Essays On Health Economics

Asal Pilehvari

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Economics and Science

Wen You, Co-chair
Xu Lin, Co-chair
Sudipta Sarangi
Susan Chen
Melinda Miller

January 22, 2021
Blacksburg, Virginia

Keywords: Retirement, Social Network, COVID-19 Pandemic, Social Distancing, Racial Disparity

Essays On Health Economics

Asal Pilehvari

(ABSTRACT)

This dissertation consists of three essays in Health Economics relating to the recent challenges in the U.S. The first essay studies the impact of retirement on subsequent health and investigates the mediation effect of social network in the relationship between retirement and health. Findings reveal that retirement adversely impacts physical and mental health outcomes and a considerable portion of these effects are explained by social network changes post-retirement. In particular, shrinkage in the size of social network post-retirement deteriorates physical health and increases depression in retirees.

In the second essay, we assess the differential effect of social distancing on the daily growth rate of COVID-19 infections in the US counties by considering the spatial pattern of COVID-19 spread. We also conduct a comparative analysis of the effect on urban versus rural counties, as well as low versus high socially vulnerable counties. Our analysis illustrates that a high level of social distancing compliance is needed in urban counties and in socially vulnerable areas to achieve the largest impact at curve flattening, whereas moderate-compliance is enough in reaching the peak marginal impact in rural regions and counties with low social vulnerability.

In the third essay, by combining multiple data sources, we investigate how racial disparities in access to healthcare contribute to the disparity in COVID-19 infections and mortality in

black versus white sub-groups. The multilevel analysis demonstrates that a higher probability of having health insurance significantly reduces disparity in COVID-19 mortality in black sub-group while it has no impact on the disparity in whites.

Essays On Health Economics

Asal Pilehvari

(GENERAL AUDIENCE ABSTRACT)

This dissertation uses various quantitative methods to investigate policy-relevant questions regarding the recent challenges in the U.S. economy. In the first chapter, we explore how the physical and mental health of individuals changes by retirement. The results show that retirement decreases physical health while increases depression and anxiety. We also analyze how social network changes after retirement might cause changes in the health of retirees. We find that retirees may experience worse physical and mental health than non-retirees due to losing some of their relationships after retirement. In particular, the loss of contacts increases depression and deteriorates general health.

In the second chapter, we investigate how compliance with social distancing within a typical county and its neighbor counties can reduce the spread of COVID-19. We examine this question for urban versus rural counties in the US and socially vulnerable versus socially not vulnerable counties. We find a high compliance level of social distancing is needed in urban counties and in socially vulnerable areas to reach the highest impact at slowing down the COVID-19 virus spread.

In the third chapter, we examine whether healthcare access inequalities (e.g., having health insurance) increase the risk of COVID-19 infections and mortality for black communities. Our results show that having health insurance decreases COVID-19 mortality in communities of color but not whites.

Dedication

To my parents

Without whom none of these would be possible

Acknowledgments

I would like to thank my advisors, Dr. Wen You and Dr. Xu Lin, for their invaluable support during the past several years. To accomplish my Ph.D. journey, I am deeply indebted to Dr. Wen You for her exceptional mentorship, support, professionalism, patience, and generosity. She went above and beyond to help me achieve my goals. She taught me not only professionalism but also personal improvements. On top of all the priceless training, she taught me how to think critically and approach systematically to answer a research question and become an independent researcher.

I am also extremely grateful to my co-advisor, Dr. Xu Lin, for her outstanding mentorship, insightful guidance, kindness, and support throughout my whole journey. I would like to express my deepest appreciation to my committee members for their guidance and look forward to further work with them in the future.

I also like to extend my gratitude to especially thank Dr. Sudipta Sarangi-Department head and my committee member- because I have benefited tremendously from his advice on my academic development and career goal setting.

My sincere gratitude goes to Dr. Achla Marathe for the opportunity to work with her and get invaluable advice on my work while I was an intern in the Biocomplexity Institute at the University of Virginia. I would like to also thank her colleagues in Biocomplexity who helped and guided me in my research, John Krulick, Professor Jiangzhuo Chen, and Dr. Srinivasan Venkatramanan. Many thanks to the faculty and staff in the Department of Economics at Virginia Tech, and my classmates.

I am thankful to my parents for their unconditional love and supports that encourages me to move forward and never stop. I would like to thank my brothers, Amir and Ahmad, for their never-ending support and faith in my capabilities that inspired me to work hard. Many thanks to my uncle-Mehdi- and his wife -Leila- for always supporting me to reach my goals. Finally, I could not have completed this journey without the support of my friends in Blacksburg who gave me the feeling of having a second family in the years away from home; especially thanks to Noushin Omidvar, Masoumeh Riazi, Sona Najafi, Parshin Shojaei, and Nazanin Sedaghatkish. I feel so lucky to find them at Virginia Tech and have them in my life.

Contents

List of Figures

List of Tables

Chapter 1

Retirement's Impact on Health: What Role does Social Network Play?

1.1 Introduction

In U.S., population aging so called silver tsunami causes deep social and political transformations, challenging society in many aspects. The continuing reduction in ratio of workers to retirees causes serious concerns about Social Security benefit sustainability, and growing of elderly population due to improvement in longevity increases the burden on medical care and pension systems. In particular, Social Security paid out more benefits than it collected in taxes in 2018 and recent prediction by Social Security Administration shows the trust will be depleted by 2034.[1]

These concerns prompt a series of policies such as increasing the eligibility age for full Social Security benefits and Medicare eligibility age. The effectiveness of increasing retirement age depends on an implicit assumption that late retirement is good for health or at least does not harm health. Retirement is a life-changing event that can improve or deteriorate elderly's health both physically and mentally (Nishimura et al., 2018; Dave et al., 2008; Coe and Lindeboom, 2008). Therefore, the precise identification of the causality impacts of retirement on the elderly's health is needed for effective policy design, implementation, and

[1] Social Security Administration Annual Report.

evaluation.

However, establishing the causal effects of retirement on health is empirically challenging for a couple of reasons. First, existing evidence suggests a reverse causal relationship between health and labor supply decisions and second, the existence of unobserved confounding factors that influence both health and retirement decisions simultaneously such as work environment and genetics.

Furthermore, as a life-changing event, retirement significantly alters retirees' daily routines, social contact, and social activities (i.e., Barnett et al., 2012; Eibich, 2015). If retirement affects social network while social network significantly impacts individual's health(Cohen, 2004), social network might be a health policy instrument that can be used to promote elderly's health: i.e. minimizing the negative impact of post-retirement's life changes on retiree's health through maintaining and fostering retirees' social capital.

However, disentangling the causality between health and social network suffers similar identification problem because the formation of social network is not random (Manski, 1993; Moffitt et al., 2001; Brock and Durlauf, 2001): individuals' unobserved heterogeneity simultaneously affects both health production and social network formation. People who have lower discount rates and put more value for future benefits are more likely to invest in health and social network that results in better health and more social capital in the future. Meanwhile, healthy individuals have more energy and time to socialize with others and enrich their social network. As a result, a model that examines the causality of retirement and social network on retirees' health outcomes will need to address the endogeneity problem that may cause inconsistent and biased estimations.

The literature has applied various identification methods to address the aforementioned empirical challenges in investigating the relationship between retirement and health but has not reached consistent findings (e.g., Kofi Charles, 2004; Neuman, 2008; Coe and Lindeboom, 2008; Coe and Zamarro, 2011; Gorry et al., 2018; Insler, 2014). The most recent studies (e.g., Gorry et al., 2018; Insler, 2014) used instrumental-variable (IV) and fixed effect (FE) methods with early and regular pension benefit eligible ages as instruments to deal with the endogeneity in the retirement status.

It is still hard to compare results across studies since they use different retirement definitions such as not working for pay, working less than 1200 hours per year, etc. Those definitions do not necessarily capture the effect of complete retirement (i.e., not working at all both for pay and for free) on health precisely.[2] Furthermore, to the best of our knowledge, there is no empirical study on the social network pathway through which retirement affects health.

The chief goal of this study is to identify the mediatory impact of social network on the subsequent health effect of retirement using the ego-centric social network and health information available in the National Social Life, Health and Aging Project (NSHAP) survey (discussed in detail in later section). For this purpose, we first investigate the retirement impact on health and address the endogeneity issue by employing the panel structure of NSHAP.[3] In particular, we tackle the endogeneity caused by unobservables and individual's heterogeneity by using eligibility age for full Social Security benefits as an instrumental variable for retirement status as well as applying Fixed-Effect methodology. Furthermore, we restrict our sample to pre-retirement healthy individuals to minimize the potential bias and inconsistency induced by reverse causality and potential weak instrument problem (i.e., people might retire in other time in the life span, not necessarily at the eligibility age for Social Security benefits). Lastly, we highlight the importance of investigating the causal relationship between retirement and health and the mediatory role of social network for public

[2] Dave et al. (2008) is the only study on US data which considers not working as the definition of retirement (using Health and Retirement Study dataset) and exploits the panel nature of data.

[3] The NSHAP has not been exploited in the literature of retirement and health studies.

policy intervention to promote healthy aging.

Our findings indicate that retirement negatively impacts physical and mental health outcomes especially self-report physical health, depression, and anxiety. Retirement significantly alters social network features, not only reduces the number of people in social network but also decreases the frequency of contacts in retirees as compared to non-retirees. The results suggest that the adverse effect of retirement on health can be partially explained by its effect on social network features. Using single and parallel mediation analysis, we find significant influence of social network size on the health of retirees. Particularly, the single mediation analysis indicates 43% of depreciation in physical health, and 24% of increment in depression post-retirement can be attributed to reduction in the social network size. In the parallel mediation identification that controls for possible dependencies between social network features, these effects are estimated to be 45.3% and 44.8%, respectively. Although the social network size reduction is found to be beneficial for ameliorating anxiety feelings, the magnitude of this effect is not sizable as compared to the large adverse effects that it imposes on physical health and depression. Findings in this paper suggest that social network interventions can mitigate the adverse health outcomes of retirement, providing valuable insights for policy design and implementation for healthy aging.

1.2 Review of Literature

1.2.1 Heath on Retirement

On top of the numerous factors that affect retirement decision, such as Social Security eligibility, financial resources, and health insurance, health is believed to be a crucial determinant. The empirical findings are consistent in supporting this conclusion. By modeling the endogenous health dynamics in a structural model of retirement, Bound et al. (2010) find that healthy people are unlikely to retire unless they have a sizable financial resource. It is more likely for those who are in poor health to retire before being eligible for any pension benefits. Capatina (2015) indicates four channels by which health affects an individual's labor supply: productivity, medical expenditures, available time, and mortality. She states that productivity and time lost to sickness are the main channels by which health affects labor supply. Similarly, Gustman and Steinmeier (2018) demonstrate that improving the overall health of the population would delay retirement by one year. Moreover, McGarry (2004) shows the impact of changes in health on retirement expectations is much greater than financial variables.

1.2.2 Retirement on Health

Recently, researchers have paid more attention to the effect of retirement on health outcomes. Several studies show a significant health improvement after retirement (e.g., Charles, 2002; Bound and Waidmann, 2007; Neuman, 2008; Coe and Lindeboom, 2008; Johnston and Lee, 2009; Coe and Zamarro, 2011; Insler, 2014), while others find that retirement significantly deteriorates health (e.g., Dave et al., 2008; Behncke, 2012). Recently, a systematic review (Nishimura et al., 2018) points out that different methodology utilized is a key factor for explaining the mixed results in the literature as well as the choice of wide ranges of control variables. For example, in a study using U.K. data, Behncke (2012) employs a non-parametric matching approach and finds no and negative effects of retirement on depression and self-report health, respectively. However, by using Regression Discontinuity Design, Johnston and Lee (2009) indicate that retirement lowers depression for a sample of men who do not have an educational degree in U.K.[4] Besides, different norms, labor market, and economic incentives embedded in the Social Security and pension system across different countries may also contribute to the inconsistency in empirical findings in the literature.

Even if we only compare the studies on US data, the findings are still mixed. For example, in terms of subjective well-being, by using Social Security normal retirement age as an instrumental variable, Kofi Charles (2004) finds that retirement has a positive effect on subjective well-being, while Dave et al. (2008) find no effect. For the self-report health, most studies find that the probability of reporting good health increases after retirement (Neuman, 2008; Coe and Lindeboom, 2008; Calvo et al., 2011; Nishimura et al., 2018), whereas some reports the opposite (Dave et al., 2008). In the case of physical health, Dave et al. (2008) report retirees are more likely to suffer from difficulties in their physical activities. While, other studies declare there is a positive relation between retirement and physical health for women but none for men (Neuman, 2008; Nishimura et al., 2018). In regards to mental health, some studies claim that retirement does not affect depression (Neuman, 2008; Coe and Lindeboom, 2008), while others find that retirement is associated with higher depression(Nishimura et al., 2018).

1.2.3 The Role of Social Network

A few studies investigate how retirement changes the lifestyle of retirees (Zantinge et al., 2013; Barnett et al., 2012). Specifically, they investigate the effect of retirement on social activities participation and social support (Barnett et al., 2012). Social network is defined as a "web of social relationships surrounding an individual and the characteristics of those ties" (Berkman et al., 2000), which are generally characterized in terms of structure, quality,

[4] There are many other studies based on the evidence of other countries such as Korea, Japan, Canada and European countries.

and function. Social network structure refers to the number of individuals in the network and the frequency of contacts an individual has with the other members.

An important function of a social network is to provide social support, especially emotional and instrumental support, to members of the network and influence mental and physical health (especially for those older individuals) (e.g., Israel, 1982; Cohen, 2004; Ronconi et al., 2012; Petrou and Kupek, 2008; Fiori and Jager, 2012; Litwin and Shiovitz-Ezra, 2010; Allen et al., 2014). Social network members can provide social support and companionship for day-to-day adaptation to new life of retirees that will contribute to better health.Berkman and Syme (1979) state that those who have larger social networks are more likely to have better health, especially when the network members are frequently contacting each other (Terhell et al., 2007). Social network involvement might include negative social interaction in which social network members behave in hurtful and inconsiderate ways that result in worse health (Krause and Rook, 2003). Noticeably, studies found that positive social interactions and social support happen much more frequently than do negative interactions (Rook, 1998).

The effect of retirement on social network is an empirical hypothesis. On the one hand, retirement reduces social interactions because of losing co-workers and work-related networks and thus shrinks the size of social network (Sugisawa et al., 1997). On the other hand, people would have more time to do voluntary works, participate in different types of social activities, and make new connections to expand their social networks (Barnett et al., 2012).[5] No matter which direction the impact is, retirement will lead to social network changes and subsequently may impact retirees' health. Therefore, it is needed to investigate the empirical evidence of social network's mediation effect in the pathway between retirement and health to understand whether social network is a viable channel to invest on for healthy aging

[5] Recently, several studies attempt to investigate how social network impacts retirement decision making (van den Berg et al., 2010; Harkonmäki et al., 2006). Findings show that social support can buffer the effect of main factors of early retirement decision such as poor health, low job satisfaction and work pressure (van den Berg et al., 2010).

promoting policies.

