared to green A areas. The coefficient for red D is -0.40 (95% CI: -0.77 to -0.026).

Table 5:
Survival Analysis

	Coefficient	Z score	95% CI	
Unadjusted model				
HOLC				
A	ref			
B	-0.138	-0.690	-0.530	0.254
C	-0.330	-1.750	-0.700	0.039
D	-0.396	-2.100	-0.765	-0.026
Adjusted model				
HOLC				
A	ref			
B	-0.130	-0.750	-0.469	0.210
C	-0.241	-1.470	-0.562	0.080

D		-0.236	-1.430	-0.559	0.087
Moved 1935-1940		-0.027	-0.490	-0.137	0.082
Race					
white	ref				
Black		-0.069	-0.660	-0.275	0.137
All other		-0.431	-1.560	-0.974	0.111
Gender					
male	ref				
female		0.345	5.710	0.227	0.464
Marital Status					
Married/Partnered	ref				
Never married		0.215	1.480	-0.069	0.500
Divorced/Separated		0.031	0.340	-0.148	0.210
Widowed		0.717	10.780	0.586	0.847
Foreign-born		-0.002	-0.010	-0.388	0.384
Total Wealth					
0 or negative	ref				
Q1		0.257	1.720	-0.036	0.551
Q2		0.422	2.860	0.133	0.711
Q3		0.696	4.700	0.406	0.986
Q4		0.922	6.040	0.623	1.221
Education					
GED or less than HS	ref				
HS grad		0.319	4.600	0.183	0.455
Insured		0.388	2.860	0.122	0.654

When controls are added (including race, gender, and wealth), no HOLC category is significantly associated with increased hazards. The top 3 wealth levels are statistically significant controls, as well as gender, being widowed, education, and health insurance status.

In the full model, with all controls and HOLC category, migration experience in the last 5 years is not statistically significant, and its inclusion in the model does not appreciably change the coefficients or significance of either our main independent variable or covariates and controls.

3.3.3. Self-rated health

Self-rated health is reported in table 6. For self-rated health, “yellow-lined” C and redlined D enumeration districts are statistically significantly associated at the 95% confidence level with decreased self-rated health. Yellow C is associated with 0.59 increased log odds (95% CI: 0.13 to 1.04) of reporting only Good, Fair, or Poor health. Red D is associated with 0.82 increased log odds (95% CI: 0.36 to 1.27) of reporting only Good, Fair, or Poor health.

Table 6:
Self rated health at entry

	Coefficient	Z score	95% CI	
Unadjusted model				
HOLC				
A	ref			
B	0.414	1.690	-0.067	0.896
C	0.586	2.520	0.130	1.042
D	0.815	3.500	0.358	1.271
Adjusted model				
HOLC				
A	ref			
B	0.344	1.360	-0.153	0.840
C	0.466	1.940	-0.004	0.936
D	0.528	2.180	0.054	1.002
Moved 1935-1940	0.245	3.100	0.090	0.400
Race				
white	ref			
Black	0.319	2.040	0.013	0.625
All other	0.032	0.080	-0.715	0.780
Gender				
male	ref			
female	-0.226	-2.610	-0.395	-0.057
Marital Status				
Married/Partnered	ref			
Never married	0.301	1.430	-0.111	0.712
Divorced/Separated	0.086	0.660	-0.169	0.341
Widowed	0.120	1.270	-0.066	0.307
Foreign-born	0.133	0.550	-0.341	0.607
Total Wealth				

0 or negative	ref				
Q1		-0.147	-0.660	-0.580	0.286
Q2		-0.420	-1.940	-0.844	0.005
Q3		-0.654	-3.020	-1.078	-0.230
Q4		-0.995	-4.490	-1.430	-0.560
Education					
GED or less than HS	ref				
HS grad		-0.880	-8.980	-1.072	-0.688
Insured		0.425	2.000	0.009	0.842

When controls are added, redlined D enumeration districts continue to be associated with worsened self-rated health, reporting 0.53 increased log odds (0.054 to 1.00) of reporting only Good, Fair or Poor health. C enumeration districts are no longer significantly associated with worsened self-rated health.

Race, migration experience in the last 5 years, gender, top 2 categories of wealth, education and insurance are all statistically significant controls in the model.

3.4 Discussion

Redlined enumeration districts have some association with health outcomes in our base models. For base models, redlined enumeration districts are associated both with increased hazards and reduced self-rated health. In the fully adjusted model, redlined enumeration districts are associated with reduced self-rated health.

Race and wealth are both non-exclusive pathways to health from HOLC in our conceptual framework, so we would expect to find the strength of any associations between redlining and health to be muted in their presence. HOLC areas were drawn with race in mind implicitly and explicitly, and HOLC's anti-Black racial attitudes and concepts of creditworthiness (along with xenophobia) were enshrined in lending and investment practices for decades that locked generations out of home ownership and thus a source of financial resources to spend on health production (Michney, 2022). This and other policies have led to systemic reductions in life expectancy and health for Black residents. At the same time, divestment and concurrent and subsequent political disenfranchisement in Black neighborhoods have been associated with worsened environmental outcomes (Pappoe, 2016; McCoy, 2017; Naman et al., 2020; Tragenstein et al., 2020) that also harm health (Bailey et al., 2017; Jacoby et al., 2018). While a mediation analysis is beyond the scope of this analysis, future work should examine race and wealth as possible mediators, or conversely examine HOLC category as a possible instrument for race and wealth.

Previous work has found that redlined areas are associated at both area (Swope et al., 2022; Richardson et al., 2020; Wilson et al., 2018; Alexander and Currie, 2017; Nardone et al., 2020; Lynch et al., 2021; Huggins, 2017) and individual levels with worsened health outcomes for present day residents (Krieger, Wright, et al., 2020; Krieger, Van Wye, et al., 2020), including on mortality and life expectancy. Our findings are broadly consistent with such findings that show that the racialized construction of geographic space is associated with ultimate health outcomes. We examine the association with individuals who were exposed shortly after redlined maps were drawn, and find associations with health outcomes attenuated by race and wealth controls.

Our conceptual framework does not explicitly account for a relationship between HOLC category and gender (though given the ubiquity of both racialized gender expressions (Steele, 2018) and gendered racial barriers (Amuchie, 2015), it’s certainly not implausible to find potential pathways). However, gender is also associated with longevity and life expectancy, and is an often used control variable.

Migration experience–as measured by whether an individual lived in the same house in 1935 as in 1940–was not statistically significant in the hazards model. One possible interpretation is that geographic mobility (or its lack) and “healthy migrant” selection effects cannot adequately explain increased hazards in redlined areas.

We have created a new data resource and method in locating and generating virtual 1940 ED’s that can be used for future research into both redlining and work examining health and geographic linkages in the 1940 census.

4. Chapter 3: Economic Outcomes and Healthcare Utilization

In chapter 3, we explore the potential pathway through economic outcomes to link redlining to ultimate health outcomes. We begin with our hypothesis, propose and run methods to test, report results, then finish the chapter with a discussion. We then wrap up the manuscript with a broader discussion.

4.1 Hypothesis:

Our hypotheses are as follows:

H0: Living in a redlined neighborhood in the 1940 census is not associated with economic outcomes and healthcare utilization in older adults today.

H1: Living in a redlined neighborhood in the 1940 census is associated with economic outcomes and healthcare utilization in older adults today.

4.2 Methodology:

4.2.1 Data Sources

We use the restricted HRS 1940 fullcount geographic enumeration district dataset merged with the HRS RAND longitudinal file as described in greater detail in chapter 2. HRS includes economic and healthcare utilization variables that could allow us to investigate our conceptual framework linking HOLC categorization through economic outcomes and to healthcare utilization. The RAND longitudinal file includes imputed economic variables, as well as measures on the use of healthcare. For geographic data on enumeration districts, we combine enumeration district data from the Urban Transitions Project, the Geographic Reference File, and the OpenCage database. For HOLC data, we use the 1930s HOLC maps digitized and georeferenced by the Mapping Inequality Project. For 1940 county boundaries, we use the NHGIS dataset.

4.2.2 Creating Measures

4.2.2.1 Independent Variable - HOLC Category

As described in chapter 2, we create our own “virtual” enumeration district dataset by using the OpenCage database to geocode 1940 addresses from the Geographic Reference File, and then creating enumeration district polygons with a convex hull algorithm around same-district addresses. We combine our “virtual” 1940 enumeration districts with the 1940 enumeration districts from the Urban Transitions Project and weight the dataset to prefer Urban Transitions Project districts when both exist. We assign HOLC categories to 1940 enumeration districts by spatially joining enumeration district polygons with 1930s HOLC maps digitized and georeferenced by the Mapping Inequality Project. Error checking was done in part by comparison to 1940 county boundaries from the NHGIS dataset. We code each participant with a four part HOLC category through their matched 1940 enumeration district.

4.2.2.2 Control Variables

The methods for creating control variables were discussed in further detail previously in chapter 2. In summary, we create controls for race, gender, marital status, foreign birth, educational attainment, and migration experience using HRS data. While net household wealth is an outcome variable for the first set of our analysis on economic outcomes in this chapter, it serves as a control for the second set of our analysis on healthcare utilization.

4.2.2.3 Main Dependent Variables - Economic Outcomes

For our main economic outcome, we focus on wealth. Most of the population is at a life stage where they begin or have already begun to prefer replacing consumption from income with consumption from savings (if those exist). Wealth and assets may be a better measure than income of the level of resources that are available for healthcare decisions at this stage (Allin et al., 2009).

We operationalize wealth using total net household wealth. The measure includes primary residence, additional real estate, liquid assets, and retirement accounts. RAND imputes this measure using nearest neighbor analysis whenever component variables are missing. We measure this at the first interview date to reduce variability that may occur over time and to reflect the level of assets that are available for prospective healthcare budgeting. The measure was collected at the household level which would reflect resource sharing that occurs within households. We use net household wealth since we're interested in the level of resources available to individuals for healthcare decision making. Similarly, we use the net wealth at the first interview to reflect the resources that are available to households for prospective healthcare budgeting (as would be indicated by the Aday-Andersen model in our conceptual framework).

To account for right skew, we collapse net household wealth into 5 sequential categories. We create one category for zero and negative net wealth (a situation where a household is low resourced or underwater with debt obligations), and create quartiles for net accumulated wealth above that point. Log transforms of accumulative variables (like income, wealth, etc.) are often used to address the right skew of these data. However, the relatively large proportion of zero and negative net wealth values in our data renders a log transformation inappropriate for our purposes.

We also measure net house or primary residence value. Many US residents rely on housing as an asset value store (in addition to its use as shelter). HOLC in particular was concerned with residential mortgage risk, and much of the public history of redlining has been focused on the impact that redlining had on home values and housing quality (Aaronson et al., 2017). Similarly to net wealth, we collapse house or primary residence values into 5 sequential categories.

4.2.2.4 Main Dependent Variables - Healthcare Utilization

We break healthcare access into two parts: health insurance and healthcare utilization.

We operationalize health insurance in two ways. We are primarily interested in whether or not an individual is insured or uninsured. Given the high uptake of Medicare the older the population gets, we looked at whether or not an individual had insurance at first interview: this is the same measure of health insurance status used as a control variable in chapter 2. We also look at whether HOLC category is associated with the type of health insurance carried at first interview, broken into Medicare, Medicaid, dual Medicare and Medicaid, other government provided insurance, private insurance, other insurance, and uninsured. When a respondent had both government provided and private insurance or other insurance (e.g., through Medicare Part C managed care plans), we categorized them into government provided insurance.

We operationalize healthcare utilization through a dichotomous variable of whether or not an individual visited the doctor over the last year. We measure at the first interview to match the respondent's health insurance status in that year.

4.2.3 Analysis Methods

4.2.3.1 HOLC categories' association with race

We run a multinomial logistic regression of HOLC grade on race of respondent with white respondents and green A HOLC category set as the reference points. Multinomial logistic regressions output relative log odds for the likelihood of belonging to a different group from the reference group for the independent variables and controls for non-binary categories. Race of the respondent as coded by HRS is a good approximation for a non-binary and non-ordinal category.

4.2.3.2 Economic Outcomes

We run ordered logistic regression of HOLC categories on sequential wealth categories. Ordered categorical variables have ordinal relationships with each other but do not have consistent "distance" between any levels. Our sequential categories of wealth fit the mold of an ordered category. Ordered logistic regressions output log odds for belonging to the next ranked group for the independent variables and controls. We then add full controls (absent wealth of course). We run ordered logistic regression of HOLC categories on sequential net house value categories with full controls.

4.2.3.3 Healthcare Utilization Outcomes

For health insurance, we run logistic regression of HOLC categories on whether an individual is insured or uninsured at the first interview date. We then add full controls. Next, we run multinomial logistic regression of HOLC categories on the type of insurance or being uninsured. We then add full controls.

For visiting the doctor, we run a logistic regression of HOLC categories on whether an individual had visited their doctor in the previous year. We then add full controls.

For all models, we test statistical significance at the 95% confidence level and report 95% confidence intervals.

4.3 Results

4.3.1 Tabulations

All tabulations are reported in table 2. 2.8% of our sample live in green A enumeration districts, 14.3% in blue B, 41.6% in yellow C, and 41.3% in red D.

9.1% of our sample is Black, 89.9% white, and 1.1% all other races. Mean age at the first interview was 68 years.

Only 3.6% of the sample is uninsured at the time of the first interview. 56.7% of the sample is on Medicare, 1.1% Medicaid, and an additional 3.5% has dual Medicare/Medicaid coverage. 35.2% is on private insurance.

87.6% of the sample had visited a doctor in the 12 months before the time of the first interview.

4.3.2 HOLC Category and Race

HOLC grade of the enumeration district is statistically significantly associated with the race of the respondent at the 95% confidence level. Respondents in redlined D enumeration districts had 2.97 (95% CI: 0.99 to 4.95) the log odds of being Black instead of white compared to respondents living in green A enumeration districts (table 7).