1.3 Data and Variables

1.3.1 Data

In our empirical analysis we use data from the National Social Life, Health, and Aging Project (NSHAP) which is conducted by National Opinion Research Center (NORC) at the University of Chicago. This is a population-based panel study of the elderly with specific purpose to investigate connections between health and social factors. The first wave of the NSHAP includes a sample of 3,005 adults aged 57 - 85 years old (born between 1920 and 1947) who were interviewed in 2005 or 2006. Wave 2 consists of 2,261 Wave 1 respondents who were re-interviewed in 2010 or 2011. Wave 2 also includes the cohabiting spouses and romantic partners of Wave 1 respondents in addition to Wave 1 non-interviewed respondents. Wave 3 constitutes of 4,777 individuals who were interviewed in wave 2 in addition to a new cohort born between 1948 and 1965 (baby boomers).[6] The construction of the sub-sample used in our study will be explained in subsection 1.3.5.

1.3.2 Health variables

As outcome variables, we consider a variety of physical and mental health measures available in NASHAP. Physical health. We examine the self-report physical health which is reported by respondents as “poor,” “fair,” “good,” “very good” or “excellent”, on scale from 1 to 5. Higher values of self-report physical health correspond to better health.

[6]We use sample weights for all the analyses.

Mental health. Respondents were asked “how do you evaluate your emotional health?” and respondents answered as “poor,” “fair,” “good,” “very good ” or “excellent” , on scale from 1 to 5. Higher value refers to better mental health. Since this measurement of mental health is not reliable (Mawani et al., 2010), we select depression and anxiety as more suitable proxy for mental health. For example, Mawani et al. (2010) state that noticeable percentage of respondents who were diagnosed to have mental illness reported their mental health as good. Also, Fleishman and Zuvekas (2007) find that measures of emotional distress are weakly correlated with self-report mental health.[7]

Depression. The NSHAP includes a depression scale introduced by Center for Epidemiologic Studies (CES) which is based on a cumulative summation over response scores to eleven questions. Respondents were asked about the frequency of certain feelings in the past week (e.g., “how often did you feel depressed in the past week?”).Certain feelings include: Depressed, Restless, Difficult, Poor appetite, Everything was an effort, Happy, Lonely, People were unfriendly, Enjoyed life, Sad, and Being disliked. There are four possible responses: “rarely or none of the time” (score: 1); “some of the time” (score: 2); “occasionally” (score: 3); and “most of the time” (score: 4).[8] Higher values of CES-D depression measurements represent more depression symptoms and worse mental health.

Anxiety. NSHAP includes seven questions related to anxiety symptoms defined by Hospital Anxiety and Depression Scale (HADS). Respondents were asked about the frequency of feeling anxiety symptoms (i.e., tense, something awful about to happen, restless, worried, relaxed, frightened, and panic) in the past week. The range of answer to these questions are similar to the CES-D depression. Hence, we calculate the anxiety measurement by summation over the response scores to the seven questions. The higher values of anxiety

[7] https://www.cdc.gov/tobacco/campaign/tips/diseases/depression-anxiety.html

[8] The answer scores to positive notion questions (e.g., “How often did you feel happy in the past week?”) are reversed to be consistent in measuring depression.

variable demonstrate worse mental health.[9]

1.3.3 Social network variables

NSHAP includes the respondent's egocentric social network. An egocentric social network includes an ego (the respondent) and a set of members. In NSHAP, respondents could name up to five people that immediately surround them in the past 12 months, but respondents were also asked to denote if they had more than five members in their networks. Also, NSHAP contains frequency of contacts among members including the respondent. Frequency of contacts are collected by asking respondents "how often do you talk to the person cited?" The responses range "have never spoken to each other (0)," "less than once a year (1)," "once a year (2)," "a couple of times a year (3)," "once a month (4)," "once every two weeks (5)," "once a week (6)," "several times a week (7)," "every day (8)." In NSHAP, respondents are asked to describe type of their relationship with each member in the network (e.g., partner, family, friend, coworker, etc.).

Based on available information about social network in NSHAP, we construct the most frequent examined network characteristics in the social network literature: size of social network, frequency of contacts, and diversity of ego's network (Carolan, 2013). Size is defined as number of members in the respondent's social network. We construct an index for measuring frequency of contacts by summation over the scores of ego's frequency of contacts (i.e., scores range from 0 to 8) with members and normalize it with size of network. This normalization makes the frequency measure less dependent on the size. To illustrate this, consider ego1 who has one member in her social network and every day is in contact with that person as compared to ego2 who has 8 members in his network and is in contact with

[9]Anxiety questions are asked in the leave-behind questionnaire, and around 2600 respondents have answered these questions in each wave. Therefore, the sub-sample for investigating the anxiety impact has the lowest sample size.

them less than a day per year. In terms of the frequency of contacts, without normalization, these two have the same frequency of contacts (i.e., 8).

Based on varieties of relationship types in NSHAP, we define 8 categories: partner, parent, child, family, friend, neighbor, coworker, and others(i.e., minister, priest, or other clergy, psychiatrist, psychologist, counselor, or therapist, caseworker/social worker, and housekeeper/home health care provider). We utilize the Index of Qualitative Variation (IQV) (Knoke and Yang, 2008) to construct the diversity measurement of social network. For the ith respondent with N members in the network, where members are classified into K categories, the IQV is defined as follows:

$$IQV = \frac{1 - \Sigma_{j=1}^{k} P_j^2}{\frac{k-1}{k}} \tag{1.1}$$

In which P_j is the percentage of members of network in the jth category. The IQV is a standardized measure ranging between 0 and 1, where 0 indicates all N members are in one category and 1 indicates members are equally dispersed across K categories.

1.3.4 Retirement and control variables

NSHAP contains demographic and socioeconomic information such as gender, age, ethnicity, marital status, educational attainment, income, employment status (i.e., currently working, retired, disabled, unemployed, homemaker, or other), etc. Dichotomous indicator is defined to be 1 for retirement if the respondent reports retired and not working, the indicator is 0 otherwise. We focus on full retirement to capture the largest effect that retirement can have on health. The NSHAP also includes respondent's health behaviors like smoking and drinking. Particularly, respondents were asked whether they smoke cigarettes currently and how many cigarettes they smoke per day, whether they currently drink alcohol and the number of drinks they consume per day. In our empirical analysis, we control for these observable characteristics.

1.3.5 Sample selection

We restrict the sample to sub-sample of healthy individuals before retirement event. Although we lose many observations, we immune our findings from potential simultaneity problem between retirement and health. There are several advantages for us to focus on pre-retirement healthy individuals: first, it is less likely that health causes retirement for these individuals (i.e., reverse causality is minimized). Second, the bias induced by unobservable confounding factors that simultaneously influence both health and retirement is also minimized. Third, this sample selection strategy also protects our estimations from weak instrument bias (more details will be provided in the empirical methodology section).

Likewise, this sub-sample selection helps minimize the bias caused by the possible endogeneity of social network as well. Remarkably, social network involvement might not only be disturbed by health problems, but also be changed by retirement. Focusing on the sub-sample of healthy individuals before retirement guarantees bad health does not interfere the social network changes. In particular, individuals are defined as healthy if they report good, very good or excellent as their self-report physical and mental health and have depression and anxiety score less than 16 and 8, respectively.[10] Consequently, our pre-retirement healthy sub-sample include 1,160 individuals. Hence, we sacrifice the variation in observations for the sake of validity of results and preventing bias and inconsistency in the estimations.

[10] CESD score lower than 16 implies no depression symptoms exist (Radloff, 1977) (Radloff, 1977). Also, anxiety scores below 8 indicates no anxiety is diagnosed (Zigmond and Snaith, 1983)

1.3.6 Summary Statistics

Table 2.1 provides data description and summary statistics of the variables in the selected sub-sample of pre-retirement healthy individuals, and compares retirees vs. non-retirees. According to Table 1, 45% of sample are retired who are on average significantly older than non-retirees (i.e., 70 years old versus 66 years old).[11] The mean of self-report physical and mental health depicts that on average retired people are in worse health condition as compare to non-retirees. This Table also indicates retirees on average experience higher levels of depression and anxiety symptoms than non-retirees do.

According to this table, on average, retirees have lower income as compared to non-retirees. With regards to health behaviors, retirees drink alcohol less than non-retirees on average, whereas no significant difference exist between these two groups in terms of smoking. While retirees on average seem to have larger social network with higher frequency of contacts and lower diversity in their network as compare to non-retirees' social network, no significant differences exist on these characteristics among the two groups.

Some limitations in the data set prevent us from a comprehensive investigation of the effects of social network. In particular, we have no information about the geographical distance between respondents and members, the form of communication between social network members whether it is by mail, internet, in person, or by phone. Also, the lack of detailed information on health condition and labor status of the social network members deprives the opportunity for examination of potential interactive and spillover social effects. Moreover, although health insurance is a key variable in late life that influences both health and retirement decision, we exclude it from our empirical analysis due to the lack of variations in the sample. Particularly, more than 80% of the sample have health insurance in the first two waves.

[11] This is according to our definition of retirement.

1.4 Empirical Methodology

1.4.1 Retirement and Health

For simplicity, consider the following linear specification of health as a function of retirement in equation (1.2):

$$H_{it} = \alpha_0 + cR_{it} + \alpha_2 X_{it} + \mu_i + \epsilon_{it} \tag{1.2}$$

$$R_{it} = \delta_0 + \delta Z_{it} + \delta_2 X_{it} + \mu_i + \nu_{it} \tag{1.3}$$

Where, H_{it} is the health status of individual i at time t and R_{it} is retirement status of individual i at time t. X_{it} denotes time-variant observable characteristics such as income, marital status and health behaviors. μ_i indicates time-invariant unobservable characteristics of individuals such as genetics, family background, and time preferences. ϵ_{it} and ν_{it} are i.i.d error terms. We are interested in unbiased and consistent estimation of c which is a challenging task due to reasons discussed earlier, i.e. the reverse causal relationship between health and labor supply decision, and the existence of unobserved confoundings that influence both health and retirement decision simultaneously.

Estimation of equation (1.3) as the first stage of the Instrumental Variables (IV) strategy aims to account for endogeneity in retirement. The eligibility age for full entitlement Social Security benefits (i.e., 65 years old) is a valid and relevant instrument. Accordingly, in equation (2), we use $Z_{it} = 1(Age \geq 65)$ as an exogenous source of variations in retirement. To be a valid instrument, an instrument must be correlated to retirement (endogenous regressor) and related to health outcomes only through the effect on retirement (i.e., $corr(Z_{it}, \epsilon_{it}) = 0$). Well-documented literature exists about the sharp changes in retirement behavior around this age (e.g., Ruhm, 1995; Coile and Gruber, 2000), which confirms this cutoff point should have the power to predict retirement decision. It is worth mentioning that an individual's health outcomes do not change substantially by officially turning one year older, even though health gradually declines as people aged.

Empirical findings using NSHAP also provide evidence for the discrete changes in retirement decision around the eligibility age for full Social Security benefits. Fig.1.1a presents the jump in retirement behavior at 65 years old, while Fig.1.1b confirms the self-report health does not necessarily decrease around this point. Therefore, eligibility age for full Social Security benefits is a valid and relevant instrument. Fixed-effect regression estimation of equation (1.3) as the first stage shows that reaching 65 years old is significantly associated with a 75 percentage-point increase in retirement for the pre-retirement healthy individuals sub-sample.

In addition to endogeneity, addressing the heterogeneity effect is also essential. Due to various job characteristics and socioeconomic background, some individuals may experience better health after retirement, while others may experience no changes or deterioration on health upon retirement. Besides, heterogeneity in health investment behaviors might be another source of variations in health after retirement (Grossman, 1972). According to Grossman's model, health is both investment and consumption goods. Upon retirement, with no incentive to invest in health to increase productivity and thus earnings, individuals may not sufficiently invest in their health , which lead to poor health post-retirement. However, health as a consumption good directly enters the utility function and retirees may invest more in it, which results in better health post-retirement.

Failing to consider these individual heterogeneity leads to inconsistent and bias estimation of retirement's impact on health. Using a vast set of controls cannot fully purge out the bias of heterogeneity and unobserved selection. Exploiting the panel nature of data allows for applying individual fixed effects (FE) method that controls for all unobserved time-invariant heterogeneity across individuals. Although using IV can help to minimize the endogeneity bias, full retirement age for eligibility of Social Security benefit might be a weak instrument for prediction of retirement because individuals retire all the time during the life span (i.e., even after full retirement age). For instance, including individuals who retired after 65 years old due to health problems results in overestimation of retirement's impact on health. To avoid the bias and inconsistency of estimations, we restrict the sample to pre-retirement healthy individuals to assure that irrelevant sample selection do not contaminate our results. For healthy individuals before retirement, retirement is less likely to be endogenous to health through unobservable factors and reverse causality. Therefore, we capture the causality of retirement on subsequent health by applying the FE-IV approach on equations (1.2) and (1.3) to the sub-sample of healthy individuals before retirement.

1.4.2 Mediation effect of social network

We are particularly interested in the mediation effect of social network on post-retirement changes in health. It is necessary here to clarify exactly what is meant by mediation. We seek to find how retirement exerts its effect on health outcomes through social network characteristics.

According to Fig.1.2, two pathways exist for retirement to influence health, the direct and indirect effect. The direct effect refers to the pathway from retirement to health without passing through social network. While the indirect effect refers to the pathway from retirement to health through social network. If retirement alters the social network characteristics (e.g., lost the contact with colleagues or raising opportunity to engage in different social activities), and social network impacts individual's health, then social network would be a health policy instrument that can be used for intervention in promoting elderly's health.

The corresponding econometric model is as follows:[12]

$$\begin{cases} SNW_{it} = \alpha_0 + aR_{it} + \alpha_1 X_{it} + \mu_i + \nu_{it} \\ \\ \quad H_{it} = \beta_0 + c^{net}R_{it} + bSNW_{it} + \beta_1 X_{it} + \mu_i + \epsilon_{it} \end{cases} \tag{1.4}$$

Where, SNW_{it} refers to a specific social network characteristic (i.e., size, frequency of contacts, or diversity), and the rest of the variables are as defined previously. ν_{it} and ϵ_{it} present the i.i.d error term in each equation, respectively.

We follow the revised method of Baron and Kenny (1986) by Zhao et al. (2010) for mediation analysis. According to Barron and Kenny (1986), the multiplication of a and b is the social network mediation on the health impact of retirement, and c^{net} is the direct impact of retirement on health holding social network variable constant. Therefore, the summation of c^{net} and $a \times b$ indicates the total effect of retirement on subsequent health. To tackle the endogeneity problem, we apply the FE-IV approach to each equation using the sub-sample of pre-retirement healthy individuals.

For policy making, it would be interesting to know if any of the social network characteristics drives the mediation more than the other two or if all three contribute to it similarly. Even though the social network characteristics are correlated with each other, no theoretical reason- to the best of our knowledge- exists to support the causality link between them. Parallel mediation analysis presented in Fig. 1.3 is a method to compare magnitudes of the indirect effects while allows for the correlations among the mediators. As shown in Fig.1.3, the absence of any arrows linking the mediators (i.e., social network characteristics) assumes no causality link exist between them.