Table 7:
Race by HOLC Composition

	Coefficient	Z score	95% CI
white	ref		
Black			

A	ref			
B	1.027	0.990	-1.015	3.069
C	0.448	0.440	-1.566	2.462
D	2.968	2.940	0.992	4.945
All other races				
A	ref			
B	0.041	0.000	-3018	3018
C	14.014	0.010	-2751	2779
D	15.208	0.010	-2750	2780

4.3.3 Economic Outcomes

Wealth outcomes are reported in table 8. For economic outcomes, HOLC grade is statistically significantly associated with the log odds of being in the next higher wealth category in the unadjusted model. Those in enumeration districts with redlined D categories had -0.69 (95% CI: -1.09 to -0.29) the log odds of belonging to the next higher wealth category compared to those in green A categorized enumeration districts. Those in B and C categories as compared to A also resulted in lower log odds of being in the next higher wealth category (with the coefficient for B being less than C being less than D), but the results were not statistically significant.

Table 8:
Total wealth 5 categories

	Coefficient	Z score	95% CI	
Unadjusted model				
HOLC				
A	ref			
B	-0.142	-0.660	-0.562	0.278
C	-0.330	-1.640	-0.726	0.065
D	-0.690	-3.410	-1.086	-0.293
Adjusted model				
HOLC				
A	ref			
B	-0.004	-0.020	-0.434	0.426
C	-0.172	-0.830	-0.578	0.235
D	-0.229	-1.100	-0.638	0.180
Moved 1935-1940	-0.128	-1.870	-0.262	0.006

Race				
white	ref			
Black	-1.428	-11.320	-1.675	-1.180
All other	-0.513	-1.460	-1.200	0.174
Gender				
male	ref			
female	-0.401	-5.400	-0.547	-0.256
Marital Status				
Married/Partnered	ref			
Never married	-1.517	-8.420	-1.870	-1.164
Divorced/Separated	-1.017	-9.070	-1.237	-0.797
Widowed	-0.505	-6.140	-0.666	-0.344
Foreign-born	-0.645	-3.310	-1.027	-0.263
Education				
GED or less than HS	ref			
HS grad	1.269	16.000	1.114	1.425
Insured	1.023	5.630	0.667	1.378

When we add full controls, the magnitude of coefficients for HOLC categories as compared to green A areas are smaller compared to the coefficients from the base model and no HOLC category remains significant. All controls were statistically significant in the model, including race, gender, marital status, foreign birth, education, and migration experience. A similar pattern holds for housing values where red D have statistically significantly lower log odds of being in the next higher housing value category when compared to green A in the unadjusted model, but is not significant in the fully adjusted model.

4.3.4 Healthcare Utilization Outcomes

HOLC grade is not statistically significantly associated with whether or not an individual had health insurance at the 95% confidence level. It was not significantly associated with the type of insurance individuals had either (table 9). This null finding holds when the full set of controls are added to either model.

Table 9:
Insured at study entry

HOLC	Coefficient	Z score	95% CI

A	ref			
B	-0.073	-0.090	-1.599	1.452
C	-0.497	-0.680	-1.929	0.936
D	-0.415	-0.570	-1.850	1.020

HOLC grade is also not significantly associated with whether or not an individual visited the doctor in the last year (table 10). Again, the null finding holds when the full set of controls are added.

Table 10:
Likelihood of seeing a doctor in last 12 months at study entry

HOLC	Coefficient	Z score	95% CI	
A	ref			
B	0.071	0.190	-0.655	0.796
C	-0.037	-0.110	-0.719	0.645
D	-0.061	-0.180	-0.743	0.621

4.4 Discussion

We find that 1940 exposure to enumeration districts predominantly consisting of redlined areas in the 1930s is statistically significantly associated with economic outcomes of those same individuals starting in 1992. Moreover, the association attenuates and is no longer significant once we control for race, educational attainment, and other controls. While a full mediation analysis is out of scope for this work, this finding is consistent with our framework that HOLC is a proxy for other policies that created structural barriers by race via geographic restrictions in the 20th century.

Previous literature (Aaronson et al., 2017) have found redlining to still be associated with area measures of economic outcomes for current residents. We find an association to exist for individual original residents as well in our base model, but the result is not significant with controls. Current residents may or may not be the same as original residents, and a future direction for research would be to identify whether residents who still live in redlined areas have different outcomes from those who have moved to non-redlined areas.

Our conceptual framework posits one pathway between economic outcomes and health is through healthcare utilization. The Aday-Andersen model (1974) suggests that reduced economic outcomes from exposure to redlined enumeration districts would translate to fewer resources (a key enabling resource) to use on healthcare, and thus reduced health access and healthcare utilization. However, we found no association between HOLC categories and either health insurance access or visiting the doctor in either unadjusted or adjusted models.

The elderly age of the population at entry may mean that much of the population who would have been uninsured at a younger age due to a lack of economic resources were covered by Medicare. Along with the high rate of Medicare uptake, this may also point to a success in an universal entitlement healthcare financing scheme in providing health insurance access despite different economic outcomes (albeit late in life), as would be suggested by Fundamental Cause theory (Link and Phelan, 1995).

Our health insurance result appears less consistent with the findings by Nardone et al. (2020) and White et al. (2021), both of which did find redlined areas to have lower rates of health insurance. However, both those studies looked at health insurance uptake across the whole population of an area, which would include everyone who does not qualify yet for Medicare and thus would not benefit from its near universal at-age coverage for an older study population in our case. Additionally, both studies looked at present day residents rather than original residents. Again, a future study should examine whether previous residents of redlined areas continue to live in those areas. Additionally, future work could examine other patterns in health insurance coverage in an older population, such as temporary periods of being uninsured and churn.

HOLC category was not associated with any visits to the doctor in the last year. As a more proximate enabling factor, high rates of health insurance might instead be a more important factor in determining healthcare utilization. Since HOLC grade does not appear to be associated with health insurance in our study, it’s possible that doctors visits are more strongly associated with health insurance or more proximate associated economic outcomes. This result is inconsistent with the finding by Poulson et al. (2021), which did find redlined areas to have lower rates of healthcare utilization in the form of preventative lung cancer screenings among Black patients. We did not test for other potential measures of healthcare utilization such as number of visits, appropriateness of visits, or type of visit. A future study might examine the number or frequency of doctors visits–while controlling for external factors that may also influence how often patients visit the doctor. Further, this may suggest that health is not solely or primarily related to medical care, which is broadly consistent with other findings about the contribution of medical care to overall health (Kaplan and Milstein, 2019).

Enumeration districts with red HOLC category are associated with the racial makeup of an area. In line with the retrenchment in American views around structural racism after 2020, much has been written in the last few years (see Michney, 2022 for a discussion of such writings) about how HOLC categories were wholly unrelated to the race of the individuals living in those neighborhoods. Leaving aside the explicitly race-based notes found in HOLC maps (Nelson et al., 2017) and new scholarship on how racial attitudes of the HOLC staff suffused future thinking on credit risk (Michney, 2022), and leaving aside the long-standing treatment of race not as a coincidental and context-less (and ahistorical) “fact” but rather a sociological construct bound up with other co-occurring sociological constructs to maintain power and economic relationships (Cox, 1948), we find evidence strongly consistent with the predominant view that HOLC categories are race-based. As such associations between redlining with economic outcomes are likely to be racially disparate even if there are Black communities that were not redlined and which may or may not attenuate such outcomes.

5. Final Discussion

This is the first work to our knowledge that explores exposure to HOLC redlining categories in 1940 by enumeration district at a large nationwide scale along with health outcomes in those same individuals in the near-present. We also created a new method and dataset for future studies that use 1940 urban enumeration districts. We find no evidence that a “race neutral” construct of population density serves as a mediator for the association between redlining and area life expectancy in Baltimore. When looking at individuals who were exposed to redlining in the first census after redlining, we find some associations with health outcomes that are attenuated by race and wealth measures. Self-rated health continues to be statistically significantly associated with 1940s experience with redlining even after applying controls. We find associations with economic outcomes without controls, but no associations for health insurance or healthcare utilization.

5.1 Limitations

Enumeration districts are area constructions that do not align perfectly with the non-cartographic categorization of HOLC areas. However, they do have a much higher resolution due to their smaller size and geographic compactness than larger geographic units such as census tracts or neighborhoods, particularly in dense urban areas.

Missing enumeration districts are not likely to be all missing completely at random since the available GRF dataset is by size of city population. As a result, we know we are missing enumeration districts for smaller incorporated cities (there were a total of 412 US cities in 1940 (Census Bureau, 1950)). However, most of these locations do not have an existing HOLC map, so our enumeration districts cover the vast majority of the geographic space covered by HOLC maps.

Among cities that are covered, some portion of enumeration district missingness is random and some is likely not. Typographic errors in recording street names (or in how they were miscoded by modern day records digitizers) are likely to be missing completely at random. On the other hand, geocoding errors–where the OpenCage API returned to us an address that we deemed not likely to be within the bounds of the 1940 enumeration district–are possibly missing because roads and originally assigned houses were moved or demolished. Thus enumeration districts that were trimmed because they failed our minimum address radius test may be more likely to be areas that underwent drastic urban renewal and reconstruction throughout the late 20th and early 21st centuries, which in turn were more likely to be areas where politically marginalized communities resided in 1940. While an analysis of the features of trimmed enumeration districts is not possible at this time (since not knowing where they are, we cannot map area features or HOLC categories to them), this is an area where future work is necessary.

HOLC categorizations are not necessarily a unified sociological construction between cities/metropolitan areas (Swope et al., 2022). As a power hierarchy, the racial power hierarchy in the US is not an internally consistent construct, but rather may vary in its porousness or the saliency of certain “rules” within each local context to maintain that power structure (Cox, 1948). As such, the decision to mark certain similar areas as red, yellow, blue, or green reflects local racial attitudes and thus varies within the general bounds of American racial policy from city to city (Michney, 2022). Additionally, local political contexts are likely to determine how resources and investments were allocated to and from politically marginalized communities from 1940 to the present day (Nelson et al., 2017). For this study, we treated HOLC categories as if they were consistent across cities, though we ran some followup analysis to investigate this assumption. In subsequent runs, we did include indicator variables on state and then on city and did not observe a meaningful or significant difference in coefficients or confidence bounds for main variables or controls from our main models (though our sample is likely not powered enough for such a method).

5.2 Future Directions

Future analysis should look at whether or not individuals still live in the same HOLC grade as they did in 1940. This would help us answer both whether associations found on the individual level are more associated with initial or continued exposure to areas redlined in the 1930s, and whether present-day area level effects in redlined neighborhoods are more associated with population change or lack of resources and investment, especially given the weak or non-existent associations found with both proximate and ultimate endpoints.

Childhood has been identified as a key period in setting trajectories of health (Krieger, Van Wye, et al., 2020). Initial exposures in housing and neighborhoods may have stronger associations due to how they act on childhood development, whether through physical environment–such as lead (Pappoe, 2016)–or social environment–such as Adverse Childhood Experiences (Felitti et al., 1998). In successive runs we did not find meaningful or significant differences from including age in the 1940 census or year of birth on our results, but a future study might more directly examine that pathway.

Future analysis should look at the individual address level, since we know georeferenced HOLC categories with far greater certainty than we do enumeration districts, and since HOLC categories are non-cartographic with respect to any other cartographic division. Exact address matching would also allow us to include other measures of interest that are strongly associated with health outcomes from a young age, including distance to automobile traffic, travel time, etc. However, due to changes in city geographies, exact address matches should be used with caution or snapped to an enumeration district before a dataset of georeferenced 1940 addresses is constructed (rather than simulated from present-day addresses). Density calculation and address trimming may be good enough to fit the geometry of an enumeration district, but would be more difficult to verify for each specific address.

A larger dataset, such as vital records, that could be matched to the 1940 fullcount census could also be a way to tie together historical exposures with health throughout the 20th and early 21st centuries and into the present.

It is clear when creating the dataset that much of the antecedent conditions stretch to the early postbellum or possibly even antebellum period. Areas where new city arrivals lived (many of whom were Black and/or immigrants) often contained the oldest housing stock (Michney, 2022; Logan et al., 2020). Additional work could be done to link histories of place, forced migration, and compulsory labor throughout the 19th and 20th centuries to help us understand the health landscape in the present day.

Since we do not theorize the HOLC maps to be in the direct causal pathway for health outcomes, but rather a proxy for existent racial policy, we focused primarily on associations. However, in future studies, an instrumental variable approach may help us control for unobserved confounding. One possible approach is to exploit enumeration district numbering schema in each city, where numbers seem to have been assigned to districts in a circular within geographic cluster and circular between cluster pattern. Thus, the modulo position of the ED's number within a geographic cluster of EDs may be related to its distance from the city center and thus its likelihood to be redlined *within the cluster* without being related to any other potential pathway. This method would require a Fourier Transform to find the period length of clustering.