The corresponding econometric model to Fig.1.3 is defined as follows:

[12] The mediation analysis models first introduced in 1986 by Baron and Kenny (1989) and developed further with less restrictive assumptions in later years, i.e., Zhao et. al., 2010.

$$\begin{bmatrix} Size_{it} \\ Frequency_{it} \\ Diversity_{it} \\ H_{it} \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ c^{net} \end{bmatrix} R_{it} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ b_1 & b_2 & b_3 \end{bmatrix} \begin{bmatrix} Size_{it} \\ Frequency_{it} \\ Diversity_{it} \end{bmatrix} + \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \mathbf{X}_{it} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \mu_i + \begin{bmatrix} \nu_{0it} \\ \nu_{1it} \\ \nu_{2it} \\ \nu_{3it} \end{bmatrix} \tag{1.5}$$

or,

$$\mathbf{Y} = \beta + \mathbf{a}R_{it} + \mathbf{B}SNW_{it} + \alpha\mathbf{X_{it}} + \mathbf{l_4}\mu_i + \nu \tag{1.6}$$

Where all variables are similar to those defined previously in equations (1.3) and (1.2). $Size_{it}$ refers to social network size of individual i at time t, $Frequency_{it}$ indicates the frequency of contacts of individual i with members of her social network at time t, and $Diversity_{it}$ refers to diversity index (i.e., IQV) for an individual i at time t.

According to Fig.1.3, we have three different indirect effects of retirement on health such that each one passes through one social network characteristic ($a_1 \times b_1$, $a_2 \times b_2$, and $a_3 \times b_3$). The sum of the three indirect effects and the direct effect of retirement gives the total effect of retirement on health.

1.5 Empirical Results

1.5.1 Total effect of retirement

The first set of analysis examines the total impact of full retirement on subsequent health. Table 1.2 presents the total effect of retirement on self-reported physical health, mental health, depression, and anxiety by FE-IV estimation of equation (1.3) and equation (1.2) as the first stage. This evidence shows, in line with the findings in Dave et al. (2008), that retirement generates a significant adverse effect on physical health, depression, and anxiety in elderly, with no effect on self-reported mental health. However, the primary focus of this study is to reveal how social network can mediate the effect of retirement on subsequent health.

1.5.2 Single mediation analysis

Estimation results of the single mediation model (i.e., Fig. 1.2) are shown in Tables 3 and 4. Each panel of the tables provides results for a specific health outcome, and each row presents the coefficient estimation of equations in system (1.4) considering one specific social network feature as the mediator variable.

We first explain how retirement affects social network features. In Tables 1.3 and 1.4, "a" refers to the estimation of retirement's impact on the size, frequency, and diversity of social network (corresponding to "a" in Fig. 1.2), separately. The result shows that retirement significantly reduces the size of social network and frequency of contacts (Sugisawa et al., 1997), while it has no significant impact on diversity in one's network.

As shown in the third column of these tables, none of the health outcomes appeared to be affected by different features of social network (i.e, b=0)- after controlling for the time-invariant effects.[13] Even if social network does not generate a significant direct impact on health, it could still perform as a mediator on the impact of retirement on health. As a matter of fact, according to Zhao et al. (2010), it is not necessary to have a significant a or b to establish a mediation effect, instead, the distribution of their products matters

[13] In Appendix , Table A.1 shows a highly significant impact of social network on health outcomes using OLS estimation which does not control for the time-invariant confounding, i.e., unobservable time-invariant confounding are the factors affecting both social network and health here.

for the existence of mediation effect. We obtain the empirical distribution of a×b using bootstrapping to calculate the standard errors and find that it is significant at the 95% level.

The fourth column display the mediatory effect of social network features on the path from retirement to health. As we see, the social network plays a significant mediatory role with the highest impact of size in comparison to the other two features. Remarkably, size shrinkage explains 43% (i.e., $\frac{0.093}{0.280}$) of reduction in physical health caused by retirement, frequency of contacts explains around 8% of it, while the mediatory impact of diversity is less than 1%.

Mental health: The adverse direct effect of retirement is canceled out by its indirect effect through social network features and as a result, the total effect of retirement on mental health is not significant.

Depression: social network size reduction explains $\frac{0.234}{0.962} = 24\%$ of the increase in depression upon retirement. Using the frequency of contacts, we find no mediatory effect, and diversity reflects less than 1 percent of increment in depression post-retirement. Thus, the size of retiree's social network explains the highest adverse effect of retirement on depression as compared to the other two features.

Anxiety: interestingly, findings show that reduction in social network size and frequency of contacts upon retirement significantly contribute to the mitigation of anxiety. This can be explained by the concept of social anxiety in elderly or negative impact of social network involvements that is discussed earlier. Social anxiety implies elderlies in contact with others often feel they are a burden on the life of people surrounding them. However, the magnitude of anxiety reduction through social network size (frequency of contacts) is about 2% (6.7%).The marginal reduction in diversity on social network by retirement significantly contributes to the increment of anxiety feelings. However, the magnitude is small. The direct impact of retirement on anxiety dominates the beneficial mediatory effect of social network characteristics and leads to overall higher anxiety symptoms in retirees as compared to non-retirees.

The last column of Tables 1.3 and 1.4 shows that the direct effect of retirement is significant on all health outcomes ($c^{net} \neq 0$) except mental health. The nonzero direct impact of retirement on health might be due to all the other mechanisms through which retirement might influence health such as financial hardships, health insurance coverage, lifestyle changes, etc.

In summary, simple mediation analysis suggests social network as a significant channel that influences physical and mental health upon retirement. In the next section, we consider the mediatory effects of all three social network features simultaneously in the model.

1.5.3 Parallel mediation analysis

According to Fig. 1.3, retirement is modeled to exert its effect on health through 4 pathways. One pathway is direct, from retirement to health without passing through any of the proposed social network mediators, and the other three pathways are indirect, each through one feature of social network. The estimation results corresponding to Fig. 1.3 (i.e., system of equations in (1.5)) are shown in Table 1.5. These parallel analyses shed more light on the social network mechanism for the post-retirement deterioration in health.

In the first row of Table 1.5, we see that retirement significantly deteriorates physical health, ceteris paribus- holding social network variables and all other explanatory variables constant. The mediation coefficient estimation of social network size implies that retirees are likely to have worse physical health than non-retirees due to a smaller social network size at the same levels of frequency of contacts and diversity. Particularly, size shrinkage explains 45.3% of reduction in self-reported physical health among retirees, while holding frequency of contacts and diversity constant. In contrast, retirees are estimated to have better physical health in comparison to non-retirees through the effect of retirement on the frequency of contacts/diversity, holding size, and everything else constant. In contrast to single mediation analysis, the parallel mediation analysis reflects that with the same social network size and diversity, higher frequency of contacts is associated with worse physical health outcome. Overall, the adverse mediatory effect of size dominates the impact of the other two features of social network.

As expected, no direct effect of retirement on subsequent mental health is observed in the parallel mediation, while the indirect effects through social network mediators are significant with the effect from size being the highest. Remarkably, parallel mediation estimation reveals that the reduction of frequency of contacts upon retirement reduces mental health in contrast to single mediation estimations. It reflects the importance of considering the correlation between the social network features in the analysis rather than assessing them in isolation from the other related features. We observe similar patterns in direct and indirect effects of depression to physical health. In particular, about 44.8% increase in depression after retirement is explained by the reduction in the size of social network, holding frequency of contacts and diversity constant.

In contrast with simple mediation analysis, smaller social network size is associated with higher levels of anxiety post-retirement, holding the frequency of contacts and diversity of network constant. The size of social network contributes to the increase in anxiety in retirees by 5.1%. Whereas, the frequency of contacts and diversity estimation show similar effects to single mediation analysis in the parallel system. Also, similar to single mediation analysis, frequency of contact has the largest mediatory impact on retiree's anxiety. Overall impact of retirement through altering social network characteristics appears to be effective for anxiety reduction.

Comparing the mediatory effects of social network features across health outcomes, we find that size of social network generates the largest effect in comparison to the other two features for health outcomes except for anxiety for which frequency of contacts have the largest absolute mediatory effect. According to the last column of Table 5, retirement deteriorates physical health, increases depression, and improves mental health and anxiety through its influence on social network features. Noticeably, the harm to physical health and depression induced by social network alterations upon retirement is much larger than its beneficial impact on anxiety and mental health,. These results imply that adverse health impacts post-retirement can be curbed by improving social network of retirees.

1.6 Discussion

Different conceptual theories in psychology and sociology address how social network influences health or vice versa (e.g., Israel, 1982), while no study empirically investigates the effect of health on social network. Yet, health might be a potential factor affecting network formation since the formation of social network is not random (Moffitt et al., 2001; Brock and Durlauf, 2001). For instance, healthier people are more likely to socialize and have more social connections. Besides, individuals' heterogeneity, such as time preferences and personality, might simultaneously contribute to the extent of social network involvement and health conditions. For example, extrovert people are more likely to participate in different social activities, meanwhile they are less likely to have depressive symptoms. Although disentangling the causal effect of social network on health suffers similar identification problems as health and retirement, the source of endogeneity is less severe for elderly.

According to socioemotional selectivity theory (SST) (Carstensen, 1991), as people age, they focus on enriching and maintaining the existing relationships rather than investing in new ones. That is, the effect of unobservables on the elderly's health is less likely to influence their social network formation, meaning that unobservable confounding factors are not significant matter of endogeneity between health and social network in this population. As we limit the sample to healthy individuals before retirement event, we further assure that health does not interfere the changes in social network post-retirement as well (i.e., no reverse causality).

Furthermore, an implicit time lag between health measurements and social network information collection in NSHAP provides a temporal exogeneity by which the current health is less likely to alter the social network structure in the past for elderly. Individuals were asked about their social network details during the past 12 months, while the health relevant questions are related to the past week. Thus, reverse causality is less source of concern in this study. However, we acknowledge that this study is limited by lack of information for developing a valid and relevant instrument for social network to sufficiently tackle the endogeneity of health and social network.

1.7 Policy Implications

Evidence provided by this study has important policy implications. According to our findings of post-retirement deterioration in physical and increment in depression and anxiety, increasing retirement age might be a prudent policy. This policy not only assures delaying Social Security trust fund depletion but also reduces health care expenditure by postponing poor health outcomes in the affected population.

The key message in this study is that considering the health challenges caused by population aging, social capital (in the form of social network) is a key policy instrument for health promotion in the elderly. Investing in the social capital of the elderly would curb the negative health effect post-retirement, even if the retirement eligibility age remains unchanged.

Our findings reveal that 43% of total deterioration in physical health after retirements is attributed to the drop in the number of people in the social network of retirees. Even if we simultaneously control for the other social network features, this amount remains to be large at 45.3 %.

Enriching social network in elderly can also ameliorate the depression symptoms after retirement. Our finding indicates that significant portion (i.e., 24% in single mediation analysis and 44.8% in parallel mediation estimation) of increment in post-retirement depression happens due to a reduction in social network size. Although reduction in social network size and frequency of contacts are found to improve anxiety feelings, studies show that negative impact of social network involvement on health outcomes happens less often than positive and supportive interactions (Rook, 1998).

Therefore, interventions that target promoting different aspects of the elderly's social capital buildup are promising ones to allocate resources into, aiming to improve elderly's health. Particularly, the government can improve health of the elderly by investing in their social capital. For example, one effective policy is to provide education on technologies that can minimize social network size shrinkage in retirees. Also, the government can directly provide subsidies to promote the elderly's social capital enrichment, for instance, by organizing community elderly activities targeting groups that share similar prior occupations, or by providing funding to stimulate community participation through voluntary organizations and community groups.

1.8 Conclusions

In this study, we investigate how retirement can impact health through altering social network characteristics, including size, diversity, and frequency of contacts. First we estimate the health impact of retirement using the FE-IV method on pre-retirement healthy individuals in NSHAP data set. Then, we investigate the mediatory role of social network on the health impact of retirement. We investigate the mediatory effect of social network characteristics in single and parallel mediation model specifications, separately. The parallel analysis has the advantage of accounting for possible dependence and correlation between different social network features, while single analysis considers one social network feature at a time.

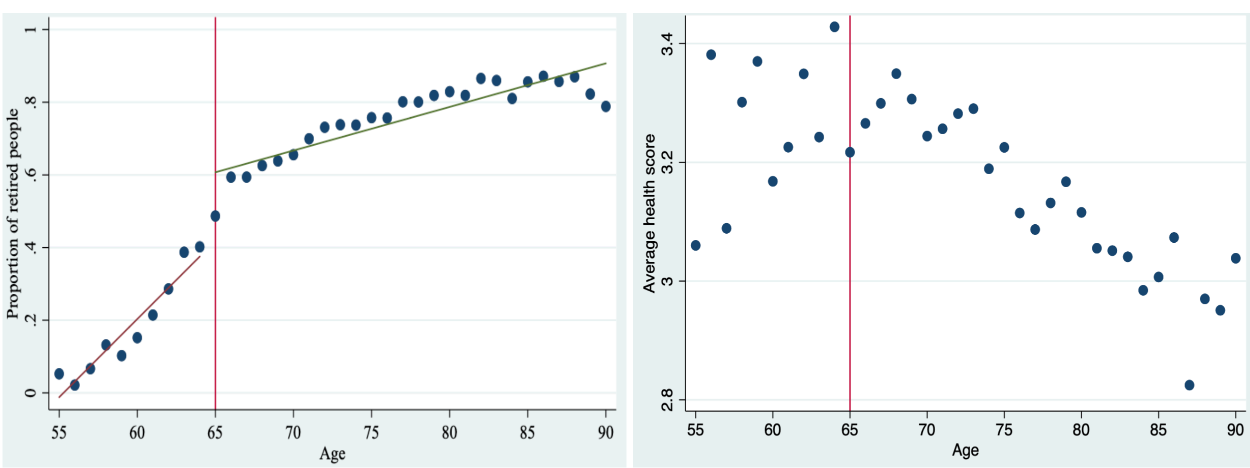
Estimations reveal a statistically significant negative effect of retirement on physical health, depression, and anxiety, with no significant effect on mental health. Findings uncover that retirees not only have fewer members in their social network but also have less frequent contacts with members as compared to non-retirees. Due to the association of social network and health, differences in social network of retirees and non-retirees explain a substantial amount of disparities in health outcomes between the two groups. Our findings indicate that a considerable portion of retirement's impact on health is mediated by social network changes.

In particular, single mediation identification suggests that 43% of depreciation in physical health and 24% of increment in depression post-retirement can be explained by the reduction in social network size. Whereas, in the parallel identification which holds the other social network features constant, the corresponding portions of the effects of social network size reduction on negative physical health outcome and depression upon retirement become 45.3% and 44.8%, respectively. We find that social network changes upon retirement might be beneficial for anxiety in retirees as compared to non-retirees. Retirement reduced anxiety through reduction in social network size and frequency of contacts which can be explained by social anxiety concept in elderly. However, this beneficial impact of social network size reduction on lessening anxiety is not comparable with the large harms it imposes on physical health and depression.

We find social network changes due to important lifestyle change after retirement and can be utilized as an effective policy instrument to buffer the adverse health outcomes of retirement. The adverse impact of retirement on health outcomes suggests that increasing retirement age might be a suitable policy to decrease the burden on health care services and increase Social Security trust fund sustainability. Social network based interventions that target social capital buildup of the elderly might be desirable for healthy aging. For instance, the government can provide subsidies for organizing and establishing different community events to stimulate social capital buildup for the elderly.

Understanding the underlying mechanisms of how social network can mediates the adverse effect of retirement on health calls for more detailed information about the structure of social network of the elderly. This will be investigated in a future project as more social network data becomes available.

1.9 Figures and Tables



(a) Retirement behavior on different ages (b) Average self-report health on different ages

Figure 1.1: Retirement and health behavior of different ages. Higher health score means better self-report health.