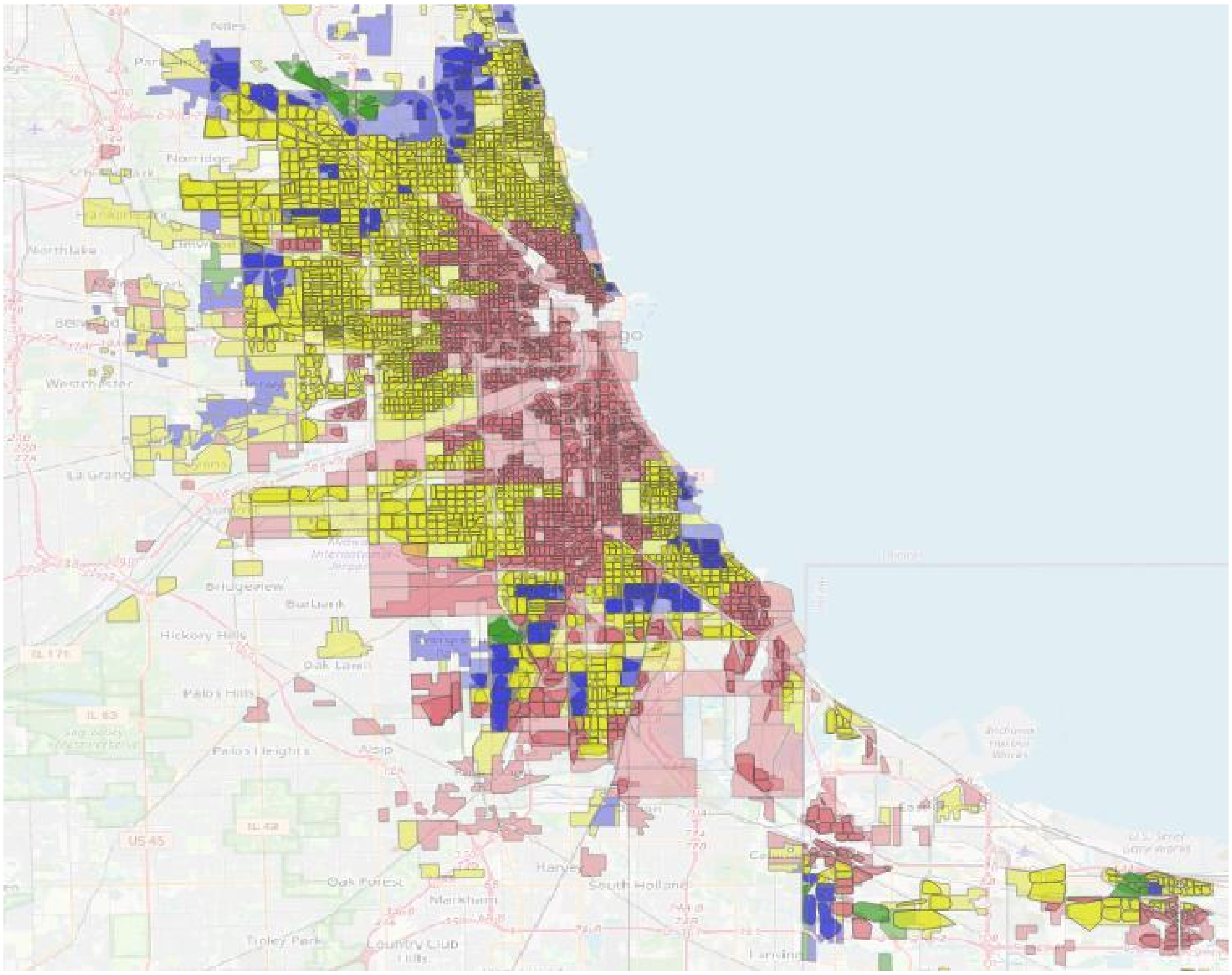
Machine learning algorithms might also potentially help us process and better draw enumeration districts. For data processing, machine learning could be used in reading and processing enumeration district map data, as well as using optical character recognition to read in handwritten full count data (such as the now available 1950 fullcount census data). Language models might also be used to better parse or correct misinterpreted atypical shorthand and address names in the fullcount data (for example in places where an enumerator drew a line to indicate the address is a continuation from the page before, or where "41st ave" was reinterpreted by a modern day coder as "41 Saint Avenue"). On the data analysis side, machine learning could be used to draw constructed enumeration districts with higher precision and to identify erroneous geocodes automatically. We briefly experimented with an isolation forest algorithm using address data geocoded from OpenCage that drew virtual enumeration districts that matched extremely well with the Urban Transition Project's enumeration districts, but the process was technically beyond our capabilities for now and computationally intensive at scale.

5.3 Policy Implications

Finally, at a community and societal level, we need to grapple with how the impacts of policy histories on health create reparative responsibilities for governance. Institutional, local, state, and national governmental policies have created a set of ongoing conditions (likely deliberately) through political disenfranchisement that continue to rob a portion of our population of the full length and quality of life that they are owed. There is no single policy that can be pointed to, but rather a matrix of path-dependent and mutually reinforcing policies through time to maintain a power structure that thus degrades health and well-being. What responsibilities to compensation, structural reshaping, and political empowerment are implied and should be demanded of government? Health equity frames can and have served as a strong impetus and rhetorical cudgel for community organizing (Pastor et al., 2018) including during the ongoing COVID-19 pandemic.

Appendix A: Figures

Figure 2: Chicago real and virtual enumeration districts tagged by HOLC. HOLC backdrop.



Legend: light transparent shade: HOLC backdrop, darker bold shade with rounded border polygons: virtual ED, straight border polygons: real ED, red: HOLC D, yellow: HOLC C, blue: HOLC B, green: HOLC A.

Figure 3: Street grid of a section of present day Philadelphia, PA, USA

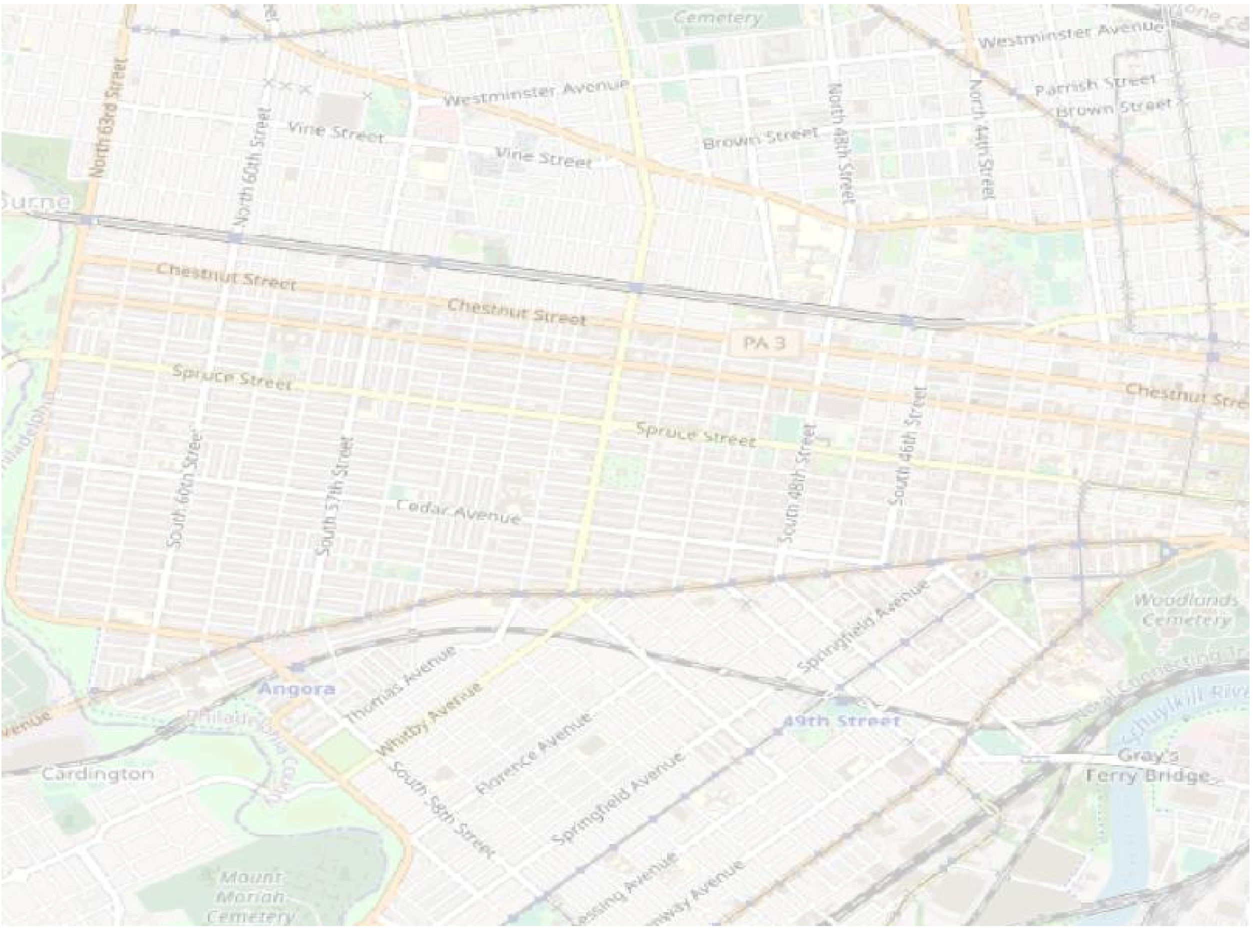
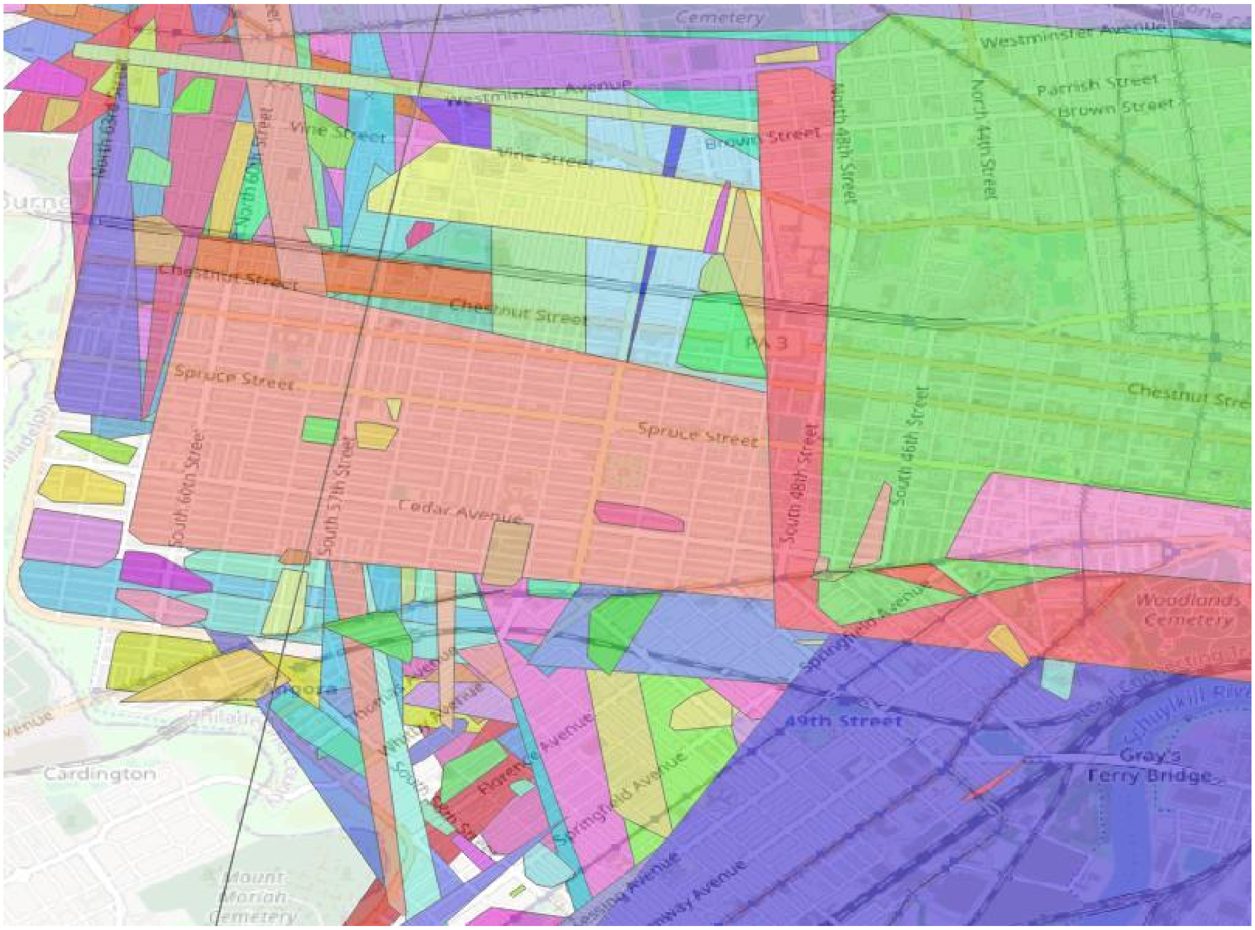


Figure 4: Philadelphia geolocated 1940 addresses by enumeration district (untrimmed)

Legend: dot: one 1940 address where color refers to enumeration district.

Figure 5: Philadelphia constructed virtual enumeration districts without trimming addresses



Legend: polygon: one virtual ED constructed from untrimmed addresses where color refers to different ED's.

Figure 6: Philadelphia trimmed address virtual enumeration districts overlaid with untrimmed addresses

Legend: dot: one 1940 address where color refers to ED, polygon: one virtual ED constructed from trimmed addresses.

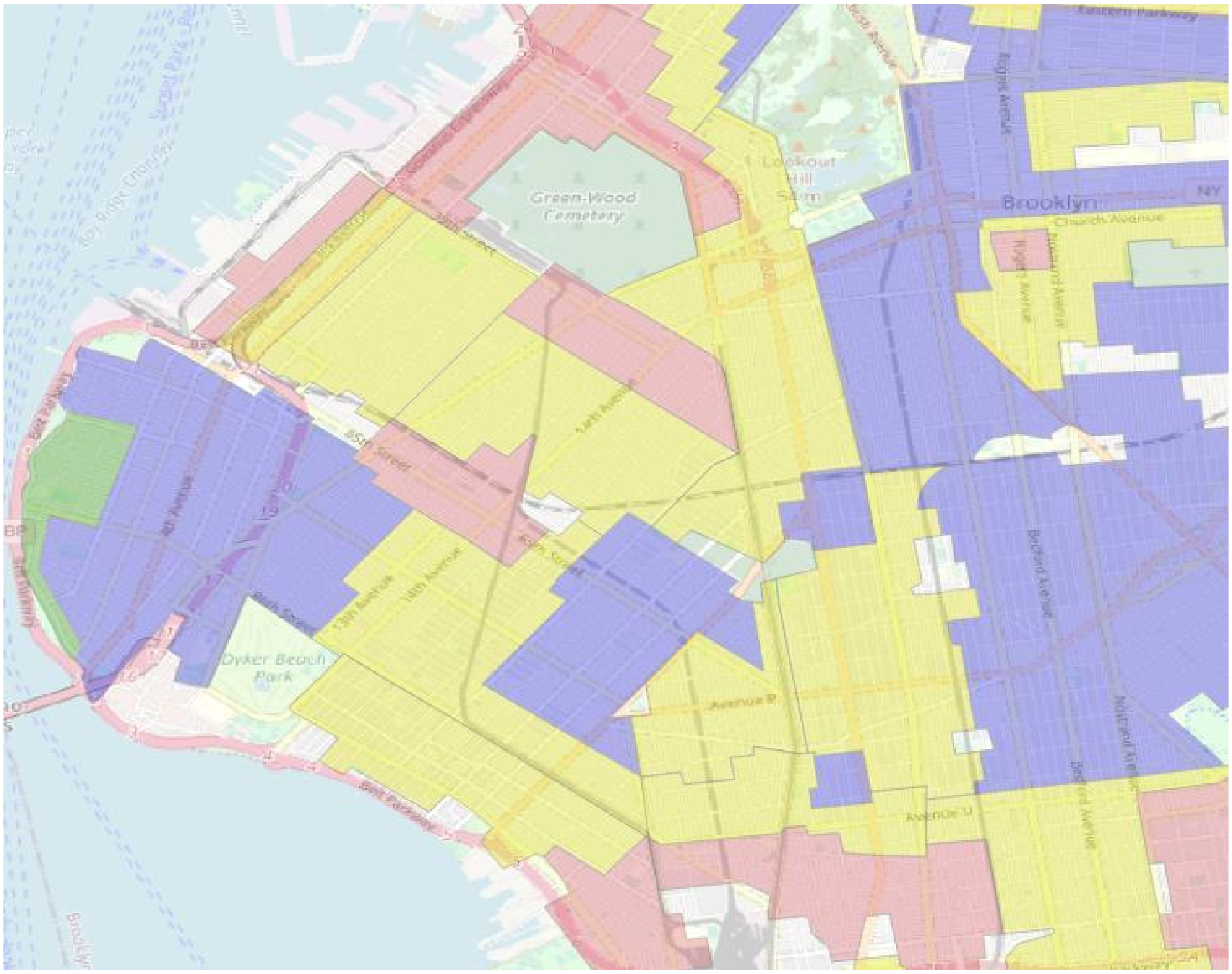
Figure 7: Philadelphia trimmed address virtual enumeration districts overlaid with real enumeration districts

Legend: rounded border darker polygon: one virtual ED constructed from trimmed addresses, straight border lighter polygon: one real ED from the Urban Transition Project.

Figure 8: Brooklyn geolocated 1940 addresses by enumeration district

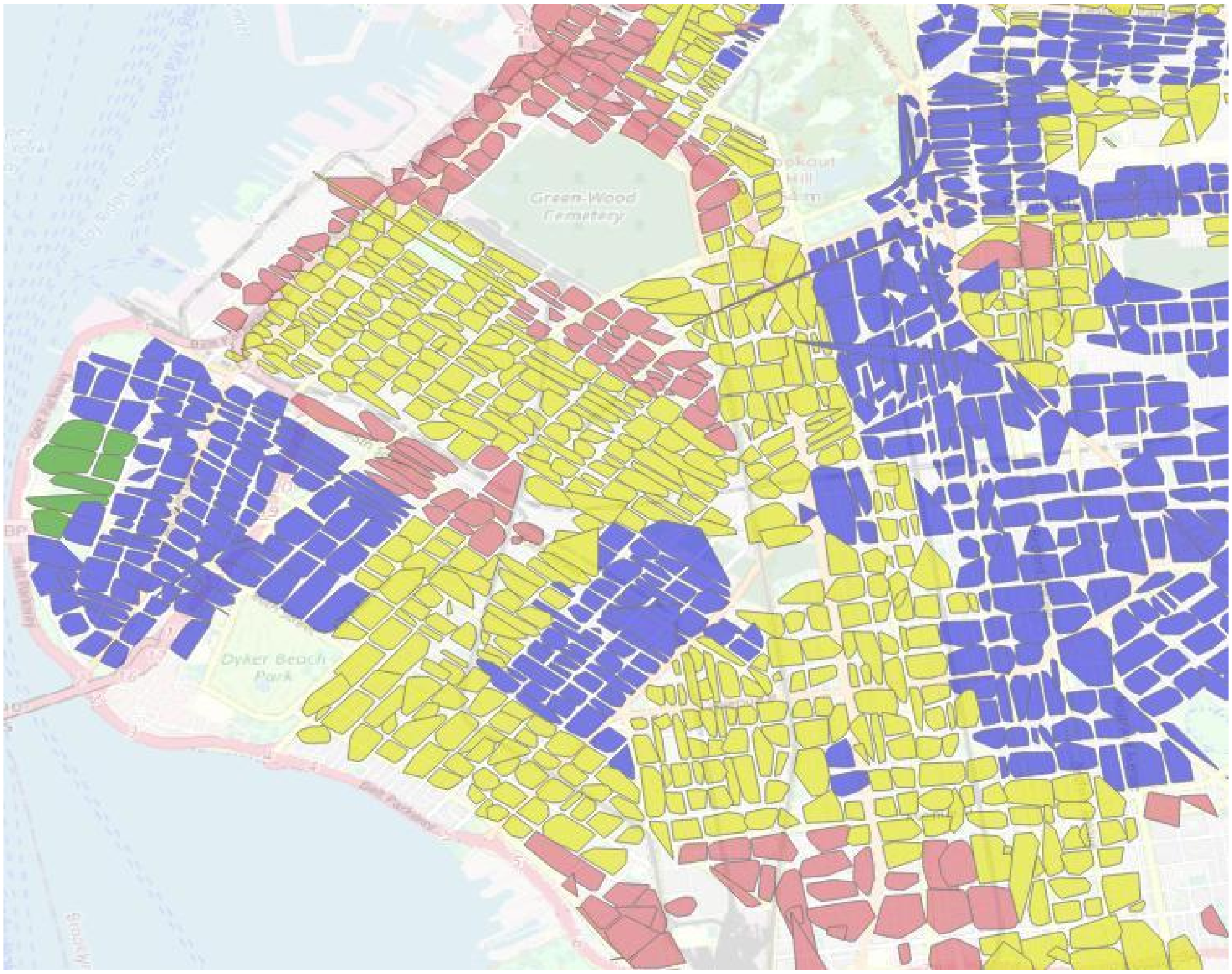
Legend: dot: one 1940 address where color refers to ED, transparent backdrop shading: HOLC map.

Figure 9: Brooklyn 1930 HOLC map



Legend: transparent polygons: HOLC map, red: HOLC D, yellow: HOLC C, blue: HOLC B, green HOLC A.

Figure 10: Brooklyn virtual enumeration districts tagged with HOLC.



Legend: rounded border polygons: virtual enumeration districts where color refers to HOLC category, red: HOLC D, yellow: HOLC C, blue: HOLC B, green HOLC A.

Appendix B: Previous paper

Huang SJ, Sehgal NJ. Association of historic redlining and present-day health in Baltimore. Plos one. 2022 Jan 19;17(1):e0261028. https://doi.org/10.1371/journal.pone.0261028

Association of historic redlining and present-day health in Baltimore

Shuo Jim Huang[1¶*], Neil Jay Sehgal[1¶]

[1]Department of Health Policy and Management, University of Maryland School of Public Health, College Park, Maryland, United States of America

*Corresponding author

E-mail: sjhuang@umd.edu (SJH)

¶ These authors contributed equally to this work.

Abstract

Background

In the 1930s, the Home Owners' Loan Corporation categorized neighborhoods by investment grade along racially discriminatory lines, a process known as redlining. Although other authors have found associations between Home Owners' Loan Corporation categories and current impacts on racial segregation, analysis of current health impacts rarely use these maps.

Objective

To study whether historical redlining in Baltimore is associated with health impacts today.

Approach

Fifty-four present-day planning board-defined community statistical areas are assigned historical Home Owners' Loan Corporation categories by area predominance. Categories are red ("hazardous"), yellow ("definitely declining") with blue/green ("still desirable"/"best") as the reference category. Community statistical area life expectancy is regressed against Home Owners' Loan Corporation category, controlling for median household income and proportion of African American residents.

Conclusion

Red categorization is associated with 4.01 year reduction (95% CI: 1.47, 6.55) and yellow categorization is associated with 5.36 year reduction (95% CI: 3.02, 7.69) in community statistical area life expectancy at baseline. When controlling for median household income and proportion of African American residents, red is associated with 5.23 year reduction (95% CI: 3.49, 6.98) and yellow with 4.93 year reduction (95% CI: 3.22, 6.23). Results add support that historical redlining is associated with health today.

Introduction

Race and its intersection with place matters when it comes to health disparities. Structural racism—the interplay of (sometimes unintentional) processes and structures that reinforce racial inequities [1]—has a large role in continued health disparities [2-5]. Place, geography, and the built environment are understudied dimensions of structural racism in terms of their impact on racial differences in health [6]. These continuing structural contributions to health disparities can often be traced to historically discriminatory policies and practices, such as redlining.

The impact of redlining is part of a broader national conversation about the impact of racially discriminatory policy and systems on disparate racial impacts, particularly differences between Black and white Americans. The 2014 Ta-Nehisi Coates article “The Case for Reparations” in *The Atlantic* [7] brought back to the fore a national conversation about and made an argument for reparative investment to correct inequities suffered by Black Americans. Part of Coates' thesis was that redlining led to systematic racial discrimination in housing throughout the twentieth century, which still impacts intergenerational wealth and life outcomes in Black communities today.

Racially discriminatory local and federal policies and redlining's role were also publicly discussed in the context and aftermath of police murders of unarmed Black civilians. After the late-summer 2014 murder of Michael Brown, Ferguson, MO underwent an uprising in protest at his death and the sense that historical and current local government policies—including redlining—were unjust and led to police suppression of the Black community [8]. In Freddie Gray's murder in Baltimore in April of 2015 and subsequent uprising, another explicit tie-in was made with the city's history of redlining [9-11]. A public link was being made between how redlining creates racial residential segregation [12, 13], and particularly the hypersegregation of Black residents from white residents [14] that concentrated and exposed Black residents both to environmental hazards and overpolicing [15].

Background and literature review

What is Redlining?

In the latter half of the 1930s, the US Federal government authorized the Home Owners’ Loan Corporation (HOLC) to draw “security” maps for investment purposes [16]. These maps were drawn with color codes to reflect areas that were more or less safe for investment. Four color codes were used: green, blue, yellow, and red. Green areas were labeled “A Best”, Blue areas “B Still Desirable”, Yellow areas “C Definitely Declining”, and Red areas “D Hazardous”. Red signified the highest risk areas to HOLC, which led to the term “redlining”. These color-coded categories reflected race and were most likely drawn to discourage investment in majority Black neighborhoods [13]. Redlined areas were explicitly labeled as such due to a large proportion of Black residents, using language characterizing “detrimental influences” such as “infiltration of Negro” while yellow categories often included the tag “infiltration of

Foreigners" [16-19].

Evidence that redlining might still have impacts today

Even time-limited urban changes can have long-lasting cumulative effects. Michaels and Rauch [20] found that the differential collapse of Western Roman urbanization in Britain and France in the 6th century CE differentially impacted the spatial efficiency of urbanization even 1500 years later in the 21st century. Similarly, evidence supports that redlining still has cumulative impacts on various social factors today. Aaronson et al. [12] found that living in close proximity on two sides of differently graded borders—as represented in the 1930s HOLC security maps—is strongly associated nation-wide with increased residential racial segregation from the 1930s to today. The effect on residential racial segregation was particularly strong from 1930 to 1970 and the effect size began to decrease after 1970. Aaronson et al. provide additional support that maps were likely drawn with race in mind: only areas marked category D had a primarily Black population in the 1930s. In addition to increased segregation, Aaronson et al. found support for reductions in home ownership, house values, and credit scores throughout the 20th century that are maintained even today nearly a century later. They also found evidence of "yellow-lining": areas marked as category C also had disparate current outcomes when compared to higher rated areas. Using a different methodology, Appel and Nickerson [21] found that redlined neighborhoods had lower home prices in 1990 compared to surrounding areas, and that these discriminatory effects remained even after nearly 60 years. The presence of these discriminatory effects can be compounded across time: Massey et al. [22] found that Black residents of redlined neighborhoods face greater barriers to residential mobility than white residents that negatively impacts Black residents' social and economic well-being.

Potential impacts of redlining on health

A growing body of scholarship examines the relationship between neighborhood characteristics and health outcomes [23], though few studies have specifically considered the legacy of redlining. The shadows of historical redlining are seen on present day health where structural racism and neighborhood segregation persist. The potential pathways for redlining's impact on health reflect those found in residential segregation [23] and fall across the entirety of the life course [9, 23-40]. Proposed pathways span both differential exposures to environmental and social hazards (including both the proximity and imposition of hazards and the reduced presence of mitigatory systems such as reduced access to health services [23, 27]) and possibly the concentration of populations that have previously suffered such hazards [25, 29]. Lynch et al. [27] find that historical redlining was associated with sustained disinvestment and lending discrimination in Milwaukee neighborhoods, as well as poorer present day physical and mental health. Similarly, neighborhood quality is thought to affect environmental health and reduce access to resources that enable protective behaviors [28-31]. Kreiger et al. [24] found that residing in a previously redlined area in Massachusetts was associated with an elevated risk for late stage at cancer diagnosis, even for residents of census tracts with present-day economic and racial privilege.

Redlining’s impact on health through disinvestment potentially starts before birth and in childhood. O'Campo et al. [29] found that neighborhood risk factors were significant in predicting low-birth weight. Specifically, they found that low-income census tracts (those with less than $8000 per capita annual income) were associated with higher risk of low birthweight. In examining birth certificate data and HOLC grade, Krieger et al. [25] concluded that historical redlining may be a structural determinant of present-day risk of preterm birth in New York City. On the other hand, Mendez et al. [31] found that redlining as represented through current mortgage discrimination was not associated with pregnancy outcomes. However, Mendez et al. stated that there was not enough variation in their sample to estimate associations and hypothesized that redlining might have measurable health impacts in other realms.