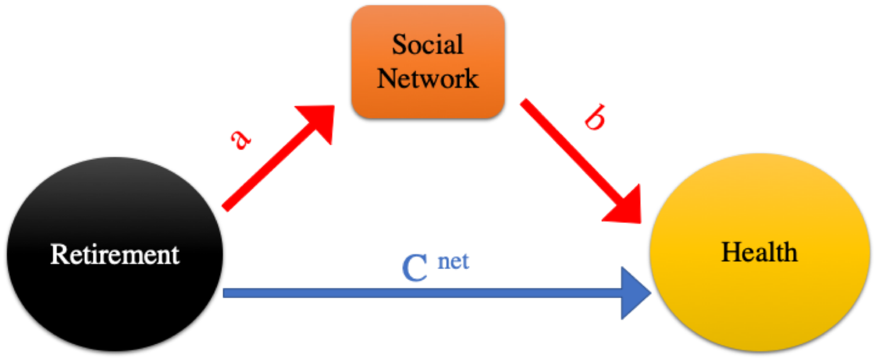


Figure 1.2: Single mediation model of social network on the relationship between retirement and health

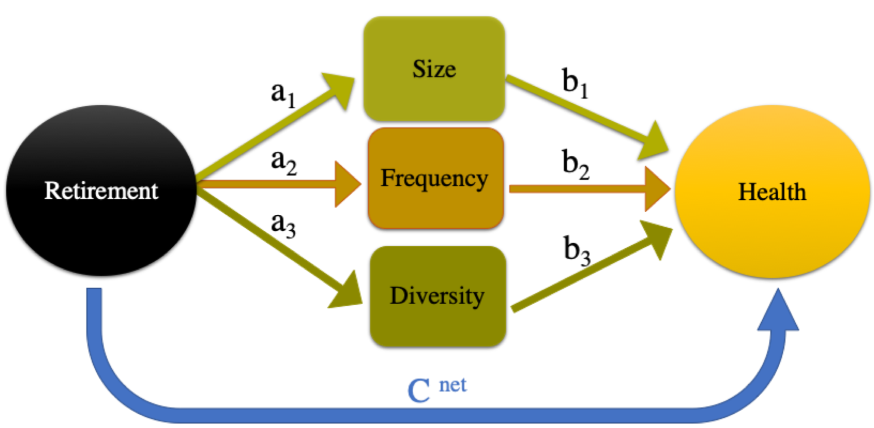


Figure 1.3: Parallel mediation model of social network on the relationship between retirement and health

Table 1.1: Summary statistics and data description of sub-sample of study

	Not-retired	Retired	Total	Difference	Min[a]	Max[b]
Employment Status						
Retirement	0	1	0.45	-1	0	1
			(0.49)			
Health						
Physical Health	3.89	3.51	3.8	0.38***	1	5
	(0.76)	(1.00)	(0.84)	(0.08)		
Mental Health	4.19	3.98	4.14	0.20***	1	5
	(0.72)	(0.83)	(0.75)	(0.08)		
Depression	12.88	13.93	13.13	-1.05***	11	29
	(1.96)	(3.16)	(2.34)	(0.23)		
Anxiety	2.39	3.48	2.65	-1.08***	0	14
	(2.23)	(3.18)	(2.53)	(0.25)		
Demographics						
Age	66.24	69.95	67.12	-3.70***	57	90
	(6.01)	(6.02)	(6.21)	(0.61)		
Female	0.46	0.44	0.46	0.02	0	1
	(0.50)	(0.50)	(0.50)	(0.05)		
Married	0.73	0.67	0.72	0.06	0	1
	(0.44)	(0.47)	(0.45)	(0.05)		
White	0.85	0.83	0.84	0.02	0	1
	(0.36)	(0.38)	(0.36)	(0.04)		
Asian	0.05	0.05	0.05	-0.01	0	1
	(0.21)	(0.23)	(0.21)	(0.02)		
High education	0.75	0.74	0.75	0	0	1
	(0.44)	(0.44)	(0.44)	(0.04)		
High income	0.88	0.81	0.86	0.06*	0	1
	(0.33)	(0.39)	(0.35)	(0.03)		
Health Behaviors						
Alcohol consumption	0.69	0.59	0.67	0.09**	0	1
	(0.46)	(0.49)	(0.47)	(0.05)		
Number of drinks	1.3	1.09	1.25	0.22	0	15
	(1.51)	(1.16)	(1.44)	(0.15)		
Smoking cigarette	0.11	0.09	0.1	0.02	0	1
	(0.31)	(0.28)	(0.31)	(0.03)		
Social Network						
Size	9.46	9.52	9.48	-0.05	2	18
	(3.08)	(3.18)	(3.10)	(0.31)		
Frequency of Contacts	14.85	15.63	15.03	-0.79	2	31.3
	(5.15)	(5.65)	(5.28)	(0.53)		
Diversity	0.67	0.64	0.66	0.03	0	0.95
	(0.19)	(0.21)	(0.20)	(0.02)		
Observations	633	537				

Note: This Table reports the summary statistics of the sub-sample of healthy (healthy in self-reported physical and mental health, no depression, and no anxiety) individuals before retirement. Retired refers to individuals who reported retired and not working. [a]: Min denotes the minimum values of each variable.[b]: Max denotes the maximum values of each variable. Standard deviations are in parentheses. For some variables, the actual sample size is less due to missing information and because some of the variables are in the leave-behind questioners. Asterisks present that the difference between the retired and non-retired samples is statistically significant as follow, * p<0.10, ** p<0.05, *** p <0.01.

Table 1.2: FE-IV estimation of total impact of retirement on subsequent health outcomes

	Physical Health	Mental Health	Depression	Anxiety
Retirement	-0.280***	-0.067	0.962***	1.320***
	(0.08)	(0.08)	(0.30)	(0.32)
Observations	1018	1012	1016	966

Note: Each cell presents the total effect of retirement on corresponding health outcomes using the Fixed-Effect Instrumental Variable estimation method. In all analyses we control for health behaviors, marital status, and income levels. Instrument is a dummy variable with value 1 if individual is older than 65 years old, and 0 otherwise. The sample is limited to pre-retirement healthy individuals in all these four health outcomes. Table A.2 in Appendix A reports the detailed estimation results. Robust individuals-clustered standard errors are in parenthesis. The significance level is defined as follows: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 1.3: Single mediation analysis: FE-IV estimation of mediatory impact of social network characteristic on retirement's effect on physical and mental health

	Retirement impact on SNW	Social Network impact on Health	Mediation impact of SNW	Direct effect of retirement on health
Physical Health	a	b	$a \times b$	c^{net}
Size	-4.300*** (0.424)	0.021* (0.011)	-0.093** (0.003)	-0.188* (0.110)
Frequency	-2.739*** (0.517)	0.006 (0.007)	-0.017** (0.001)	-0.264*** (0.088)
Diversity	-0.035 (0.023)	-0.042 (0.149)	0.001** (0.000)	-0.282*** (0.081)
Observations	1018			
Mental health	a	b	$a \times b$	c^{net}
Size	-4.244*** (0.428)	-0.023** (0.011)	0.099*** (0.002)	-0.167 (0.112)
Frequency	-2.664*** (0.523)	-0.005 (0.007)	0.013** (0.001)	-0.081 (0.090)
Diversity	-0.035 (0.024)	-0.191 (0.151)	0.007** (0.001)	-0.074 (0.083)
Observations	1012			

Note: Each row reports the Fixed-Effect Instrumental Variable estimation results of equations corresponding to Fig. 1.2. The sample is limited to pre-retirement healthy individuals in all these four health outcomes. Instrument is a dummy variable with value 1 if individual is older than 65 years old, and 0 otherwise. Tables A.3 and A.4 in Appendix A report the detailed estimation results. Robust individuals-clustered standard errors are in parenthesis. The significance level is defined as follows: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 1.4: Single mediation analysis: FE-IV estimation of mediatory impact of social network characteristic on retirement's effect on depression and anxiety

	Retirement impact on SNW	Social Network impact on Health	Mediation impact of SNW	Direct effect of retirement on health
Depression	a	b	$a \times b$	c^{net}
Size	-4.304*** (0.424)	-0.053 (0.042)	0.234** (0.006)	0.733* (0.409)
Frequency	-2.747*** (0.517)	-0.001 (0.027)	0.003 (0.003)	0.961*** (0.328)
Diversity	-0.036 (0.023)	0.450 (0.553)	0.016** (0.001)	0.978*** (0.301)
Observations	1016			
Anxiety	a	b	$a \times b$	c^{net}
Size	-4.334*** (0.440)	0.007 (0.045)	-0.034** (0.010)	1.350*** (0.438)
Frequency	-2.775*** (0.532)	0.033 (0.029)	-0.089** 0.000	1.412*** (0.350)
Diversity	-0.038* (0.023)	-0.374 (0.620)	0.013** (0.002)	1.306*** (0.324)
Observations	966			

Note: Each row reports the Fixed-Effect Instrumental Variable estimation results of equations corresponding to Fig. 1.2. The sample is limited to pre-retirement healthy individuals in all these four health outcomes. Instrument is a dummy variable with value 1 if individual is older than 65 years old, and 0 otherwise. Tables A.5 and A.6 in Appendix A report the detailed estimation results. Robust individuals-clustered standard errors are in parenthesis. The significance level is defined as follows: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 1.5: Parallel mediation analysis: FE-IV estimation of mediatory impact of social network characteristics on retirement's effect on physical and mental health

Health Outcomes		Mediation effects			
		Size	Frequency	Diversity	Total
	c^{net}	$a_1 \times b_1$	$a_2 \times b_2$	$a_3 \times b_3$	indirect effect
Physical health	-0.174*	-0.127**	0.013**	0.007**	-0.106**
	(0.097)	(0.002)	(0.001)	(0.000)	(0.002)
Mental health	-0.168	0.117**	-0.018**	0.002**	0.101**
	(0.118)	(0.002)	(0.001)	(0.000)	(0.002)
Depression	0.661*	0.431**	-0.096**	-0.033**	0.302**
	(0.400)	(0.007)	(0.003)	(0.001)	(0.004)
Anxiety	1.355***	0.067**	-0.118**	0.017**	-0.034**
	(0.455)	(0.009)	(0.003)	(0.001)	(0.005)

Note: Each row reports the Fixed-Effect Instrumental Variable estimation results of equations corresponding to Fig. 1.3. The sample is limited to pre-retirement healthy individuals in all these four health outcomes. Instrument is a dummy variable with value 1 if individual is older than 65 years old, and 0 otherwise. Tables A.7, A.8, A.9, and A.10 in Appendix A report the detailed estimation results. Robust individuals-clustered standard errors are in parenthesis. The significance level is defined as follows: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Bibliography

Allen, J., Balfour, R., Bell, R., and Marmot, M. (2014). Social determinants of mental health. *International review of psychiatry*, 26(4):392–407.

Barnett, I., Guell, C., and Ogilvie, D. (2012). The experience of physical activity and the transition to retirement: a systematic review and integrative synthesis of qualitative and quantitative evidence. *International Journal of Behavioral Nutrition and Physical Activity*, 9(1):97.

Baron, R. M. and Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6):1173.

Behncke, S. (2012). Does retirement trigger ill health? *Health economics*, 21(3):282–300.

Berkman, L. F., Glass, T., Brissette, I., and Seeman, T. E. (2000). From social integration to health: Durkheim in the new millennium . *Social science & medicine*, 51(6):843–857.

Berkman, L. F. and Syme, S. L. (1979). Social networks, host resistance, and mortality: a nine-year follow-up study of alameda county residents. *American journal of Epidemiology*, 109(2):186–204.

Bound, J., Stinebrickner, T., and Waidmann, T. (2010). Health, economic resources and the work decisions of older men. *Journal of Econometrics*, 156(1):106–129.

Bound, J. and Waidmann, T. (2007). Estimating the health effects of retirement.

Brock, W. A. and Durlauf, S. N. (2001). Interactions-based models. In *Handbook of econometrics*, volume 5, pages 3297–3380. Elsevier.

Calvo, E., Sarkisian, N., and Tamborini, C. R. (2011). Causal effects of retirement timing on subjective well-being: The role of cultural norms and institutional policies. *Doc-umento Trabajo (7.)*.

Capatina, E. (2015). Life-cycle effects of health risk. *Journal of Monetary Economics*, 74:67–88.

Carolan, B. V. (2013). *Social network analysis and education: Theory, methods & applications*. Sage Publications.

Carstensen, L. L. (1991). Selectivity theory: Social activity in life-span context. *Annual review of gerontology and geriatrics*, 11(1):195–217.

Charles, K. K. (2002). Is retirement depressing?: Labor force inactivity and psychological well-being in later life. Technical report, National Bureau of Economic Research.

Coe, N. B. and Lindeboom, M. (2008). Does retirement kill you? evidence from early retirement windows.

Coe, N. B. and Zamarro, G. (2011). Retirement effects on health in europe. *Journal of health economics*, 30(1):77–86.

Cohen, S. (2004). Social relationships and health. *American psychologist*, 59(8):676.

Coile, C. and Gruber, J. (2000). Social security and retirement. Technical report, National Bureau of Economic Research.

Dave, D., Rashad, R. I., and Spasojevic, J. (2008). The effects of retirement on physical and mental health outcomes. *Southern Economic Journal*, 75(2):497–523.

Eibich, P. (2015). Understanding the effect of retirement on health: Mechanisms and heterogeneity. *Journal of health economics*, 43:1–12.

Fiori, K. L. and Jager, J. (2012). The impact of social support networks on mental and physical health in the transition to older adulthood: A longitudinal, pattern-centered approach. *International Journal of Behavioral Development*, 36(2):117–129.

Fleishman, J. A. and Zuvekas, S. H. (2007). Global self-rated mental health: associations with other mental health measures and with role functioning. *Medical Care*, pages 602–609.

Gorry, A., Gorry, D., and Slavov, S. N. (2018). Does retirement improve health and life satisfaction? *Health economics*, 27(12):2067–2086.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political economy*, 80(2):223–255.

Gustman, A. L. and Steinmeier, T. L. (2018). The role of health in retirement . In *Transitions through the Labor Market: Work, Occupation, Earnings and Retirement*, pages 229–297. Emerald Publishing Limited.

Harkonmäki, K., Rahkonen, O., Martikainen, P., Silventoinen, K., and Lahelma, E. (2006). Associations of sf-36 mental health functioning and work and family related factors with intentions to retire early among employees. *Occupational and environmental medicine*, 63(8):558–563.

Insler, M. (2014). The health consequences of retirement. *Journal of Human Resources*, 49(1):195–233.

Israel, B. A. (1982). Social networks and health status: Linking theory, research, and practice. *Patient counselling and health education*, 4(2):65–79.

Johnston, D. W. and Lee, W.-S. (2009). Retiring to the good life? the short-term effects of retirement on health. *Economics Letters*, 103(1):8–11.

Knoke, D. and Yang, S. (2008). Network fundamentals. *Social network analysis*, (154):3–14.

Kofi Charles, K. (2004). Is retirement depressing?: Labor force inactivity and psychological well-being in later life. In *Accounting for Worker Well-Being*, pages 269–299. Emerald Group Publishing Limited.

Krause, N. and Rook, K. S. (2003). Negative interaction in late life: Issues in the stability and generalizability of conflict across relationships. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 58(2):P88–P99.

Litwin, H. and Shiovitz-Ezra, S. (2010). Social network type and subjective well-being in a national sample of older americans. *The Gerontologist*, 51(3):379–388.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3):531–542.

Mawani, F. N., Gilmour, H., et al. (2010). Validation of self-rated mental health. *Health Rep*, 21(3):61–75.

McGarry, K. (2004). Health and retirement do changes in health affect retirement expectations? *Journal of Human Resources*, 39(3):624–648.

Moffitt, R. A. et al. (2001). Policy interventions, low-level equilibria, and social interactions. *Social dynamics*, 4(45-82):6–17.

Neuman, K. (2008). Quit your job and get healthier? the effect of retirement on health. *Journal of Labor Research*, 29(2):177–201.

Nishimura, Y., Oikawa, M., and Motegi, H. (2018). What explains the difference in the effect of retirement on health? evidence from global aging data. *Journal of Economic Surveys*, 32(3):792–847.

Petrou, S. and Kupek, E. (2008). Social capital and its relationship with measures of health status: evidence from the health survey for england 2003. *Health economics*, 17(1):127–143.

Radloff, L. S. (1977). The ces-d scale: A self-report depression scale for research in the general population. *Applied psychological measurement*, 1(3):385–401.

Ronconi, L., Brown, T. T., and Scheffler, R. M. (2012). Social capital and self-rated health in argentina. *Health economics*, 21(2):201–208.

Rook, K. S. (1998). Investigating the positive and negative sides of personal relationships: Through a lens darkly?

Ruhm, C. J. (1995). Secular changes in the work and retirement patterns of older men. *Journal of Human Resources*, pages 362–385.

Sugisawa, A., Sugisawa, H., Nakatani, Y., and Shibata, H. (1997). Effect of retirement on mental health and social well-being among elderly japanese. *[Nihon koshu eisei zasshi] Japanese journal of public health*, 44(2):123.

Terhell, E. L., Van Groenou, M. I. B., and Van Tilburg, T. (2007). Network contact changes in early and later postseparation years. *Social networks*, 29(1):11–24.

van den Berg, T. I., Elders, L. A., and Burdorf, A. (2010). Influence of health and work on early retirement. *Journal of occupational and environmental medicine*, 52(6):576–583.

Zantinge, E. M., van den Berg, M., Smit, H. A., and Picavet, H. S. J. (2013). Retirement and a healthy lifestyle: opportunity or pitfall? a narrative review of the literature. *The European Journal of Public Health*, 24(3):433–439.

Zhao, X., Lynch Jr, J. G., and Chen, Q. (2010). Reconsidering baron and kenny: Myths and truths about mediation analysis. *Journal of consumer research*, 37(2):197–206.

Zigmond, A. S. and Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta psychiatrica scandinavica*, 67(6):361–370.

Appendix A

(First Chapter Appendix)

Table A.1: Ordinary Least Square estimation of impact of social network characteristics on health outcomes

	Physical Health	Mental Health	Depression	Anxiety
Size[a]	0.031***	-0.018*	-0.077**	-0.04
	(0.01)	(0.01)	(0.03)	(0.04)
Frequency[b]	-0.016**	0.004	0.034*	0.007
	(0.01)	(0.01)	(0.02)	(0.02)
Diversity[c]	-0.162	0.147	0.336	0.212
	(0.15)	(0.14)	(0.47)	(0.48)
Female	0.135**	-0.031	0.179	0.177
	(0.06)	(0.05)	(0.19)	(0.20)
Age	-0.010**	-0.006	0.054***	0.065***
	(0.00)	(0.00)	(0.01)	(0.01)
High Education[d]	0.133**	0.128**	-0.096	-0.431*
	(0.06)	(0.06)	(0.20)	(0.22)
Income	0.080***	0.094***	-0.102	0.003
	(0.03)	(0.03)	(0.10)	(0.11)
Married[e]	0.11	-0.045	-0.596**	0.376
	(0.07)	(0.06)	(0.24)	(0.23)
Number of drinks	0.060***	-0.001	0.008	-0.101
	(0.02)	(0.02)	(0.06)	(0.06)
Smoking cigarette	-0.026	-0.139*	0.181	0.238
	(0.09)	(0.08)	(0.35)	(0.30)
Constant	3.979***	4.316***	10.175***	-1.321
	(0.34)	(0.33)	(1.04)	(1.18)
Observation	1018	1012	1016	966

Note: This Table reports the OLS estimation results of social network characteristics on different health outcomes.[a] Size refers to the number of members in the individual's social network. [b] Frequency refers to the frequency of contacts with members of one's social network. [c] Diversity refers to diversity in the types of relationships in individual's social network that is measured by Index of Qualitative Variation (IQV). [d] Education levels equal to some college or higher is considered as high education. [e] Married is a dummy variable that takes value of 1 if individual is married, and 0 otherwise. The sample is limited to pre-retirement healthy individuals in all these four health outcomes. Robust individuals-clustered standard errors are in parenthesis. The significance level is defined as follows: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table A.2: FE-IV estimation of total impact of retirement on health outcomes

	Physical Health	Mental Health	Depression	Anxiety
Retirement[a]	-0.280***	-0.067	0.962***	1.320***
	(0.07)	(0.09)	(0.28)	(0.31)
Income	-0.007	-0.071	0.032	0.387**
	(0.04)	(0.04)	(0.15)	(0.15)
Married[b]	0.12	0.251**	-2.220***	-0.853
	(0.12)	(0.13)	(0.56)	(0.58)
Number of drinks	0.044*	0.036	-0.167	-0.13
	(0.03)	(0.03)	(0.11)	(0.09)
Smoking cigarette	0.194	-0.186	-1.077*	-1.103
	(0.23)	(0.18)	(0.61)	(0.80)
Constant	3.686***	4.123***	14.826***	2.514***
	(0.14)	(0.16)	(0.57)	(0.60)
Observations	1018	1012	1016	966

Note: This Table reports the Fixed-Effect Instrumental Variable estimation of retirement on health outcomes. Instrument is a dummy variable with value 1 if individual is older than 65 years old, and 0 otherwise.[a] Retirement is 1 if the individual reports retired and not working, and 0 otherwise. [b] Married is a dummy variable that takes value of 1 if individual is married, and 0 otherwise. The sample is limited to pre-retirement healthy individuals in all these four health outcomes. Robust individuals-clustered standard errors are in parenthesis. The significance level is defined as follows: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table A.3: Single mediation analysis: FE-IV estimation of mediatory impact of social network characteristic on retirement's effect on physical health

	Size[a]	Physical Health	Frequency[b]	Physical Health	Diversity[c]	Physical Health
Retirement	-4.300***	-0.188*	-2.739***	-0.264***	-0.035	-0.282***
	(0.42)	(0.11)	(0.52)	(0.09)	(0.02)	(0.08)
Size		0.021*				
		(0.01)				
Frequency				0.006		
				(0.01)		
Diversity						-0.042
						(0.15)
Income	0.35	-0.015	0.162	-0.008	0.007	-0.007
	(0.22)	(0.04)	(0.27)	(0.04)	(0.01)	(0.04)
Married[d]	0.529	0.109	2.619***	0.104	0.099***	0.124
	(0.66)	(0.12)	(0.80)	(0.13)	(0.04)	(0.13)
Number of drinks	0.084	0.042	0.025	0.044	-0.005	0.044
	(0.16)	(0.03)	(0.19)	(0.03)	(0.01)	(0.03)
Smoking cigarette	-0.065	0.196	-0.136	0.195	0.107**	0.199
	(0.86)	(0.16)	(1.05)	(0.16)	(0.05)	(0.17)
Constant	8.271***	3.509***	12.796***	3.608***	0.568***	3.710***
	(0.80)	(0.20)	(0.98)	(0.19)	(0.04)	(0.18)
Observations	1018	1018	1018	1018	1018	1018

Note: This Table presents the Fixed-Effect Instrumental Variable estimation results of single mediation equations corresponding to Fig. 1.2. Instrument is a dummy variable with value 1 if individual is older than 65 years old, and 0 otherwise. For instance, column 2 presents the estimation result of retirement on the size of social network and column 3 presents the estimation of retirement and size of social network on physical health. [a] Size refers to the number of members in the individual's social network. [b] Frequency refers to the frequency of contacts with members of one's social network. [c] Diversity refers to diversity in the types of relationships in individual's social network that is measured by Index of Qualitative Variation (IQV). [d] Married is a dummy variable that takes value of 1 if individual is married, and 0 otherwise. The sample is limited to pre-retirement healthy individuals in all health outcomes in this study. Robust individuals-clustered standard errors are in parenthesis. The significance level is defined as follows: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table A.4: Single mediation analysis: FE-IV estimation of mediatory impact of social network characteristic on retirement's effect on mental health

	Size[a]	Mental Health	Frequency[b]	Mental Health	Diversity[c]	Mental Health
Retirement	-4.244***	-0.167	-2.664***	-0.081	-0.035	-0.074
	(0.43)	(0.11)	(0.52)	(0.09)	(0.02)	(0.08)
Size		-0.023**				
		(0.01)				
Frequency				-0.005		
				(0.01)		
Diversity						-0.191
						(0.15)
Income	0.313	-0.064	0.059	-0.071	0.008	-0.07
	(0.22)	(0.04)	(0.27)	(0.04)	(0.01)	(0.04)
Married[d]	0.574	0.264**	2.700***	0.264**	0.098***	0.269**
	(0.66)	(0.13)	(0.80)	(0.13)	(0.04)	(0.13)
Number of drinks	0.081	0.038	0.017	0.036	-0.005	0.035
	(0.16)	(0.03)	(0.19)	(0.03)	(0.01)	(0.03)
Smoking cigarette	-0.037	-0.187	-0.089	-0.186	0.107**	-0.165
	(0.86)	(0.17)	(1.05)	(0.17)	(0.05)	(0.17)
Constant	8.308***	4.318***	12.963***	4.189***	0.564***	4.231***
	(0.80)	(0.20)	(0.98)	(0.20)	(0.04)	(0.18)
Observations	1012	1012	1012	1012	1012	1012

Note: This Table presents the Fixed-Effect Instrumental Variable estimation results of single mediation equations corresponding to Fig. 1.2. Instrument is a dummy variable with value 1 if individual is older than 65 years old, and 0 otherwise. For instance, column 2 presents the estimation result of retirement on the size of social network and column 3 presents the estimation of retirement and size of social network on mental health. [a] Size refers to the number of members in the individual's social network. [b] Frequency refers to the frequency of contacts with members of one's social network. [c] Diversity refers to diversity in the types of relationships in individual's social network that is measured by Index of Qualitative Variation (IQV). [d] Married is a dummy variable that takes value of 1 if individual is married, and 0 otherwise. The sample is limited to pre-retirement healthy individuals in all health outcomes in this study. Robust individuals-clustered standard errors are in parenthesis. The significance level is defined as follows: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table A.5: Single mediation analysis: FE-IV estimation of mediatory impact of social network characteristic on retirement's effect on Depression

	Size[a]	Physical Health	Frequency[b]	Physical Health	Diversity[d]	Physical Health
Retirement	-4.304***	0.733*	-2.747***	0.961***	-0.036	0.978***
	(0.42)	(0.41)	(0.52)	(0.33)	(0.02)	(0.30)
Size		-0.053				
		(0.04)				
Frequency				-0.001		
				(0.03)		
Diversity						0.45
						(0.55)
Income	0.331	0.05	0.095	0.032	0.007	0.029
	(0.22)	(0.16)	(0.27)	(0.16)	(0.01)	(0.16)
Married[d]	0.555	-2.191***	2.806***	-2.219***	0.091**	-2.261***
	(0.67)	(0.47)	(0.81)	(0.47)	(0.04)	(0.47)
Number of drinks	0.083	-0.163	0.032	-0.167	-0.005	-0.165
	(0.16)	(0.11)	(0.19)	(0.11)	(0.01)	(0.11)
Smoking cigarette	-0.066	-1.081*	-0.149	-1.077*	0.107**	-1.125*
	(0.86)	(0.61)	(1.05)	(0.61)	(0.05)	(0.61)
Constant	8.306***	15.268***	12.828***	14.833***	0.574***	14.567***
	(0.80)	(0.74)	(0.98)	(0.71)	(0.04)	(0.66)
Observations	1016	1016	1016	1016	1016	1016

Note: This Table presents the Fixed-Effect Instrumental Variable estimation results of single mediation equations corresponding to Fig. 1.2. Instrument is a dummy variable with value 1 if individual is older than 65 years old, and 0 otherwise. For instance, column 2 presents the estimation result of retirement on the size of social network and column 3 presents the estimation of retirement and size of social network on Depression. [a] Size refers to the number of members in the individual's social network. [b] Frequency refers to the frequency of contacts with members of one's social network. [c] Diversity refers to diversity in the types of relationships in individual's social network that is measured by Index of Qualitative Variation (IQV). [d] Married is a dummy variable that takes value of 1 if individual is married, and 0 otherwise. The sample is limited to pre-retirement healthy individuals in all health outcomes in this study. Robust individuals-clustered standard errors are in parenthesis. The significance level is defined as follows: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table A.6: Single mediation analysis: FE-IV estimation of mediatory impact of social network characteristic on retirement's effect on Anxiety

	Size[a]	Mental Health	Frequency[b]	Mental Health	Diversity[c]	Mental Health
Retirement	-4.334***	1.350***	-2.775***	1.412***	-0.038*	1.306***
	(0.44)	(0.44)	(0.53)	(0.35)	(0.02)	(0.32)
Size		0.007				
		(0.05)				
Frequency				0.033		
				(0.03)		
Diversity						-0.374
						(0.62)
Income	0.338	0.385**	0.074	0.385**	0.008	0.391**
	(0.23)	(0.17)	(0.28)	(0.17)	(0.01)	(0.17)
Married[d]	0.636	-0.857*	2.673***	-0.941*	0.101***	-0.815
	(0.69)	(0.50)	(0.83)	(0.50)	(0.04)	(0.51)
Number of drinks	0.046	-0.13	-0.065	-0.128	-0.007	-0.132
	(0.16)	(0.12)	(0.19)	(0.12)	(0.01)	(0.12)
Smoking cigarette	0.081	-1.104*	-0.183	-1.097*	0.104**	-1.065
	(0.90)	(0.66)	(1.09)	(0.66)	(0.05)	(0.66)
Constant	8.240***	2.456***	13.048***	2.083***	0.568***	2.726***
	(0.82)	(0.78)	(0.99)	(0.76)	(0.04)	(0.70)
Observations	966	966	966	966	966	966

Note: This Table presents the Fixed-Effect Instrumental Variable estimation results of single mediation equations corresponding to Fig. 1.2. Instrument is a dummy variable with value 1 if individual is older than 65 years old, and 0 otherwise. For instance, column 2 presents the estimation result of retirement on the size of social network and column 3 presents the estimation of retirement and size of social network on Anxiety. [a] Size refers to the number of members in the individual's social network. [b] Frequency refers to the frequency of contacts with members of one's social network. [c] Diversity refers to diversity in the types of relationships in individual's social network that is measured by Index of Qualitative Variation (IQV). [d] Married is a dummy variable that takes value of 1 if individual is married, and 0 otherwise. The sample is limited to pre-retirement healthy individuals in all health outcomes in this study. Robust individuals-clustered standard errors are in parenthesis. The significance level is defined as follows: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table A.7: Parallel mediation analysis: FE-IV estimation of mediatory impact of social network characteristics on retirement's effect on physical health