Wilson et al. [28] show the pathway from historic redlining to worsened pollution and housing stock and ultimately to poor asthma outcomes. Similarly, Alexander and Currie [30] found that sustained racial segregation in neighborhoods in New Jersey is associated with increased prevalence of asthma in low birthweight children. Nardone et al. [26] report that historically redlined census tracts in California have significantly higher rates of emergency department visits due to asthma.

Neighborhood quality and redlining might have an impact on physical activity as well. Boone et al. [32] find that redlined neighborhoods in Baltimore had reduced availability of greenspaces. The availability of greenspaces, walkability, other neighborhood built environmental factors, and neighborhood income in Baltimore and Seattle were found to be significantly associated with physical activity and better physical quality of life [6], which are factors that are important for the prevention and management of chronic diseases [41].

Potentially negative health impacts can also be a manifestation of direct state subjugation of and lack of political responsiveness to marginalized populations through environmental injustice. McCoy [34] found that coal-fired plants were more likely to be sited in communities of color. Trangenstein et al. [33] find HOLC category best predicts slow responsiveness of liquor store closures to depopulation and store clustering in Baltimore. Pappoe [9] in examining Freddie Gray's case in Baltimore, argues that government-enabled and sponsored housing segregation and concomitant impacts of underfunded education, overpolicing, and deliberate impoverishment is the best frame to understand both structural and interpersonal violence (including state-sponsored violence) against Black people living in Baltimore.

The totality of potential health impacts across the life course may be reflected in life-expectancy [25, 40] and age-differentiated mortality. Richardson et al. find large differences in life expectancy by HOLC grade [40]. To our knowledge there has not been a study yet on associations with age-differentiated mortality.

Methodological issues in measuring redlining

HOLC map categories do not map well onto other, more conventional geographical units. HOLC categories cross census tracts (both contemporaneous and current), as well as larger neighborhood boundaries. This non-cartographic categorization can cause issues in classifying neighborhoods according to HOLC category as well as obtaining representative sets of sociodemographic characteristics of neighborhoods in ecologic studies. Predominance, or the use of the largest category feature to assign a non-cartographic category to a conventional geographic unit, is commonly used across disciplines [42-46]. In studies where the unit of analysis is an individual or their address, these categorization issues are no longer relevant, but such access is almost always subject to restrictions. Studies under those restrictions for individual-level data have sometimes used multilevel models for area-level effects while using addresses only for precision in geographical unit assignment [24, 25]. While smaller sizes of conventional geographical units, such as census tracts or block groups, may offer better resolution than larger units, such as CSAs, they are still susceptible to the ecological fallacy.

HOLC categories also do not necessarily function as “equal steps” away from each other [47]. Some studies have treated HOLC categories as intervals specifically for the purpose of examining potential determinants of those HOLC categories, but have also noted the limitations of taking such an approach

[47]. Although changes in neighborhood demographic composition are by no means exogenous, HOLC maps cannot directly capture the demographic changes that have occurred over the last century nor other federal and local policies [12] that may either augment or reduce the potential demographic and health impacts of redlining in a given area.

Objective

The objective of this paper is to examine whether redlining policies from nearly a century ago are associated with a lingering disparate impact on health outcomes at the community statistical area (CSA) level today. The specific case of Baltimore is examined both because of its place in the national conversation about redlining, and because of its availability of public data on both HOLC security maps and CSA health indicators including life expectancy, maternal and child health outcomes, and age-differentiated mortality (which has not been previously studied).

Materials and Methods

Data Sources and Variables

CSA neighborhood maps were taken from the Baltimore Neighborhood Indicators Alliance (BNIA) [48]. This is a common set of CSA polygons that are each composed of several masked census tracts. These shapefiles include neighborhood characteristics such as median household income uptake and racial composition for each CSA in 2013. They also include various neighborhood health outcomes including 2013 life expectancy at birth (calculated using prevailing rates of death through the life span), teen pregnancy (per 1000 female residents aged 15-19), percentage of births at term, percentage of births at satisfactory birth weight, cases of elevated childhood lead levels per 1000 children 0-6, and mortality per 10,000 deaths in a five-year period at various age categories. Additionally, there are health process outcomes such as the uptake of prenatal care, as well as neighborhood features such as rate of liquor stores and fast-food restaurants. Shapefiles that include this data in a map were downloaded from BNIA.

The redlining map was taken from the Mapping Inequality [16] project on HOLC security maps. HOLC categories are assigned to areas within the current city of Baltimore geography and extend beyond the city of Baltimore's boundaries. Given the age of the maps, not all of present-day Baltimore is categorized. Shapefiles that include this data in a map were downloaded from the University of

Richmond's Digital Scholarship Lab.

Methods

Transformations and overlay

All mapping analysis was done in QGIS 2.18.13. BNIA's CSA shapefiles were projected using WGS84, these were converted into WGS84 pseudo-Mercator projection to be compatible with the Mapping Inequality HOLC shapefiles so that intersect overlays could be performed.

In order to determine which map sections lined up with each other, the intersection was taken of HOLC categories and Baltimore CSAs. The intersection of two fields in two maps returns all sub areas.

In line with methods for determining the classification of minimum mapping areas [42, 43], the largest HOLC category by area within a CSA was then taken to represent the CSA's HOLC categorization. This method is used across several disciplines [44-46]. We do not follow Hillier's [47] usage of interval HOLC categories: they were trying to find determinants of HOLC categories and had noted the problematic limitations with treating HOLC categories as interval rather than ordinal. Aaronson et al. [12] were looking specifically at boundary blocks and tracts, which is beyond the scope of this analysis.

Statistical analysis

All statistical analysis was done in StataMP 16. The main independent variable of interest was HOLC category. Summary statistics were taken of HOLC category, life expectancy, median household income, and proportion of African American residents.

Next, our main dependent variable of life expectancy was regressed with ordinary least squares (OLS) against the main independent variable. Red and yellow lined CSAs were analyzed against a reference category of combined green/blue CSAs (for sample size reasons). A second OLS analysis was run with the control variable of median household income—to account for the independent non-HOLC effects of CSA income on health—and the control variable of proportion of African American residents—to account for other forms of anti-Black structural racism and racial composition change on health [2, 29]. These measures cannot be fully disentangled from redlining, so we test for multicollinearity using variance inflation factor. We define our threshold of statistical significance to be $p < 0.05$.

Various other health outcomes across the life course were also regressed on HOLC category.

These outcomes measures included mortality in BNIA-defined age bins (infancy, 1-14, 15-24, 25-44, 45-64, 65-84, and 85+), teen birthrate, percentage of births at term, percentage of babies born with satisfactory weight, and percentage of children testing positive for elevated blood lead levels. Process measures that were also regressed as dependent variables against HOLC category included percentage of mothers receiving prenatal care. Neighborhood factors that might impact health behaviors were regressed including the per population rate of liquor stores and fast-food restaurants in a CSA.

Finally, we ran a sensitivity analysis by redoing our first OLS analysis while excluding CSAs that have less than 25% HOLC coverage to see if our analysis is sensitive to CSAs that have low HOLC coverage.

Results

Mapping sample

The BNIA shapefiles included 55 CSA polygons for Baltimore (Fig 1, Fig 2). The Mapping Inequality shapefiles for Baltimore included 42 HOLC categorized polygons (Fig 1). A total of 156 polygons were created with an intersect overlay between CSA polygons and HOLC polygons (Fig 3). One Baltimore CSA polygon (#7, Cherry Hill) was dropped from the analysis since there were no intersections with the Mapping Inequality shapefiles (the area was not assigned any HOLC categories in the 1930s). 54 primary polygons represented the remaining CSAs with HOLC categories assigned.

Fig 1: 2013 Baltimore CSA neighborhoods (numbers refer to CSA names) in light purple overlaid on top of HOLC redlining map (red: “hazardous”, yellow: “declining”, blue: “still desirable”, green: “best”).

Fig 2: 2013 Baltimore CSAs by life expectancy. Darker blue is longer life expectancy. Numbers refer to CSA names.

Fig 3: 2013 Baltimore CSAs’ life expectancy overlaid on top of HOLC redlining map.

There were 14 CSAs that were classified as red using the predominant polygon by area. 19 were classified as yellow, 17 as blue, and 4 as green. Red and yellow polygons tended to be at the city core, while green and blue polygons tended to be at the northern peripheries (Fig 1).

Summary statistics

CSA characteristics are fairly heterogeneous in Baltimore. The mean life expectancy at birth in the 54 CSAs was 73.3 years, with a standard deviation of 4.33 years, and ranging from a low of 66.0 to a high of 85.3. Mean of median CSA household income was $44,610 with a low of $13,887 to a high of $106,771. Mean African American proportion of the population was 61.2%, with a low of 2.44% to a high of 97.4% (Table 1).

Table 1: Summary statistics for 2013 Baltimore CSAs

	Life Expectancy-years			Median HH Income-$			African American-%		
n=54	Mean	Std Dev	Range	Mean	Std Dev	Range	Mean	Std Dev	Range
Total	73.3	4.33	66.0, 85.3	44610	20162	13877, 106771	61.2	33.2	2.44, 97.4
HOLC red (n=14)	72.3	4.41	66.0, 78.8	47539	28853	13887, 92130	46.9	37.8	2.44, 96.4
HOLC yellow (n=19)	70.9	2.32	66.8, 76.4	35482	9630	24175, 60104	62.5	31.8	10.7, 97.4
HOLC blue (n=17)	75.9	4.16	71.2, 85.3	49430	19154	31701, 106771	70.1	29.1	6.86, 96.4
HOLC green (n=4)	77.9	3.91	75.4, 83.8	57225	12220	47124, 74277	67.4	34.7	15.9, 90.0

OLS regression

CSAs with red classification were statistically significantly associated at the 95% confidence level with a mean 4.01 years reduction (95% CI: 1.47, 6.55) in CSA life expectancy and CSAs with a yellow classification were associated with a mean 5.36 years reduction (95% CI: 3.02, 7.69) in CSA life expectancy (Table 2). Once CSA proportion of African Americans and CSA median household income were added to the model as controls, the effect was slightly stronger in red CSAs with a mean reduction of 5.23 years (95% CI: 3.49, 6.98) and slightly weaker in yellow CSAs with a mean reduction of 4.93 years (95% CI: 3.22, 6.63). This model accounted for 73% of the variability of the life expectancy response data around the mean. The Variance Inflation Factor of this model was 2.17, so multicollinearity seems less likely to be an issue. In the fully adjusted model, yellow classification is also significantly associated with mortality between the ages of 15-24 (mean 4.65 additional deaths per 10,0000 people), while yellow and red classifications are also significantly associated with mortality between the ages of 25 and 44 (mean 10.54 (yellow) and 6.19 (red) additional deaths per 10,000 people), mortality between the ages of 45 and 64 (mean 30.99 (yellow) and 36.4 (red) additional deaths per 10,000 people), mortality between the ages of 65 and 84 (mean 74.65 (yellow) and 95.47 (red) additional deaths per 10,000 people) (Table 3).

Table 2: Multivariable ordinary least squares analysis of Baltimore HOLC and 2013 life expectancy

Independent variables: n=54	Model 1: life expectancy-years		Model 2: life expectancy-years	
	Effect [95% CI]	P-value	Effect [95% CI]	P-value
HOLC Blue/Green	ref		ref	
HOLC Yellow	**-5.36 [-7.69, -3.02] ***	p<0.001	**-4.93 [-6.63, -3.22] ***	p<0.001
HOLC Red	**-4.01 [-6.55, -1.47] ***	0.003	**-5.23 [-6.98, -3.49] ***	p<0.001
Median HH Income-$1000			**0.06 [0.01, 0.11] ***	0.028
African American-%			**-0.06 [-0.09, -0.03] ***	p<0.001

* Significant at the 95% confidence level. Unit of analysis is the community statistical area. Dependent variables in columns, independent variables and covariates in rows. Mean and 95% CI reported. Model 1: Life expectancy regressed on HOLC yellow and HOLC red vs. HOLC blue/green. Model 2: Added neighborhood median household income and proportion African American as controls. Median household income and percent African American are controls and effects should not be interpreted independently.

Table 3: Multivariable ordinary least squares analysis of Baltimore HOLC and 2013 mortality by age bins

HOLC Blue/Green reference	HOLC Yellow		HOLC Red		Median HH Income-$1000		African American-%	
Dependent variables: n=54	Effect [95% CI]	P-value	Effect [95% CI]	P-value	Effect [95% CI]	P-value	Effect [95% CI]	P-value
mortality up to age 1 deaths in 5 years-per 10000	1.98 [-0.39, 4.35]	0.100	2.01 [-0.42, 4.44]	0.102	-0.01 [-0.08, 0.06]	0.872	**0.11 [0.07, 0.15] ***	**p<0.001**
mortality age 1-14 deaths in 5 years-per	0.13 [-2.16, 2.43]	0.907	1.66 [-0.69, 4.01]	0.163	-0.05 [-0.11, 0.02]	0.177	-0.01 [-0.05, 0.03]	0.647

10000								
mortality age 15-24 deaths in 5 years-per 10000	**4.65 [1.19, 8.1] ***	**0.009**	3.51 [-0.02, 7.05]	0.051	-0.02 [-0.12, 0.09]	0.744	**0.09 [0.03, 0.15] ***	**0.004**
mortality age 25-44 deaths in 5 years-per 10000	**10.54 [6.51, 14.57] ***	**p<0.001**	**6.19 [2.07, 10.32] ***	**0.004**	-0.07 [-0.19, 0.05]	0.266	**0.24 [0.17, 0.31] ***	**p<0.001**
mortality age 45-64 deaths in 5 years-per 10000	**30.99 [18.64, 43.35] ***	**p<0.001**	**36.4 [23.75, 49.06] ***	**p<0.001**	**-0.74 [-1.1, -0.37] ***	**p<0.001**	0.19 [-0.03, 0.4]	0.09
mortality age 65-84 deaths in 5 years-per 10000	**74.65 [32.22, 117.07] ***	**0.001**	**95.47 [52.03, 138.91] ***	**p<0.001**	-0.8 [-2.05, 0.45]	0.205	0.4 [-0.34, 1.14]	0.284
mortality age 85 and up deaths in 5 years-per 10000	136.9 [-4.86, 278.65]	0.058	103.16 [-41.99, 248.31]	0.16	**4.31 [0.13, 8.49] ***	**0.044**	**2.58 [0.1, 5.06] ***	**0.042**

* Significant at the 95% confidence level. Unit of analysis is the community statistical area. Dependent variables in rows, independent variables and covariates in columns. Mean and 95% CI reported. BNIA pre-defined age bin mortality rates are individually regressed on HOLC yellow and HOLC red vs. blue/green, with covariates of neighborhood median household income, proportion of African American residents. Effects are reported as additional deaths in a 5 year period per 10000 residents within an age group for the relevant HOLC grade (yellow or red) as compared to the reference category of HOLC blue/green. Median household income and percent African American are controls and effects should not be interpreted independently.

HOLC classification was not significantly associated with infant mortality, mortality from ages 1-14, or mortality over the age of 85 (Table 3). Red or yellow classification was not significantly associated with the health outcomes of percent of births given at term, percent of births at satisfactory weight, or number of children testing positive for elevated blood lead levels (Table 4). Red and yellow classification was significantly associated with the teen birthrate (additional 23.61 (yellow) and 20.36 (red) births per 1000) (Table 4). The process outcome of percentage of births with prenatal care was not significant for red classification but was for yellow (5.06 percent fewer births) (Table 4). For CSA characteristics, red or yellow classification was associated with a significant increase of 0.87 (yellow) and 2.18 (red) additional liquor stores per 1000 residents but was not significantly associated with the rate of fast-food restaurants

(Table 5).

Table 4: Multivariable ordinary least squares analysis of Baltimore HOLC and 2013 maternal and child health outcomes and processes

HOLC Blue/Green ref	HOLC Yellow		HOLC Red		Median HH Income-$1000		African American-%	
Dependent variables: n=54	Effect [95% CI]	P-value	Effect [95% CI]	P-value	Effect [95% CI]	P-value	Effect [95% CI]	P-value
teen pregnancy per 1000 female residents 15-19	**23.61 [7.97, 39.26] ***	**0.004**	**20.36 [4.34, 36.37] ***	**0.014**	-0.03 [-0.49, 0.43]	0.896	0.18 [-0.09, 0.45]	0.192
term birth-% of births	-1.49 [-3.79, 0.82]	0.201	-0.66 [-3.02, 1.7]	0.575	-0.04 [-0.11, 0.02]	0.198	**-0.11 [-0.15, -0.07] ***	**p<0.001**
satisfactory birth weight-% of births	-1.67 [-3.76, 0.41]	0.113	-0.89 [-3.03, 1.25]	0.406	-0.01 [-0.07, 0.05]	0.795	**-0.09 [-0.13, -0.06] ***	**p<0.001**
Received 1st trimester prenatal care-% of births	**-5.06 [-8.7, -1.43] ***	**0.007**	-1.82 [-5.54, 1.9]	0.331	**0.19 [0.08, 0.3] ***	**0.001**	**-0.14 [-0.21, -0.08] ***	**p<0.001**
age 0-6 elevated blood lead levels-%	0.67 [-0.67, 2]	0.32	0.19 [-1.15, 1.52]	0.78	0 [-0.04, 0.04]	0.987	0.01 [-0.01, 0.04]	0.215

* Significant at the 95% confidence level. Unit of analysis is the community statistical area. Dependent variables in rows, independent variables and covariates in columns. Mean and 95% CI reported. Maternal and child health variables are individually regressed on HOLC yellow and HOLC red vs. blue/green, with covariates of neighborhood median household income, proportion of African American residents. Teen pregnancy effects are reported as additional pregnancies per 1000 female residents for the relevant HOLC grade (yellow or red) as compared to the reference category of HOLC blue/green. All other effects are reported as additional percentage points for the relevant HOLC grade (yellow or red) as compared to the reference category of HOLC blue/green. Median household income and percent African American are controls and effects should not be interpreted independently.

Table 5: Multivariable ordinary least squares analysis of Baltimore HOLC and 2013 neighborhood characteristics

HOLC Blue/Green ref	HOLC Yellow		HOLC Red		Median HH Income-$1000		African American-%	
Dependent variables: n=54	Effect [95% CI]	P-value	Effect [95% CI]	P-value	Effect [95% CI]	P-value	Effect [95% CI]	P-value
liquor outlets-per 1000 residents	**0.87 [0.09, 1.65] ***	**0.03**	**2.18 [1.38, 2.98] ***	**p<0.001**	0.00 [-0.02, 0.02]	0.95	-0.01 [-0.03, 0.001]	0.08
fast food outlets-per 1000 residents	0.07 [-2.43, 2.57]	0.954	2.49 [-0.07, 5.05]	0.056	2.49 [-0.07, 5.05]	0.056	-0.03 [-0.07, 0.02]	0.229

* Significant at the 95% confidence level. Unit of analysis is the community statistical area. Dependent variables in columns, independent variables and covariates in rows. Mean and 95% CI reported. CSA neighborhood factors are individually regressed on HOLC yellow and HOLC red vs. blue/green, with covariates of neighborhood median household income, proportion of African American residents. Effects are reported as additional outlets per 1000 residents for the relevant HOLC grade (yellow or red) as compared to the reference category of HOLC blue/green. Median household income and percent African American are controls and effects should not be interpreted independently.

For our sensitivity analysis, 8 CSAs were excluded for having less than 25% HOLC coverage (CSAs 4, 9, 12, 30, 38, 44, 45, and 50). Five excluded CSAs were categorized as yellow, and the remaining 3 were distributed evenly among red, blue, and green. Rerunning our first model while excluding these CSAs resulted in effect sizes with the same directionality and significance, and the effect size was mildly stronger with yellow CSAs associated with a mean 5.78 years reduction (95% CI: 3.01, 7.69) and red CSAs associated with a mean 4.44 years reduction (95% CI: 1.61, 7.27).

Discussion

Summary

These results lend support to the current health impact of living in areas of Baltimore that were categorized as “hazardous” or “declining” by the HOLC in the 1930s, specifically with regards to life expectancy and age-differentiated mortality. To our knowledge, this is the first study to examine the association of redlining with age-differentiated mortality.

This study adds an additional health dimension of support to broader claims that past redlining still has impacts today. While Aaronson et al. [12] and Appel and Nickerson [21] focus on the current impact of redlining and yellowlining on residential segregation and household values, and Boone et al. [32] focus on the impact of redlining on greenspaces, this study lends support that there are associated ultimate impacts on life expectancy and mortality as well, at least in Baltimore. It also lends support to the hypothesis posed by Mendez et al. [31] that redlining has an impact on health outcomes. However, instead of using modern loan applications as a proxy for historical redlining [31], these results are based on the actual HOLC security maps. More broadly, these results provide support for the argument that neighborhood level characteristics [23, 28-31] have an impact on health characteristics.

In addition to categorization, the HOLC maps themselves provide insight on the bias implicit in grade-assignment and how that historical social bias is associated with present day health effects. The descriptions in the HOLC maps note “detrimental influences” such as a “heavy concentration of foreigners” or characterising a preponderance of African American residents as an “infiltration”. At the time, HOLC grades suggested that whiter neighborhoods were “desirable” and colored neighborhoods “hazardous”. The gravity of these historical assignments is felt even today, as the HOLC maps were both a consequence of existing discriminatory lending practices and a cause of sustained disinvestment in neighborhoods.

Current associations of redlining with life expectancy

Even after controlling for racial composition and median household income, CSAs that were categorized as red or yellow had a 5-year shorter life expectancy than CSAs that had been categorized as green or blue. While the effect is small and we are not testing a full model for racial composition, racial composition as a control continues to have a significant association with life expectancy, which suggests that red/yellow categorization impact on life expectancy is not entirely due to the racial composition of CSAs. These results are consistent with Richardson et al. [40] which also found significant differences in life expectancy between the most and least redlined quartiles of census tracts nationwide.

Care must be taken in saying that the HOLC security maps by themselves caused the decline in life expectancy or increases in mortality at various age groups. Instead, as discussed by others [12, 47], the HOLC security maps are likely a proxy for de jure and de facto government policy throughout the 20th

century that systematically disenfranchised and harmed the largely Black residents of those neighborhoods [9]. These include—but aren't limited to—lending practices, restrictive racial covenants, and political choices for municipal investment not just in the 1930s but throughout the 20th century and into today [47]. However, the initial state of racial discrimination in the policymaking of the 1930s enabled many of these other racist policies and still leaves a mark today. Aaronson et al. [12] found discontinuous city size-dependent differences in residential segregation by exploiting how HOLC maps were used in differently sized cities. This suggests that HOLC security maps may have been a driving force behind broader policies hostile to Black residents.

Maternal and child health outcomes

O'Campo et al. [29] provide strong support for the notion that low birthweight was dependent on macro level neighborhood health factors at the census tract level in Baltimore. Krieger et al. [25] likewise provide evidence for New York City. However, HOLC categories in Baltimore are not associated with low birthweight or term birth. It is possible that HOLC categories at the CSA level are not sensitive enough to variations in birthweight by census tract, or that the sample size is too small to detect such variations. It is also possible that while historical HOLC categories are related to certain neighborhood characteristics (such as the significantly associated increased prevalence of liquor stores), they might not be related to current neighborhood characteristics that are relevant for a safe birthweight. We did find associations with increased teen pregnancy and reduced prenatal care in the first trimester. While we did not assess measures of the quality of maternal care, others have noted that redlining and residential segregation may reduce access to appropriate healthcare services [23, 25, 29].

While we found no association with infant mortality, Lynch et al. [27] noted that current investment patterns in census tracts interact with redlining and population displacement of Black residents by white residents that may drive infant mortality more than historical redlining patterns alone. This may suggest similar patterns for other maternal and child health outcomes. While analysis including the interaction of current investment patterns with redlining is beyond the scope of this paper, it deserves further study in the future.

Age-differentiated increased mortality

The primary driver of increased mortality came from those in the 25-84 age range, especially in the 65-84 age range. There are multiple possible explanations, though the data do not capture cause of death. People living in redlined neighborhoods had less access to green spaces [32] and physical activity that might have an impact on the prevention and management of chronic conditions [6]. Multiple, lifelong encounters with hostile systems and racism can also lead to increased and premature mortality in adults [2, 5].

Liquor stores

CSAs categorized as red had more per capita liquor stores than those categorized as green or blue, while CSAs categorized as yellow had fewer. While our results for red categorized CSAs are congruent with findings from Trangenstein et al. [33], we found the opposite effect for yellow categorized CSAs. However, Trangenstein et al. used census tracts rather than CSAs.

Limitations

The unit of analysis was the CSA, not the individual. As a result, there is a risk of ecological fallacy in these results and analysis. Rather, redlining at the neighborhood level may not be the same construct [49] as an individual's exposure to redlining. Small sample size from using CSAs as the unit of analysis is a study limitation as well, and may especially be an issue with relatively rare outcomes such as elevated blood lead levels in any given year. CSAs are also larger than other potential geographic units of analysis like census tracts: a more geographically concentrated unit of analysis may offer better resolution and cleaner categorization of HOLC categories than CSAs could. Future studies could use restricted-access census tract level mortality data to reduce the impact of these limitations. Life expectancy data was calculated by Baltimore's Health Department using 2013 mortality rates within each age category, and thus may not fully capture "true" life expectancy.

The study scope is limited to Baltimore, which reduces the generalizability of these results. Other redlining maps from across the US are available for use in a future study. However, Baltimore provided a convenience sample that included publicly available mortality and redlining data and has been central to the national conversation around redlining and residential segregation.

The main limitation is the passage of time. While this study adds to the literature that suggests that nearly century-old policies in Baltimore still have measurable and significant impacts on residents' lives today, the time lag masks a great number of social changes that have taken place in the interim.

These include the strong possibility of endogenous and path-dependent [20] changes such as gentrification and concomitant displacement of populations that were living in redlined neighborhoods in the 1930s. Certainly, the demographic makeup of Baltimore has changed drastically even in the last decade [50]. Additionally, the study does not account for the length of time residing in home/residence for Baltimore residents, and potential CSA or city out-migration; if there were mobility across CSAs in Baltimore, this would potentially serve as a mediator of the effect of redlining. These changes to the social and built environment may be either enhancing or blunting the impacts of policies set in the 1930s. We are unable to determine from this analysis whether the measured reduction in life expectancy in redlined CSAs is due to CSAs' differing resource characteristics—as suggested by several other authors [12, 21, 28, 29, 32]—or simply changes in the demographic profile of a CSA through displacement and concentration.

Policy implications

Further study to overcome these limitations is needed to draw more concrete policy lessons. However, potential policy implications exist given that the results of this study found disparate and significant differences in life expectancy. The primary policy implication is that discriminatory public or social policy—whether or not such a policy is intentionally discriminatory—in a realm such as housing can potentially have long-lasting, disparate, and large impacts on health. Additionally, contemporary African American activists and organizations in the 1930s presciently recognized the discriminatory impacts of how HOLC made its policy decisions [16]: members of marginalized and/or impacted communities should have control over every major policy-making process.

Since many of these issues are structural and historical, health interventions that do not focus on disparate health outcomes risk being "weighed down" by structural problems that predispose a population to worse health. If we take health equity and disparities research seriously, we should be examining the possibility of interventions that attempt to tackle some of these structural factors, including the possibility of reparations [7]. What Link and Phelan propose as "fundamental causes" [51] of health disparities—such as a lack of flexible resources of money, knowledge, social connections, and political power—may have even more fundamental causes rooted in power hierarchies such as racial, class, and gender subordination.

Like other studies of redlining, this study also may be relevant to ongoing policy disputes in Baltimore. Multiple current policy issues in Baltimore including mitigation policies against displacement, redevelopment in the city core of formerly redlined neighborhoods particularly by health services provisioners operating theoretically under the aegis of Maryland's hospital Community Benefit policies, and the creation of a new police force to control access to those redeveloped areas, should be backdropped against these nearly century-long policies and their associations with current health.