	Size[a]	Frequency[b]	Diversity[c]	Physical Health
Retirement	-4.300***	-2.739***	-0.035	-0.174*
	(0.35)	(0.52)	(0.02)	(0.10)
Size				0.029**
				(0.01)
Frequency				-0.005
				(0.01)
Diversity				-0.205
				(0.18)
Income	0.35	0.162	0.007	-0.016
	(0.22)	(0.27)	(0.01)	(0.04)
Married[d]	0.529	2.619***	0.099**	0.137
	(0.67)	(0.76)	(0.04)	(0.12)
Number of drinks	0.084	0.025	-0.005	0.041
	(0.18)	(0.18)	(0.01)	(0.03)
Smoking cigarette	-0.065	-0.136	0.107**	0.218
	(0.63)	(0.86)	(0.05)	(0.23)
Constant	8.271***	12.796***	0.568***	3.620***
	(0.80)	(1.00)	(0.05)	(0.20)
Observations	1018	1018	1018	1018

Note: This Table presents the Fixed-Effect Instrumental Variable estimation results of parallel mediation equations corresponding to Fig. 1.3 with physical health as the outcome variable. Instrument is a dummy variable with value 1 if individual is older than 65 years old, and 0 otherwise. Columns 2-4 presents the FE-IV estimation of social network variables on the retirement, corresponding to the paths in Fig. 1.3. [a] Size refers to the number of members in the individual's social network. [b] Frequency refers to the frequency of contacts with members of one's social network. [c] Diversity refers to diversity in the types of relationships in individual's social network that is measured by Index of Qualitative Variation (IQV). [d] Married is a dummy variable that takes value of 1 if individual is married, and 0 otherwise. The sample is limited to pre-retirement healthy individuals in all health outcomes in this study. Robust individuals-clustered standard errors are in parenthesis. The significance level is defined as follows: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table A.8: Parallel mediation analysis: FE-IV estimation of mediatory impact of social network characteristics on retirement's effect on mental health

	Size [a]	Frequency[b]	Diversity[c]	Mental Health
Retirement	-4.244***	-2.664***	-0.035	-0.168
	(0.36)	(0.53)	(0.02)	(0.12)
Size				-0.028**
				(0.01)
Frequency				0.007
				(0.01)
Diversity				-0.047
				(0.18)
Income	0.313	0.059	0.008	-0.063
	(0.22)	(0.28)	(0.01)	(0.04)
Married[d]	0.574	2.700***	0.098**	0.253*
	(0.71)	(0.78)	(0.04)	(0.13)
Number of drinks	0.081	0.017	-0.005	0.038
	(0.18)	(0.18)	(0.01)	(0.03)
Smoking cigarette	-0.037	-0.089	0.107**	-0.181
	(0.61)	(0.88)	(0.05)	(0.19)
Constant	8.308***	12.963***	0.564***	4.290***
	(0.80)	(1.01)	(0.05)	(0.22)
Observations	1012	1012	1012	1012

Note: This Table presents the Fixed-Effect Instrumental Variable estimation results of parallel mediation equations corresponding to Fig. 1.3 with mental health as the outcome variable. Instrument is a dummy variable with value 1 if individual is older than 65 years old, and 0 otherwise. Columns 2-4 presents the FE-IV estimation of social network variables on the retirement, corresponding to the paths in Fig. 1.3. [a] Size refers to the number of members in the individual's social network. [b] Frequency refers to the frequency of contacts with members of one's social network. [c] Diversity refers to diversity in the types of relationships in individual's social network that is measured by Index of Qualitative Variation (IQV). [d] Married is a dummy variable that takes value of 1 if individual is married, and 0 otherwise. The sample is limited to pre-retirement healthy individuals in all health outcomes in this study. Robust individuals-clustered standard errors are in parenthesis. The significance level is defined as follows: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table A.9: Parallel mediation analysis: FE-IV estimation of mediatory impact of social network characteristics on retirement's effect on depression

	Size [a]	Frequency[b]	Diversity[c]	Depression
Retirement	-4.304***	-2.747***	-0.036*	0.661*
	(0.35)	(0.52)	(0.02)	(0.40)
Size				-0.100**
				(0.05)
Frequency				0.035
				(0.03)
Diversity				0.934
				(0.64)
Income	0.331	0.095	0.007	0.056
	(0.22)	(0.27)	(0.01)	(0.15)
Married[d]	0.555	2.806***	0.091**	-2.347***
	(0.69)	(0.79)	(0.04)	(0.58)
Number of drinks	0.083	0.032	-0.005	-0.156
	(0.18)	(0.18)	(0.01)	(0.13)
Smoking cigarette	-0.066	-0.149	0.107**	-1.179*
	(0.63)	(0.86)	(0.05)	(0.61)
Constant	8.306***	12.828***	0.574***	14.673***
	(0.80)	(1.00)	(0.05)	(0.84)
Observations	1016	1016	1016	1016

Note: This Table presents the Fixed-Effect Instrumental Variable estimation results of parallel mediation equations corresponding to Fig. 1.3 with depression as the outcome variable. Instrument is a dummy variable with value 1 if individual is older than 65 years old, and 0 otherwise. Columns 2-4 presents the FE-IV estimation of social network variables on the retirement, corresponding to the paths in Fig. 1.3. [a] Size refers to the number of members in the individual's social network. [b] Frequency refers to the frequency of contacts with members of one's social network. [c] Diversity refers to diversity in the types of relationships in individual's social network that is measured by Index of Qualitative Variation (IQV). [d] Married is a dummy variable that takes value of 1 if individual is married, and 0 otherwise. The sample is limited to pre-retirement healthy individuals in all health outcomes in this study. Robust individuals-clustered standard errors are in parenthesis. The significance level is defined as follows: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table A.10: Parallel mediation analysis: FE-IV estimation of mediatory impact of social network characteristics on retirement's effect on anxiety

	Size[a]	Frequency[b]	Diversity[c]	Anxiety
Retirement	-4.334***	-2.775***	-0.038*	1.355***
	(0.37)	(0.55)	(0.02)	(0.46)
Size				-0.015
				(0.06)
Frequency				0.042
				(0.04)
Diversity				-0.43
				(0.67)
Income	0.338	0.074	0.008	0.393**
	(0.22)	(0.28)	(0.01)	(0.16)
Married[d]	0.636	2.673***	0.101**	-0.913
	(0.78)	(0.86)	(0.05)	(0.61)
Number of drinks	0.046	-0.065	-0.007	-0.129
	(0.18)	(0.18)	(0.01)	(0.10)
Smoking cigarette	0.081	-0.183	0.104*	-1.05
	(0.67)	(0.85)	(0.06)	(0.81)
Constant	8.240***	13.048***	0.568***	2.331***
	(0.84)	(1.07)	(0.05)	(0.77)
Observations	966	966	966	966

Note: This Table presents the Fixed-Effect Instrumental Variable estimation results of parallel mediation equations corresponding to Fig. 1.3 with anxiety as the outcome variable. Instrument is a dummy variable with value 1 if individual is older than 65 years old, and 0 otherwise. Columns 2-4 presents the FE-IV estimation of social network variables on the retirement, corresponding to the paths in Fig. 1.3. [a] Size refers to the number of members in the individual's social network. [b] Frequency refers to the frequency of contacts with members of one's social network. [c] Diversity refers to diversity in the types of relationships in individual's social network that is measured by Index of Qualitative Variation (IQV). [d] Married is a dummy variable that takes value of 1 if individual is married, and 0 otherwise. The sample is limited to pre-retirement healthy individuals in all health outcomes in this study. Robust individuals-clustered standard errors are in parenthesis. The significance level is defined as follows: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Chapter 2

Differential impact of social distancing on COVID-19 spread in the US: by rurality and social vulnerability

2.1 Introduction

By the end of July 2020, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has infected more than 4.6 million individuals and caused more than 155,000 deaths in the US. The outbreak is not only threatening the healthcare system globally but also devastating all segments of the global economy such as high employment rate, sharp drop of GDPs, widened disparities, etc. (Baker et al., 2020; Cajner et al., 2020; Coibion et al., 2020; Dunn et al., 2020; Luo and Tsang, 2020). The current knowledge of the SARS-CoV-2 reveal that the virus is easily transmitted via person-to-person contact and is sustained for an alarmingly long period of time.[1] In the absence of effective and safe vaccines and pharmacological solutions, social distancing remains the most viable option to stop the speed of the pandemic. Therefore, procedures that aim at minimizing people contact have been implemented worldwide as the first line of defense such as emergency lock-downs of hot spots, none-essential businesses closures, shelter-in-place and stay-at-home orders, etc. According to US Centers for Disease

[1]https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/social-distancing.html

Control and Prevention (CDC), all those procedures that achieve social distancing among communities and throughout countries are the best way to prevent the spread of COVID-19.

The social distancing measure at individual level usually refers to the recommendation of maintaining a minimum of 6-feet distance between each other when interacting face-to-face. At a group level (i.e., community, county, State etc.), it generally contains a composite of several physical mobility limitations: stay-at-home order and minimizing work and non-work trips. There are other interventions implemented aiming at limiting social gatherings through targeting restaurants, entertainment venues, and schools, etc. All those policies intend to target the direct transmission channel of the virus with the obvious costs of economy disruption.

Although the observed evidence all support the use of social distancing, there is limited quantitative conclusion on how effective it is in fighting the spread of the virus. A few studies have investigated the impact of various social distancing based interventions on the spread of COVID-19 infection. In particular, a study finds that shelter-in-place order along with closure of bars, restaurants and entertainment outlets on average reduces daily growth rate of confirmed COVID-19 cases by 5.4% after 1–5 days, 6.8% after 6–10 days, 8.2% after 11–15 days, and 9.1% after 16–20 days across US counties (Courtemanche et al., 2020).

However, it has been recognized that it is difficult to establish causal inferences in this context due to the presence of spatial and temporal confounders that influence the speed of the spread simultaneously alongside with the social distancing (Courtemanche et al., 2020). Examples of such confounders are: state/county specific prevention measures such as requirement for wearing masks, restrictions on international and domestic travel, testing, contact tracing, and local norms and perceptions towards social distancing, etc. There is a significant heterogeneity in levels of compliance to social distancing directives throughout the US and the distribution is shown to be correlated with local norms, perceptions, social-economic profiles and political leanings (Wright et al., 2020; Lou et al., 2020; Bodas and Peleg, 2020).

A lot of these confounders are either unobserved or difficult to measure directly. In this study we examine the impact of social distancing on the growth rate of daily confirmed COVID-19 cases at county level. To address the challenges of directly measuring and controlling confounders, we implement Spatial Durbin Models with county fixed-effects to indirectly control for time-invariant and county-specific unobservables while accounting for spatial dependence between counties. The spatial econometric model indirectly controls for spatial confounders through recognizing the spatial contiguity of counties and through county level fixed effects. The models acknowledge that infectious diseases, such as COVID-19, follow spatial patterns i.e., the infection rates are correlated with geographic proximity and connectivity through mobility among communities (Langford, 2002).

Furthermore, COVID-19 pandemic has been found to disproportionately hit the low income, minority and vulnerable populations. Low-income sub-populations have exhibited lower levels of compliance to social distancing due to occupation specific inflexibility (Wright et al., 2020; Lou et al., 2020). Stay-at-home orders did not significantly alter work trip patterns for essential businesses such as grocery stores, garbage collection, postal services, construction work, etc. which account for majority of the low-income populations' employment (Lou et al., 2020; Drago and Miller, 2010). The lack of compliance to social distancing along with health care access inequalities result in widened infection risk gaps across socio-economic and racial-ethnicity subgroups (Link, 2008; Leclere et al., 1994). For example, recent research showed that counties with higher minority population and lower socio-economic status experienced significantly higher COVID-19 death and infection rates (Nayak et al., 2020).

To quantify the social distancing impact heterogeneity, this study conducts subgroup analysis to examine differences between urban vs. rural and between those areas with high vulnerability vs those areas with low vulnerability. This sensitivity analysis is motivated by the fact that region-specific factors such as healthcare access, population density, racial/ethnicity composition can be quite different when examining urban vs. rural. For instance, rural areas are characterized by sparse population and lower housing density in contradiction to urban areas .[2] Therefore it is not surprising that regional patterns of disease spreading and social distancing compliance will be different. Similarly those differences are expected to be observed in low versus high vulnerability areas, as characterized by the social vulnerability index (SVI) provided by the CDC. The SVI for each US county takes into account community level poverty, transportation access, disabilities and housing composition, among other variables.[3] It provides another way to examine potential subgroup differences of social distancing policy on the spread of COVID-19 beyond traditional dimension of mere population size. Moreover, quantifying impact heterogeneity across areas may provide evidence for an area-based customized policy over a "one-size-fits-all" approach.

Figure 2.1 shows cumulative infections at the beginning (left) and at the end of the study period (right) for urban vs. rural counties. Figure 2.2 shows cumulative infections at the beginning (left) and at the end of the study period (right) for high vs. low SVI counties. As can be seen in these plots, the infected areas are highly spatially clustered, meaning adjacent counties experience similar levels of infections. Therefore, conventional regression models (e.g. Ordinary Least Squares) that impose independence assumption on observations and ignore the spatial dependence among them will produce biased and inconsistent inferences (Wooldridge, 2016). Another reason is that spatially adjacent counties are more likely to share similar norms and perceptions which can include acceptance towards social distancing. Therefore in the context of infectious disease such as COVID-19 that is spread primarily via close contact, it is important to quantify and disentangle direct effects (within county

[2] https://www.census.gov/library/stories/2017/08/rural-america.html
[3] https://svi.cdc.gov/index.html

and feedback effects) and indirect effects (across counties spillover effects) for understanding the value of coordinated efforts among neighboring counties.

In this study, we examine the impact of county-level social distancing compliance on the county-level daily growth rate of confirmed COVID-19 cases. We implement spatial panel data methods to indirectly control for time-invariant and county-specific confounding unobservables while accounting for spatial dependence between counties. Our goal is not only to quantify the direct and indirect marginal effects of social distancing compliance in reducing the growth rate in infections, but also to assess the disparity in marginal effects between rural and urban areas, and between areas with high or low social vulnerability. Furthermore, by examining quartiles along the social distancing compliance distribution, we investigate the most effective levels of social distancing. Our analysis focuses on the time period corresponding to the *first wave* of COVID-19 U.S. pandemic which goes from mid-March till mid-June. We chose this time period based on the fact that procedures such as social distancing are the first line of defense in the early phase and quantitative evidence is needed for current and future infectious disease public health prevention and intervention.

Our findings support the importance of social distancing in slowing down infections during the early phase of the pandemic. Notably, the spatial estimation uncovers significant spillover effects on infection rate reduction in a typical county (i.e., rate reduction caused by social distancing compliance in its neighboring counties) and provides support for multi-county social distancing coordination efforts. Our results further reveal that those counties with high social distancing index (SDI) (i.e., those with SDI level higher than the median of the SDI distribution) on average experienced about 1.84% lower daily infection growth rates compared to those counties with low SDI. This difference is larger when comparing urban areas to rural ones (i.e., 1.88% rate differences), and even larger when comparing areas with high vs. low social vulnerability index (i.e. 2.00% rate differences). Moderate level of social distancing is found to be most effective (i.e., generating the largest rate differences as compared to those non-compliance counties) in rural and low SVI areas and contributes to reduction in daily infection growth rate by 1.5% and 1.2%, respectively. Results of this paper highlight the importance of collateral planning and coordinating with the geographically adjacent counties in flattening the epidemic curve.