Future studies

Given these policy questions and some of the study limitations, there are multiple avenues for further study. A future study could use census tract-level mortality data and use census tracts or blocks instead of CSAs as the unit of analysis and could look nationwide beyond Baltimore. It could also use tract or block area data to look at demographic changes [50] to gauge whether or not a neighborhood's community has been displaced.

Multiple authors [12, 21, 47] argue that continued diminished investment in neighborhoods due to redlining might be what leads to lower housing values and lower quality neighborhoods today. Gentrification then might be a countervailing trend that brings in resources that might ameliorate any potential neighborhood-level health impacts of redlining. Adding measures of gentrification over time might be a way to account for how neighborhood health status changes over time. However, gentrification also displaces populations from geographically delineated neighborhoods [50]. Redlining may also have current health and non-health impacts through intergenerational effects: housing and neighborhood quality impacts education and wealth accumulation [7, 9] which in turn impacts the ability to provide for the next generation. In this case, while gentrification might change a neighborhood's health status, it would not be due to improvements to infrastructure or wealth accumulation of the impacted community but rather by their replacement. Adding measures of community composition changes could be a way to account for these changes [50]. Additionally, adding analysis of current investment patterns [27] may help elucidate the ways that redlining interacted with population displacement.

Potential pathways from redlining through sustained disinvestment leading to exposure to hazards have been both theorized [9, 23] and in some cases explored [27, 28]. Future studies could look more closely at neighborhood factors like documented environmental hazards, exposure to policing, healthcare utilization, or mobility patterns to clarify potential causal pathways, including specifically in Baltimore.

Additional topics to explore on neighborhood quality could include the association of HOLC categories with hospital availability and quality, or the response times of EMS. Geographic data on police interactions could also be used to examine the relationship between HOLC categories and where communities are policed. Future studies could also more thoroughly examine childhood health outcomes and factors including asthma. Geocoded social data might also be an avenue for exploration: Chae et al. [52] found a strong geographical association between area racism as the proportion of Google searches containing the “N-word” in 196 designated market areas and Black mortality. Such an analysis could be extended to redlined neighborhoods.

Conclusion

HOLC assigned categories for neighborhood quality in the 1930s are associated with present-day impacts on life expectancy, age-differentiated mortality, and other health outcomes.

Acknowledgements

The authors would like to acknowledge the Baltimore Neighborhood Indicators Alliance and the Mapping Inequality Project for their publicly available data.

References

1. Powell JA. Structural racism: Building upon the insights of John Calmore. NCL Rev. 2007;86:791.

2. Gee GC, Ford CL. Structural racism and health inequities. Du Bois review: social science research on race. 2011;8(1):115–132.

3. Lukachko A, Hatzenbuehler ML, Keyes KM. Structural racism and myocardial infarction in the United States. Social Science & Medicine. 2014 Feb 1;103:42-50.

4. Gee GC, Walsemann KM, Brondolo E. A life course perspective on how racism may be related to health inequities. American Journal of Public Health. 2012;102(5):967–974.

5. Williams DR. Race, socioeconomic status, and health the added effects of racism and discrimination. Annals of the New York Academy of Sciences. 1999;896(1):173–188.

6. Sallis JF, Saelens BE, Frank LD, Conway TL, Slymen DJ, Cain KL, et al. Neighborhood built environment and income: examining multiple health outcomes. Social science & medicine. 2009;68(7):1285–1293.

7. Coates T-N. The Case for Reparations. The Atlantic [Internet]. 2014 Jun [cited 2017 Nov 6]; Available from: https://www.theatlantic.com/magazine/archive/2014/06/the-case-for-reparations/361631/

8. Coates T-N. The Racist Housing Policies That Built Ferguson. The Atlantic [Internet]. 2014 Oct 17 [cited 2017 Dec 16]; Available from: https://www.theatlantic.com/business/archive/2014/10/the-racist-housing-policies-that-built-ferguson/381595/

9. Pappoe YN. Remedying the Effects of Government-Sanctioned Segregation in a Post-Freddie Gray Baltimore. U Md LJ Race, Religion, Gender & Class. 2016;16:115.

10. Rosenblatt P, DeLuca S. What happened in Sandtown-Winchester? Understanding the impacts of a comprehensive community initiative. Urban Affairs Review. 2017 May;53(3):463-94.

11. Barry-Jester AM. Baltimore's Toxic Legacy Of Lead Paint [Internet]. FiveThirtyEight. 2015 [cited 2017 Dec 16]. Available from: https://fivethirtyeight.com/features/baltimores-toxic-legacy-of-lead-paint/

12. Aaronson D, Hartley DA, Mazumder B. The Effects of the 1930s HOLC 'Redlining' Maps. 2017;

13. Rothstein R. The color of law: A forgotten history of how our government segregated America. Liveright Publishing; 2017 May 2.

14. Massey DS. The legacy of the 1968 fair housing act. InSociological Forum 2015 Jun (Vol. 30, pp. 571-588).

15. Thomas ME, Moye R, Henderson L, Horton HD. Separate and unequal: The impact of socioeconomic status, segregation, and the Great Recession on racial disparities in housing values. Sociology of Race and Ethnicity. 2018 Apr;4(2):229-44.

16. Nelson RK, Winling L, Marciano R, Connolly N, et al. Mapping Inequality. n.d. [cited 2017 Nov 6]. In: Nelson RK, Ayers EL, eds. American Panorama [Internet]. Richmond: University of Richmond Digital Scholarship Lab. Available from: https://dsl.richmond.edu/panorama/redlining/#loc=11/39.293/-76.808&city=baltimore-md

17. Hillier AE. Redlining and the home owners' loan corporation. Journal of Urban History.

2003;29(4):394–420.

18. Woods LL. The Federal Home Loan Bank Board, redlining, and the national proliferation of racial lending discrimination, 1921–1950. Journal of Urban History. 2012 Nov;38(6):1036-59.

19. Jackson KT. Race, ethnicity, and real estate appraisal: The home owners loan corporation and the federal housing administration. Journal of Urban History. 1980;6(4):419–452.

20. Michaels G, Rauch F. Resetting the Urban Network: 117-2012. The Economic Journal. 2016;

21. Appel I, Nickerson J. Pockets of Poverty: The Long-Term Effects of Redlining [Internet]. Rochester, NY: Social Science Research Network; 2016 Oct [cited 2017 Dec 16]. Report No.: ID 2852856. Available from: https://papers.ssrn.com/abstract=2852856

22. Massey DS, Condran GA, Denton NA. The effect of residential segregation on black social and economic well-being. Social Forces. 1987;66(1):29–56.

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

24. Krieger N, Wright E, Chen JT, Waterman PD, Huntley ER, Arcaya M. Cancer stage at diagnosis, historical redlining, and current neighborhood characteristics: Breast, cervical, lung, and colorectal cancers, Massachusetts, 2001–2015. American journal of epidemiology. 2020 Oct;189(10):1065-75.

25. Krieger N, Van Wye G, Huynh M, Waterman PD, Maduro G, Li W, Gwynn RC, Barbot O, Bassett MT. Structural racism, historical redlining, and risk of preterm birth in New York City, 2013–2017. American journal of public health. 2020 Jul;110(7):1046-53.

26. Nardone A, Casey JA, Morello-Frosch R, Mujahid M, Balmes JR, Thakur N. Associations between historical residential redlining and current age-adjusted rates of emergency department visits due to asthma across eight cities in California: an ecological study. The Lancet Planetary Health. 2020 Jan 1;4(1):e24-31.

27. Lynch EE, Malcoe LH, Laurent SE, Richardson J, Mitchell BC, Meier HC. The legacy of structural racism: Associations between historic redlining, current mortgage lending, and health. SSM-population health. 2021 Jun 1;14:100793.

28. Wilson B, Bowles B, Eaton P, Wilson N, Brown N. Housing, health and history: Interdisciplinary spatial analysis in pursuit of equity for future generations. Intergenerational responsibility in the 21st century.

2018 Jan 30.

29. O'Campo P, Xue X, Wang M-C, Caughy M. Neighborhood risk factors for low birthweight in Baltimore: a multilevel analysis. American journal of public health. 1997;87(7):1113–1118.

30. Alexander D, Currie J. Is it who you are or where you live? Residential segregation and racial gaps in childhood asthma. Journal of health economics. 2017 Sep 1;55:186-200.

31. Mendez DD, Hogan VK, Culhane J. Institutional racism and pregnancy health: using Home Mortgage Disclosure Act data to develop an index for mortgage discrimination at the community level. Public Health Reports. 2011;126(3_suppl):102–114.

32. Boone CG, Buckley GL, Grove JM, Sister C. Parks and People: An Environmental Justice Inquiry in Baltimore, Maryland. Annals of the Association of American Geographers. 2009 Sep 17;99(4):767–87.

33. Trangenstein PJ, Gray C, Rossheim ME, Sadler R, Jernigan DH. Alcohol outlet clusters and population disparities. Journal of urban health. 2020 Feb;97(1):123-36.

34. McCoy E. Which Came First, Coal-Fired Power Plants or Communities of Color? Assessing the Disparate Siting Hypothesis of Environmental Justice. 2017.

35. Bailey ZD, Krieger N, Agénor M, Graves J, Linos N, Bassett MT. Structural racism and health inequities in the USA: evidence and interventions. The Lancet. 2017 Apr 8;389(10077):1453-63.

36. Jacoby SF, Dong B, Beard JH, Wiebe DJ, Morrison CN. The enduring impact of historical and structural racism on urban violence in Philadelphia. Social Science & Medicine. 2018 Feb 1;199:87-95.

37. Namin S, Xu W, Zhou Y, Beyer K. The legacy of the Home Owners' Loan Corporation and the political ecology of urban trees and air pollution in the United States. Social Science & Medicine. 2020 Feb 1;246:112758.

38. Hoffman JS, Shandas V, Pendleton N. The effects of historical housing policies on resident exposure to intra-urban heat: a study of 108 US urban areas. Climate. 2020 Jan;8(1):12.

39. McClure E, Feinstein L, Cordoba E, Douglas C, Emch M, Robinson W, Galea S, Aiello AE. The legacy of redlining in the effect of foreclosures on Detroit residents' self-rated health. Health & place. 2019 Jan 1;55:9-19.

40. Richardson J, Mitchell BC, Meier HC, Lynch EE, Edlebi J. The lasting impact of historic" redlining" on neighborhood health: Higher prevalence of COVID-19 risk factors.

41. Dietz WH, Douglas CE, Brownson RC. Chronic disease prevention: tobacco avoidance, physical activity, and nutrition for a healthy start. Jama. 2016;316(16):1645–1646.

42. Lindgren D. Land use planning and remote sensing. Springer Science & Business Media; 2013. 180 p.

43. Schneider A, Friedl MA, Potere D. A new map of global urban extent from MODIS satellite data. Environmental Research Letters. 2009;4(4):044003.

44. itd. MassGIS Data - Land Use (1951-1999) [Internet]. Administration and Finance. 2012 [cited 2017 Dec 15]. Available from: http://www.mass.gov/anf/research-and-tech/it-serv-and-support/application-serv/office-of-geographic-information-massgis/datalayers/lus.html

45. Panhalkar S. Land capability classification for integrated watershed development by applying remote sensing and gis techniques. Journal of Agricultural and Biological Science. 2011;6(4):46–55.

46. Saura S. Effects of minimum mapping unit on land cover data spatial configuration and composition. International Journal of Remote Sensing. 2002 Jan 1;23(22):4853–80.

47. Hillier AE. Residential Security Maps and Neighborhood Appraisals. Social Science History. 2005;29(2):207–233.

48. Vital Signs 13 Children and Family Health | Baltimore Neighborhood Indicators Alliance [Internet]. [cited 2017 Dec 16]. Available from: https://data-bniajfi.opendata.arcgis.com/datasets/vital-signs-13-children-and-family-health?geometry=-77.06%2C39.192%2C-76.192%2C39.378&selectedAttribute=LifeExp13

49. Schwartz S. The fallacy of the ecological fallacy: the potential misuse of a concept and the consequences. American journal of public health. 1994 May;84(5):819-24.

50. Stancil W. American Neighborhood Change in the 21st Century. Institute on Metropolitan Opportunity, University of Minnesota Law School. 2019 Apr. Available from: https://www.law.umn.edu/sites/law.umn.edu/files/metro-files/american_neighborhood_change_in_the_21st_century_-_full_report_-_4-1-2019.pdf

51. Link BG, Phelan J. Social conditions as fundamental causes of disease. Journal of health and social behavior. 1995 Jan 1:80-94.

52. Chae DH, Clouston S, Hatzenbuehler ML, Kramer MR, Cooper HL, Wilson SM, et al. Association between an internet-based measure of area racism and black mortality. PloS one. 2015;10(4):e0122963.

References

- Aaronson D, Hartley DA, Mazumder B. The Effects of the 1930s HOLC 'Redlining' Maps. 2017.
- Aday, L. A., & Andersen, R. (1974). A framework for the study of access to medical care. Health services research, 9(3), 208.
- Alexander D, Currie J. Is it who you are or where you live? Residential segregation and racial gaps in childhood asthma. Journal of health economics. 2017 Sep 1;55:186–200. pmid:28802746
- Allin S, Masseria C, Mossialos E. Measuring socioeconomic differences in use of health care services by wealth versus by income. American journal of public health. 2009;99(10):1849–55.
- Amuchie N. The forgotten victims how racialized gender stereotypes lead to police violence against black women and girls: Incorporating an analysis of police violence into feminist jurisprudence and community activism. Seattle J Soc Just. 2015;14:617.
- Appel I, Nickerson J. Pockets of Poverty: The Long-Term Effects of Redlining [Internet]. Rochester, NY: Social Science Research Network; 2016 Oct [cited 2017 Dec 16]. Report No.: ID 2852856. Available from: https://papers.ssrn.com/abstract=2852856.
- Asher L. The relationship between historical redlining and Census Bureau Community Resilience Estimates in Columbus, Ohio. Environ Plan A. 2021 Nov 1;53(8):1859–61.
- Bailey ZD, Krieger N, Agénor M, Graves J, Linos N, Bassett MT. Structural racism and health inequities in the USA: evidence and interventions. The Lancet. 2017 Apr 8;389(10077):1453–63. pmid:28402827
- Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of personality and social psychology. 1986;51(6):1173.
- Beenackers MA, Groeniger JO, Kamphuis CB, Van Lenthe FJ. Urban population density and mortality in a compact Dutch city: 23-year follow-up of the Dutch GLOBE study. Health & Place. 2018;53:79–85.
- Bowling A. Measuring Health. McGraw-Hill Education (UK); 2004. 226 p.
- Bugliari D, CARROLL J, Hayden O, Hayes J, HURD MD, LEE S, et al. RAND HRS Longitudinal File 2020 (V1) Documentation. Aging. 2023;
- Cain KC, Harlow SD, Little RJ, Nan B, Yosef M, Taffe JR, et al. Bias Due to Left Truncation and Left Censoring in Longitudinal Studies of Developmental and Disease Processes. Am J Epidemiol. 2011 May 1;173(9):1078–84.
- Callahan D. The WHO definition of'health'. Hastings Center Studies. 1973;77–87.
- Cox OC. Caste, class, & race: A study in social dynamics. New York, Monthly Review P; 1948.
- cpython/Lib/asyncio at 3.11 · python/cpython [Internet]. GitHub. [cited 2023 Jun 26]. Available from: https://github.com/python/cpython/tree/3.11/Lib/asyncio/
- Factor RM, Waldron I. Contemporary population densities and human health. Nature. 1973;243:381–4.
- Felitti VJ, Anda RF, Nordenberg D, Williamson DF, Spitz AM, Edwards V, et al. Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: The Adverse Childhood Experiences (ACE) Study. American journal of preventive medicine. 1998;14(4):245–58.
- Ferguson, S. J. (2013). Race, gender, sexuality, and social class: Dimensions of inequality. Sage.
- Fisher GG, Ryan LH. Overview of the health and retirement study and introduction to the special issue. Work, aging and retirement. 2018;4(1):1–9.
- Freedman JL, Heshka S, Levy A. Population density and pathology: Is there a relationship? Journal of Experimental Social Psychology. 1975;11(6):539–52.
- Gee, G. C., Walsemann, K. M., & Brondolo, E. (2012). A life course perspective on how racism may be related to health inequities. American journal of public health, 102(5), 967-974.
- Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020;585:357–62.
- Hillier AE. Redlining and the home owners' loan corporation. Journal of Urban History. 2003;29(4):394–420.

- Hillier AE. Residential Security Maps and Neighborhood Appraisals. Social Science History. 2005;29(2):207–233.
- Huang SJ, Sehgal NJ. Association of historic redlining and present-day health in Baltimore. Plos one. 2022 Jan 19;17(1):e0261028.
- Huber M, Knottnerus JA, Green L, Van Der Horst H, Jadad AR, Kromhout D, et al. How should we define health? Bmj. 2011;343.
- Huggins JC. A Cartographic Perspective on the Correlation Between Redlining and Public Health in Austin, Texas–1951. Cityscape. 2017;19(2):267–80.
- Hurst R. The international disability rights movement and the ICF. Disability and rehabilitation. 2003;25(11–12):572–6.
- Jacoby SF, Dong B, Beard JH, Wiebe DJ, Morrison CN. The enduring impact of historical and structural racism on urban violence in Philadelphia. Social Science & Medicine. 2018 Feb 1;199:87–95. pmid:28579093
- Jones CP. Levels of racism: a theoretic framework and a gardener's tale. American journal of public health. 2000 Aug;90(8):1212.
- Jordahl K. GeoPandas: Python tools for geographic data. URL: https://github com/geopandas/geopandas. 2014;
- Kaplan RM, Milstein A. Contributions of health care to longevity: a review of 4 estimation methods. The Annals of Family Medicine. 2019;17(3):267–72.
- Krieger N, Van Wye G, Huynh M, Waterman PD, Maduro G, Li W, Gwynn RC, Barbot O, Bassett MT. Structural racism, historical redlining, and risk of preterm birth in New York City, 2013–2017. American journal of public health. 2020 Jul;110(7):1046–53. pmid:32437270
- Krieger N, Wright E, Chen JT, Waterman PD, Huntley ER, Arcaya M. Cancer stage at diagnosis, historical redlining, and current neighborhood characteristics: Breast, cervical, lung, and colorectal cancers, Massachusetts, 2001–2015. American journal of epidemiology. 2020 Oct;189(10):1065–75. pmid:32219369
- Krimmel J. Persistence of Prejudice: Estimating the Long Term Effects of Redlining [Internet]. SocArXiv; 2018 Mar [cited 2023 Jun 23]. Available from: https://osf.io/jdmq9
- Linde S, Walker RJ, Campbell JA, Egede LE. Historic Residential Redlining and Present-Day Social Determinants of Health, Home Evictions, and Food Insecurity within US Neighborhoods. Journal of General Internal Medicine. 2023;1–8.
- Lindgren D. Land use planning and remote sensing. Springer Science & Business Media; 2013. 180 p.
- Link BG, Phelan J. Social conditions as fundamental causes of disease. Journal of health and social behavior. 1995;80–94.
- Logan JR, Bellman B, Minca E. From Side Street to Ghetto: Understanding the Rising Levels and Changing Spatial Pattern of Segregation, 1900–1940. Working Paper. Retrieved September 14, 2022. https://s4. ad. brown. edu …; 2020.
- Logan JR, Zhang W. Developing GIS maps for US cities in 1930 and 1940. In: Gregory I, DeBats D, Lafreniere D, editors. The Routledge Companion to Spatial History [Internet]. 1st ed. Milton Park, Abingdon, Oxon; New York, NY: Routledge, 2018.: Routledge; 2018 [cited 2023 Mar 24]. p. 229–49. Available from: https://www.taylorfrancis.com/books/9781351584142/chapters/10.4324/9781315099781-15
- Lynch EE, Malcoe LH, Laurent SE, Richardson J, Mitchell BC, Meier HC. The legacy of structural racism: Associations between historic redlining, current mortgage lending, and health. SSM-population health. 2021 Jun 1;14:100793. pmid:33997243
- Manor O, Matthews S, Power C. Dichotomous or categorical response? Analysing self-rated health and lifetime social class. International journal of epidemiology. 2000;29(1):149–57.
- Mayeda ER, Banack HR, Bibbins-Domingo K, Al Hazzouri AZ, Marden JR, Whitmer RA, et al. Can survival bias explain the age attenuation of racial inequalities in stroke incidence? A simulation study. Epidemiology. 2018 Jul;29(4):525–32.
- McClure E, Feinstein L, Cordoba E, Douglas C, Emch M, Robinson W, Galea S, Aiello AE. The legacy of redlining in the effect of foreclosures on Detroit residents' self-rated health. Health & place. 2019 Jan 1;55:9-19.
- McCoy E. Which Came First, Coal-Fired Power Plants or Communities of Color? Assessing the Disparate Siting Hypothesis of Environmental Justice. 2017.

- McKinney W, others. Data structures for statistical computing in python. In: Proceedings of the 9th Python in Science Conference. 2010. p. 51–6.
- McMaster RB, Lindberg M, Van Riper D. The national historical geographic information system (NHGIS). In: Proc International Cartographic Conference (ICC). 2003. p. 88–102.
- Mendez DD, Hogan VK, Culhane J. Institutional racism and pregnancy health: using Home Mortgage Disclosure Act data to develop an index for mortgage discrimination at the community level. Public Health Reports. 2011;126(3_suppl):102–114.
- Michaels G, Rauch F. Resetting the Urban Network: 117-2012. The Economic Journal. 2016
- Michney TM. How the City Survey's Redlining Maps Were Made: A Closer Look at HOLC's Mortgagee Rehabilitation Division. Journal of Planning History. 2022 Nov 1;21(4):316–44.
- Morse S. One-Step Web Pages: A Potpouri of Genealogical Search Tools. [National Genealogical Society]]; 2007.
- Mossey JM, Shapiro E. Self-rated health: a predictor of mortality among the elderly. American journal of public health. 1982;72(8):800–8.
- Namin S, Xu W, Zhou Y, Beyer K. The legacy of the Home Owners' Loan Corporation and the political ecology of urban trees and air pollution in the United States. Social Science & Medicine. 2020 Feb 1;246:112758. pmid:31884239
- Nardone A, Casey JA, Morello-Frosch R, Mujahid M, Balmes JR, Thakur N. Associations between historical residential redlining and current age-adjusted rates of emergency department visits due to asthma across eight cities in California: an ecological study. The Lancet Planetary Health. 2020 Jan 1;4(1):e24–31. pmid:31999951
- Nearest neighbor analysis with large datasets [Internet]. [cited 2023 Jun 26]. Available from: https://autogis-site.readthedocs.io/en/2019/notebooks/L3/nearest-neighbor-faster.html
- Nelson RK, Winling L, Marciano R, Connolly N, et al. Mapping Inequality. n.d. [cited 2017 Nov 6]. In: Nelson RK, Ayers EL, eds. American Panorama [Internet]. Richmond: University of Richmond Digital Scholarship Lab. Available from: https://dsl.richmond.edu/panorama/redlining/#loc=11/39.293/-76.808&city=...
- Nyman JA. Is 'Moral Hazard' Inefficient? The Policy Implications Of A New Theory. Health Affairs. 2004 Sep;23(5):194–9.
- O'Campo P, Xue X, Wang M-C, Caughy M. Neighborhood risk factors for low birthweight in Baltimore: a multilevel analysis. American journal of public health. 1997;87(7):1113–1118. pmid:9240099
- OpenCage - Easy, Open, Worldwide, Affordable Geocoding and Geosearch [Internet]. [cited 2023 Jun 26]. Available from: https://opencagedata.com/
- OpenCage GmbH [Internet]. [cited 2023 Jun 26]. Available from: https://github.com/opencagedata
- Organization WH. Global health risks: mortality and burden of disease attributable to selected major risks. World Health Organization; 2009.
- Pappoe YN. Remedying the Effects of Government-Sanctioned Segregation in a Post-Freddie Gray Baltimore. U Md LJ Race, Religion, Gender & Class. 2016;16:115.
- Pastor M, Terriquez V, Lin M. How Community Organizing Promotes Health Equity, And How Health Equity Affects Organizing. Health Affairs. 2018 Mar;37(3):358–63.
- Pedregosa F, Varoquaux, Ga"el, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. Journal of machine learning research. 2011;12(Oct):2825–30.
- Personal Correspondence. PLOS paper on redlining in Baltimore. 2022.
- Population of Cities of 25,000 or More [Internet]. Washington, D.C.: US Department of Commerce: Bureau of the Census; 1950 Dec [cited 2023 Jun 26] p. 1–5. (1950 Census of Population Preliminary Counts). Report No.: Series PC-3, No. 6. Available from: https://www2.census.gov/library/publications/decennial/1950/pc-03/pc-3-06.pdf
- Poulson MR, Kenzik KM, Singh S, Pavesi F, Steiling K, Litle VR, et al. Redlining, structural racism, and lung cancer screening disparities. The Journal of thoracic and cardiovascular surgery. 2022;163(6):1920–30.
- Richardson J, Mitchell BC, Meier HC, Lynch EE, Edlebi J. The lasting impact of historic" redlining" on neighborhood health: Higher prevalence of COVID-19 risk factors. 2020.
- Rothstein R. The color of law: A forgotten history of how our government segregated America. Liveright Publishing; 2017 May 2.

- Santucci L. Documenting Racially Restrictive Covenants in 20th Century Philadelphia. Cityscape. 2020;22(3):241–68.
- Sinclair S. Improved health state descriptions will not benefit disabled patients under QALY-based assessment. Journal of Medical Ethics. 2018;44(11):797–8.
- Smith MD, Drummond M, Brixner D. Moving the QALY forward: rationale for change. Value in Health. 2009;12:S1–4.
- Stancil W. American Neighborhood Change in the 21st Century. Institute on Metropolitan Opportunity, University of Minnesota Law School. 2019 Apr. Available from: https://www.law.umn.edu/sites/law.umn.edu/files/metro-files/american_nei...
- Steele AJ. Enacting new worlds of gender: Nonbinary speakers, racialized gender, and anti-colonialism. 2018;
- Suissa S. Immortal time bias in pharmacoepidemiology. American journal of epidemiology. 2008;167(4):492–9.
- Swope CB, Hernández D, Cushing LJ. The relationship of historical redlining with present-day neighborhood environmental and health outcomes: a scoping review and conceptual model. Journal of Urban Health. 2022;99(6):959–83.
- Sy KTL, White LF, Nichols BE. Population density and basic reproductive number of COVID-19 across United States counties. PloS one. 2021;16(4):e0249271.
- Trangenstein PJ, Gray C, Rossheim ME, Sadler R, Jernigan DH. Alcohol outlet clusters and population disparities. Journal of urban health. 2020 Feb;97(1):123–36. pmid:31264024
- Tuljapurkar S. The final inequality: variance in age at death. In: Demography and the Economy. University of Chicago Press; 2010. p. 209–21.
- Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, Van Der Walt SJ. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature methods. 2020 Mar;17(3):261-72.
- Warren JR, Pfeffer F, Helgertz J, Xu D. HRS Documentation Report. 2020;
- White AG, Guikema SD, Logan TM. Urban population characteristics and their correlation with historic discriminatory housing practices. Applied Geography. 2021;132:102445.
- White K, Haas JS, Williams DR. Elucidating the role of place in health care disparities: the example of racial/ethnic residential segregation. Health services research. 2012 Jun;47(3pt2):1278–99. pmid:22515933
- Williams DR. Race, socioeconomic status, and health the added effects of racism and discrimination. Annals of the New York Academy of Sciences. 1999;896(1):173–188. pmid:10681897
- Williams, D. R., & Jackson, P. B. (2005). Social sources of racial disparities in health. Health affairs, 24(2), 325-334.
- Wilson B, Bowles B, Eaton P, Wilson N, Brown N. Housing, health and history: Interdisciplinary spatial analysis in pursuit of equity for future generations. Intergenerational responsibility in the 21st century. 2018 Jan 30.