2.2 Data and Summary Statistics

We use data from the "COVID-19 Impact Analysis Platform" made publicly available by the University of Maryland (UMD) (Zhang et al., 2020; Institute, 2020). This database provides information on COVID-19 related infections, deaths, social and economic indicators, social mobility, as well as testing and tracing information at a county level for the entire US on a daily basis. Specifically, it includes daily number of newly confirmed COVID-19 infections, cumulative death rate, number of active cases etc. It also provides several mobility and social distancing measures such as population movement within and out of county, percentage of residents staying at home, average number of all trips per person per day, daily average person-miles traveled, daily percentage of all trips that cross county borders, percentage of all trips that cross state borders per day, and daily social distancing index.

Furthermore, the UMD data set also contains the Social Distancing Index (SDI) developed by the Maryland Transportation Institute (MTI) using a series of county-level mobility measurements and changes compared to pre-COVID-19 levels, with highest weight given to stay-at-home levels.[4] The SDI provides a way to quantify the degree of residents' compliance to social distancing in a county. It is an integer between 0 (no social distancing at all) and

[4] SDI = 0.8*[% stay-at-home + 0.01*(100 - %stay-at-home)*(0.1*% reduction all trips + 0.2*% reduction work trips + 0.4*% reduction non-work trips + 0.3*% reduction travel distance)] + 0.2*% reduction out-of-county trips.

100 (100% of the residents in the county follow social distancing). It is also important to control for the counties' virus exposure level and testing capacity so that the social distancing effect is compared among counties with similar exposure and prevention efforts. Therefore, we control in our models the following variables: number of COVID-19 tests done per 1000 people (with a 2 days lag) and number of residents already exposed to COVID-19 (with a 14 days lag).[5]

The outcome variable of interest is the daily growth rate of confirmed COVID-19 cases. If a county had zero confirmed COVID-19 cases, we consider its infection growth rate as zero. However, since explanatory variables are entered into the model with 14-day lag, our daily growth rate of COVID-19 cases starts from late March when only 2% of the counties had zero confirmed COVID-19 cases. Even though the death count is relatively more accurate than the case incidence rate, we focus on studying the impact of the social distancing on the growth rate of infections since the social distancing measures are meant to mainly slow down the spread of the virus and not directly affect the death rate. Although the number of cases correlates with the number of deaths, the death rate depends on a variety of other factors as well such as age, comorbidities, access to health care, availability of hospital beds, access to health insurance and ventilators etc.Guan et al. (2020)

The period of analysis starts on March 19 (i.e. the start of the lock down period), and ends on June 12, 2020 (end of first wave). In order to avoid the linear and constant marginal impact assumption when using the SDI as a continuous variable, we convert the SDI into discrete categories. It is more informative to quantify the impact of SDI in a relative sense between groups since there is no evidence to support the exact 'optimal' level of SDI. For example, a binary version of SDI would measure the extent to which more compliant counties do better in terms of slowing down the infection rates compared to counties that are less compliant.

[5] Please see the detail justification of the lag choices in the last paragraph of this section

Based on these reasons, we examine two relative measures of SDI. First, we use a dummy variable that indicates whether or not the county's SDI level is greater than the median of the SDI distribution of all U.S. counties. To facilitate readability, counties located in the upper 50% of SDI distribution are called 'compliant counties' and in the bottom half of the distribution are called 'non-compliant counties'. To further explore nonlinear impact of SDI and facilitate policy-making, we group counties by SDI quartiles and represent those quartile indicators in the empirical models using dummies. We call counties in the first/lower quartile "non-compliant counties", in the second quartile "low-compliance counties", in the third quartile "moderate-compliance counties" and in the fourth/upper quartile "high-compliance counties".

To conduct subgroup analysis, we first classify counties as rural or urban using the Rural Urban Continuum Codes (2013) developed by the Economic Research Services of the Department of Agriculture and the Rural Health Research Center at the University of Washington.[6] Counties with code values of 1 to 5 are classified as urban and those with code values of 6 to 9 are considered as rural. In a total of 3,142 counties across U.S., 1,472 counties are classified as urban and 1,668 as rural. [7]

The other subgroup analysis is based on the Social Vulnerability Index (SVI). The CDC has developed the SVI for US counties to capture social determinants of health-based inequalities which in turn impact communities' ability to deal with health crisis (Marmot et al., 2008). A high value of SVI implies a high level of vulnerability in the county. It is based on 15 social factors that measure four dimensions of overall vulnerability: socioeconomic status, household composition and disability, minority status and language, housing type, and transportation. Needless to say that levels of social distancing varies with different social

[6] https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/

[7] Two counties (Kusilval Census Area in Alaska and Oglala Lakota County in South Dakota) that were not classified in RUCC 2003 are excluded from our sub-sample analysis.

vulnerability status. Therefore, it is important to understand the differential impact SDI has across counties with different SVI status. We use the 2018 SVI data to group counties into high (i.e., those with the SVI level above the median of the SVI distribution) and low (i.e., those with the SVI in the bottom half of the SVI distribution) socially vulnerable counties.[8]

Table 2.1 provides data description and summary statistics of the variables, and compares rural vs. urban and low vs. high SVI counties. On average, daily infection growth rate is 5.33% with a maximum of 2833%. During the time period of analysis, the average social distancing index level is 33.52.[9] The average percentage of daily trips across state borders is 5.96%, and the mean daily number of residents exposed to COVID-19 is 5.33. On average 23.05 COVID-19 tests were done per 1000 people across all counties. All variables exhibit statistically significant differences between rural and urban, and between low and high SVI counties. On average, rural areas see lower daily infection growth rates compared to urban areas while rural areas exhibit relatively lower SDI level as compared to urban areas. Residents in rural areas had more out of state trips than residents in urban areas. As expected, high SVI counties have larger daily infections growth rate (i.e., 5.83 vs 4.83) and lower levels of SDI as compared to low SVI ones (i.e., 32.28 vs 34.76). These disparities support our subgroup-based approach to quantify the SDI impact heterogeneity.

Our empirical analysis accounts for potential observed confounders including daily number of COVID-19 tests (per 1,000 people), number of point of interests for crowd gathering in counties (per 1,000 people), number of residents exposed to COVID-19 (per 1,000 people), and percentage of daily trips across state borders.[10] Due to 14-day incubation period of

[8] The latest available SVI is for year 2018 at the time of this study. Also, the SVI is not defined for Rio Arriba county in New Mexico which is then excluded in our SVI subgroup analysis.

[9] Maximum SDI score of 100 is only observed in three counties in Alaska and Hawaii. These observations do not affect estimations in terms of magnitude, size, and statistical significance levels therefore are included in our analysis.

[10] Percentage of daily trip across states have not been used in the calculation of SDI and that is why included here.

COVID-19, these variables as well as social distancing compliance enter into our empirical models with 14-day lag, except for number of testing which is used with a 2-day lag.[11] On average, it took about two days to get COVID-19 test results.[12] The county fixed-effect controls all county-level time-invariant characteristics such as racial and ethnicity profiles, socio-economic levels, etc.

2.3 Methodology

Spatial econometrics models are gaining popularity in the investigation of disease outbreak especially those mainly transmitted via in-person contacts (Brockmann et al., 2009). It not only provides consistent estimates of the intervention effects by accounting for spatial dependency but also has the ability to disentangle direct (which includes feedback effects) and indirect effects. Specifically, the method enables us to calculate a total direct effect of county A's social distancing on daily infection growth rate that includes the feedback/boomerang effect caused by spatial dependency (i.e., the part of its infection growth rate reduction reflected back from its neighboring counties' reduction in infection rate which in turn was caused by county A's compliance). Furthermore, we can calculate the indirect effect (or spillover effect) which is the impact of county A's neighboring counties' social distancing compliance on county A's daily growth rate of infections.

Specifically, we estimate Spatial Durbin Models with county fixed-effects to extract the spatial dependencies in daily growth rate of infections in US counties during the initial phase of COVID-19 spread (LeSage and Pace, 2009). Our model includes county fixed effects but

[11] The period of this study starts from March 19 and the first observation of the growth rate used in estimation is April 1 due to the 14-day lag of most of the control variables. The 2-day lag of the testing variable used in the model starts from March 30th. By then data on all counties have become available.

[12] For incubation period: https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html; for testing: https://www.cdc.gov/coronavirus/2019-ncov/testing/diagnostic-testing.html

not time fixed effect to allow more efficiency. Furthermore, the daily growth rate calculation has already taking time trend into consideration. Time fixed effect is not included specifically because the surge in infections follows spatial pattern rather than temporal: i.e., the rise and plateau in Seattle and west coast are not synchronized with east coast temporally. The model specification is defined as follows:

$$\mathbf{y}_{nt} = \rho \mathbf{W} \mathbf{y}_{nt} + \alpha l_n + \mathbf{X}_{nt-14}\beta + \mathbf{W}\mathbf{X}_{nt-14}\theta + \mu_n + \epsilon_{nt} \tag{2.1}$$

where, $\mathbf{y}_{nt} = (y_{1t}, y_{2t}, y_{3t}, ..., y_{nt})'$ is an $\mathbf{n} \times 1$ vector of county-specific daily growth rate of COVID-19 infections in period t. $\mathbf{W}$ is the spatial weight matrix and $\mathbf{W}\mathbf{y}_{nt}$ is the spatial lag, ρ reflects the strength of spatial dependence between counties, and l_n denotes $n \times 1$ vector of ones. $\mathbf{X}_{nt-14}$ contains time varying exogenous explanatory variables with 14 days lag to impose temporal exogeneity and avoid the simultaneity problem of social distancing compliance with COVID-19 Cases.[13] θ is a $k \times 1$ vector of parameters that measure the strength of spatial dependence of explanatory variables between counties. μ contains time-invariant observed and unobserved county-specific features and ϵ_t represents the i.i.d error term.

To capture spatial linkages between counties i and j, we use contiguity matrix that contains values of either 0 or 1, where 1 refers to the case that county i and j are adjacent (i.e., share geographic borders) and, 0 otherwise.[14] The spatial weight matrix $\mathbf{W}$ is a row-stochastic conversion of contiguity matrix of 0 and 1 that captures spatial linkages between counties i and j (with $i, j = 1, ..., n$). $\mathbf{W}$ is positive definite with zero diagonal.[15] The contiguity matrix is obtained using geographical coordinates of counties available in Census 2018-shape

[13]Only testing (per 1000 people) is used with a 2-day lag.

[14]Our results are robust across different spatial weights (i.e., inverse distance and commuting flow (2015) https://www.census.gov/topics/employment/commuting/guidance/flows.html

[15]The main diagonal of $\mathbf{W}$ has all values of zero to prevent a county from being treated as a neighbor to itself.

files.[16] We use Maximum Likelihood Estimation method to estimate parameters in spatial regression. The null hypothesis of cross-sectional independence is rejected[17] in the full sample and all sub-samples which provide statistical support for spatial specification and its importance for consistent inference. To begin, we first report estimation results based on naive pooled Ordinary Least Squares (OLS) regression which ignores spatial dependency and then compare with the spatially dependent analysis. In all empirical analysis, we present robust standard errors that recognize that the daily data is clustered within county level.

2.4 Empirical Results

2.4.1 No Spatial Dependence Assumed Between Counties

Figure 2.3a presents parameter estimates of the impact of social distancing on daily growth rate of infections at a county level, using naive pooled OLS regression which ignores spatial dependence between counties.[18] The results show that counties with SDI above median levels, on average experience 0.43% lower daily infection growth rate as compared to counties with SDI levels in the bottom half of the SDI distribution. The sub-sample analysis shows that, compared to rural counties, urban areas on average experience a larger infection growth rate (about 0.65% in urban and 0.28% in rural). The more compliant counties in socially vulnerable areas (i.e., high SVI areas) experience larger reduction in infection growth rate (about 0.71%) whereas SDI compliance does not have statistically significant impact on low SVI areas.

We further refine the SDI compliance groups by SDI distribution quartiles. Results are

[16] https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html
[17] All p-values for the tests are less than 0.001
[18] Appendix Table B.1 represents the regression estimation results in details.

presented in Figure 2.3b for four groups of counties with different relative SDI levels (the base comparison group contain counties with SDI in the bottom quartile of the distribution ('non-compliant' ones)).[19] It shows that in order to have a statistically significant effect on flattening the curve, counties need to be at least moderately compliant to social distancing (i.e., SDI level should be in or above third quartile of the distribution).[20] Similar to what was observed in Figure 2.3a where SDI took a binary value, urban areas have a larger reduction in infection rate from SDI compliance as compared to rural areas, and it has the largest impact in socially vulnerable areas (1.11% daily infection growth rate gap between high-compliance counties and non-compliant counties). However, this naive regression model does not control for spatial dependence between counties and cannot disentangle the direct and indirect effects which are of public health policy importance.

2.4.2 With Spatial Dependence Considered

Figure 2.4 presents results of empirical estimation based on Spatial Durbin Model for the US, urban, rural, high SVI and low SVI regions.[21] The direct effect shows social distancing in a county significantly reduces rate of infections within the same county. In particular, on average, social distancing compliant counties experience about 0.25% lower infection growth rate compared to non-compliant ones in the national sample.

More importantly, the indirect effect of social distancing compliance is much larger: i.e., neighboring counties' SDI as a whole will result in about 1.02% reduction in a typical county's daily infection growth rate. This indirect effect is more than four times the direct effect. The much larger indirect effect is no surprising since the indirect effect is the sum of all the

[19] Appendix Table B.1 represents the regression estimation results in details.

[20] The SDI level thresholds for each quartile are: SDI=22 at 25th percentile, SDI=31 at 50th percentile, SDI=43 at 75th percentile.

[21] Detailed results are shown in the Table B.2 in the Appendix.

neighboring counties' effect. This evidence will be masked in models that do not account for spatial dependency. The both statistically significant and public health impact relevant indirect effect of SDI compliance highlights the importance of coordinated planning among counties that share common geographic boarders to combat the spread of COVID-19. The total effect of social distancing combines the direct effect and indirect spillover effect and shows a difference of 1.28% in daily infection growth rate between compliant and non-compliant counties.

This pattern of relatively larger indirect effects persists in sub-sample analysis of urban, rural and socially vulnerable counties. Most direct and indirect effects are all statistically significant with the only exception that the direct effects are not statistically significant in rural areas and areas with low SVI. Even though rural and low SVI areas do not see statistically significant impact of their own SDI compliance on slowing down the infections (i.e., null direct effects), there are significant indirect effects on flattening the epidemic curve as a results of neighboring counties' efforts. Furthermore the magnitudes of the indirect effects are similar in urban, rural and socially vulnerable areas signaling the importance of coordinated efforts in those areas.

Socially vulnerable areas see the largest total effect of social distancing compliance (about 1.45% growth rate differences between high and low compliance counties) followed by urban areas (about 1.41% growth rate differences between high and low compliance counties). In case of the OLS regression which ignores the spatial dependency between neighboring counties, we do not find statistical evidence to support the effectiveness of social distancing compliance in low SVI areas. However, the SDM model which accounts for spatial dependency finds statistically significant indirect effects of social distancing on infection control and flattening of the curve.

To further investigate potentially heterogeneous effect of social distancing, we estimate the spatial models with three group dummies that categorize counties according to their SDI quartiles and report the results in Figure 2.5 (Details of estimation results are provided in Table B.3 in the Appendix). Results show that high-compliance counties (i.e., those with SDI in the fourth quartile of the distribution) are the ones experiencing statistically significant direct effect on reduction of infection growth rates as compared to non-compliant counties (i.e., those with SDI in the first quartile of the distribution): about 0.4% lower infection growth rate as a result of their own compliance to social distancing, compared to non-compliant counties. This pattern also holds in urban areas (about 0.73% lower infection growth rate in high compliance counties compared to non-compliant ones). Similar to the binary SDI model, rural and low SVI areas do not see statistically significant direct effect of social distancing. Noticeably, moderate and high-compliance counties show significant direct effect on infection control in socially vulnerable areas i.e., compared to non-compliance counties, there are an average of -0.43% in moderate-compliance counties and -0.80% in high-compliance counties.

Similar to the binary SDI model, all indirect spillover effects induced from social distancing compliance in neighboring counties are significant and show much larger public health impact significance (i.e. larger magnitudes). Noticeably, moderate-compliance level has the largest indirect effects and total effects on infection control in the national sample and in rural and socially less vulnerable areas, whereas high-compliance level is needed to achieve the largest effects in urban and socially vulnerable areas. Possible explanations for this result could be the fact that rural and low SVI areas have lower population density, sparse housing and more open spaces that social distancing compliance effect can be achieved without having to abide to exact measures. For example, stay-at-home may not be needed for rural areas where the population density and housing density are low to achieve crowd avoiding and physical distancing goals as compared to urban areas which have high population density and more mobile.

In socially vulnerable areas, complex interaction of socioeconomic factors (such as wage jobs, low income, inflexible work conditions, etc.) contribute to the need for higher levels of social distancing to curb the spread of COVID-19 infection. For example, people with low socioeconomic profiles are often working for essential businesses which are exempt from lock downs; or they are working in jobs that require human contact such as child-care, nursing, house-cleaning, cooking etc. which limits their ability to fully comply with social distancing regulations and results in greater exposure to risk of infection.

Both Figure 2.4 and Figure 2.5 show that the direct effect of social distancing within counties may be minimal and/or null but it produces meaningful indirect effects at lowering COVID-19 infection growth rate due to spatial dependence of counties. Figure 2.5 further confirms the importance of compliance to social distancing. Based on the total effect results, counties with moderate social distancing compliance levels show a sizable reduction in infection growth rates. Even counties that have low compliance, there is a sizable and significant reduction nationwide and in sub-samples, except for urban areas.

Considering the daily fluctuation in the growth rate of infection, we also examine the impact of social distancing compliance on a 3-day moving average of infection growth rates. As shown in Appendix Tables B.5 and B.6, results remain unchanged except a slight reduction in magnitudes.

2.5 Discussion and Conclusions

In this paper, we investigate the efficacy of social distancing in combating the COVID-19 pandemic. Using spatial econometric methods, we find significant reduction in COVID-19 infection growth rate as a result of social distancing in the US counties, especially among urban counties and counties with high SVI. Our results demonstrate significant and sizable spillover effects induced by social distancing in neighboring counties for nationwide samples and all sub-samples, even when within county effects are absent. Within county social distancing significantly lowers the growth rate of infections except in rural and low SVI areas. Our findings provide empirical evidence of spatial dependence between counties and highlight the need to coordinate planning and mitigation efforts with the geographically adjacent counties to successfully combat COVID-19 pandemic.

Policies that enforce social distancing, such as stay-at-home, are economically costly (Birge et al., 2020), and therefore we need to identify areas which are likely to benefit most from such policies. Overall, the effect of social distancing has been found to be higher in urban than in rural areas. Results indicate highest levels of social distancing compliance is needed in urban areas, whereas moderate levels of social distancing compliance will suffice in controlling the spread in rural areas, probably because of the low population and housing density in those counties. Similarly, highest levels of social distancing is needed in flattening the epidemic curve in high SVI areas but moderate levels may be sufficient in low SVI areas. The examination of areas with high and low SVI adds socio-economic and disaster readiness dimensions to the mainly population size categorization done by rural vs. urban analysis.

Those subgroup analysis will guide policy makers and public health officials in setting up appropriate levels of social distancing goals (not one-size-fits-all approach), at least until a medical solution is available to control the pandemic. It also highlights the importance of having a clear and coherent national plan that can be implemented consistently across the country. A patchwork of policies that are applied piece meal by counties and states who independently decide what is best for them, has not been working and will not work as shown by this analysis. There are a lot of spatial dependencies between regions that must be taken into account for an effective public health response.

There are a number of limitations to this analysis. First, it is well-known that number of confirmed COVID-19 cases may not represent the actual cases of COVID-19 infections due to lack of testing capacity, asymptomatic infections, access to healthcare etc. (Stock et al., 2020). For instance, it is highly likely that people with mild symptoms did not seek medical care and hence did not get counted among the confirmed cases. However, our paper findings on the marginal impact of SDI and the disparity of the impact across rural/urban and low/high vulnerability areas will still be valid and robust as long as the under-reporting is uniformly distributed Second, the social distancing index may not precisely capture the level of social distancing in a society. The SDI metric is based on mobility information obtained through mobile devices that need users to enable their devices for data collection (e.g., GPS history). This type of data collection by nature will exclude people who do not have smart phones, or do not have their devices enabled for data collection. To the best of our knowledge, UMD social distancing index not only includes the most comprehensive and relevant dimensions especially those capturing spatial dependencies in mobility (e.g., percentage of out-of-county trips) but also covers all 3,142 counties across the U.S. The other available social distancing related datasets such as the Google community mobility data and Apple mobility data are either proprietary, or not available at all county levels (e.g., Google community mobility data covers only 2,795 counties in the U.S.

2.6 Figures and Tables

Figure 2.1: Spatial spread of cumulative infections per 100,000 people by urban and rural counties in the US

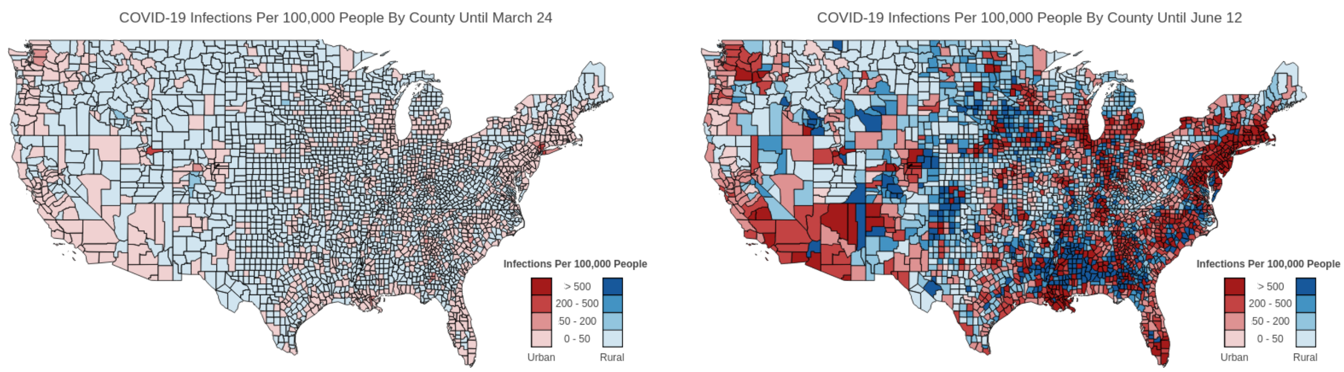
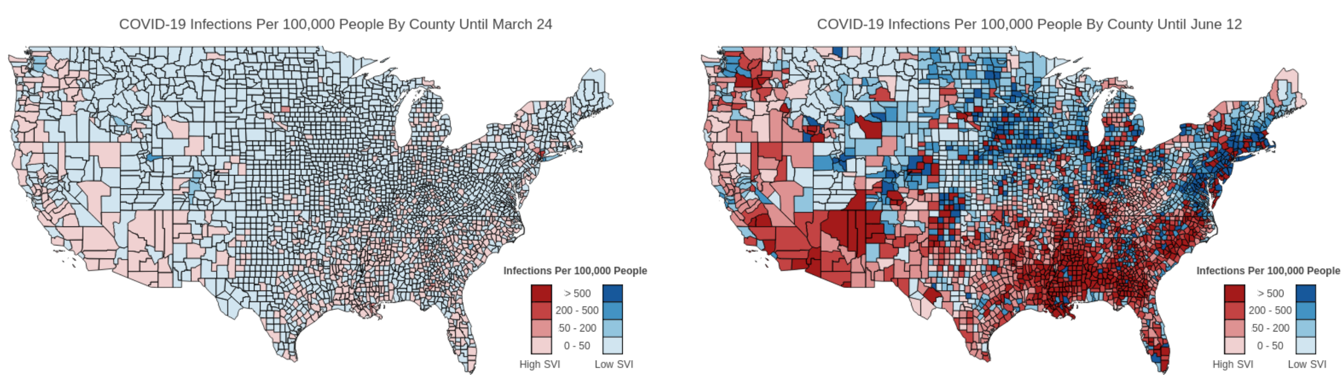


Figure 2.2: Spatial spread of cumulative infections per 100,000 people by counties with high and low social vulnerability index (SVI) in the US.



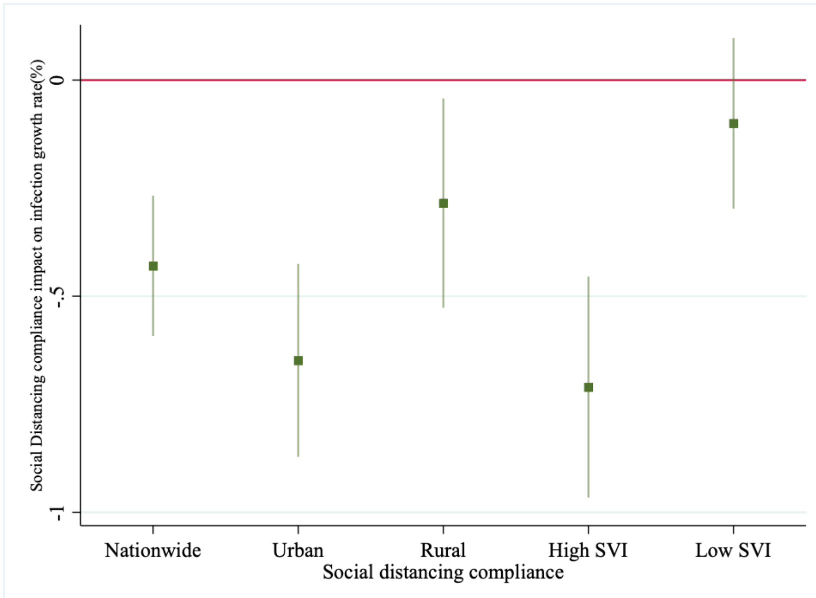
NOTES Spatial spread of cumulative infections per 100,000 people by counties with high and low social vulnerability index (SVI) in the US. High SVI counties are the ones in the top half of SVI distribution and Low SVI counties refer to those in the bottom 50% of the SVI distribution.

Table 2.1: Data description and Summary Statistics for pooled daily data (March 19 - June 12, 2020)

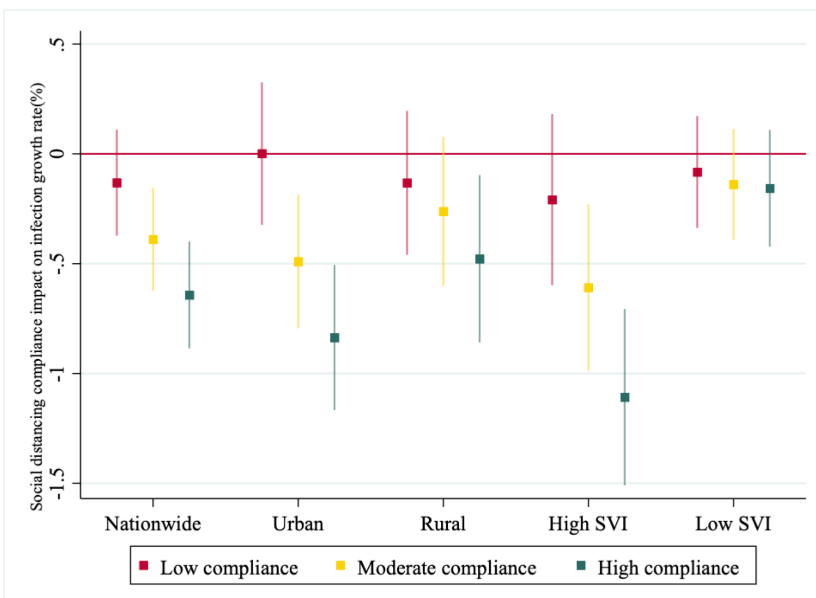
	Nationwide			Rural	Urban	Urban-Rural	Low SVI[a]	High SVI[b]	Low SVI-High SVI
	Mean	Minimum	Maximum	Mean	Mean	disparity	Mean	Mean	disparity
Infection growth rate(%)	5.33	0	2833	4.03	6.8	-2.77***	4.83	5.83	-1***
	(22.21)			(21.15)	(23.30)	(0.085)	(20.00)	(24.23)	(0.085)
Social distancing index[c]	33.52	0	100	30.31	37.15	-6.835***	34.76	32.28	2.481***
	(15.14)			(14.18)	(15.38)	(0.06)	(15.56)	(14.61)	(0.06)
Out of state trips(%)[d]	5.96	0	100	6.592	5.247	1.345***	6.298	5.624	0.675***
	(8.13)			(8.63)	(7.48)	(0.03)	(8.26)	(7.99)	(0.03)
COVID-19 Exposure per 1000 people[e]	5.33	0.01	44.32	4.634	6.123	-1.488***	5.731	4.932	0.799***
	(6.46)			(5.04)	(7.68)	(0.02)	(6.86)	(6.01)	(0.02)
Number of Tests done per 1000 people	23.05	0.02	181.1	22.55	23.6	-1.046***	23.25	22.83	0.418***
	(22.00)			(21.33)	(22.73)	(0.08)	(22.40)	(21.58)	(0.08)
People older than 60(%)	25.27	6	65.00	27.28	23.02	4.257***	26.31	24.23	2.078***
	(5.65)			(5.49)	(4.90)	(0.02)	(6.07)	(4.98)	(0.02)
African-Americans(%)	8.92	0	87.40	7.70	10.32	-2.611***	3.37	14.47	-11.10***
	(14.46)			(15.18)	(13.48)	(0.06)	(5.81)	(17.97)	(0.05)
Hispanic-Americans(%)	9.26	0	99.10	8.52	10.08	-1.604***	5.885	12.6	-6.714***
	(13.79)			(14.16)	(13.22)	(0.05)	(7.43)	(17.32)	(0.05)
Median income	51,580.44	20,188	136,268	46,686.20	57,161.30	-10,465.5***	58,650.9	44,520	14,130.9***
	(13700)			(10414.80)	(14800.90)	(48.83)	(13630.10)	(9479.70)	(45.17)
Male	50.09	41.39	79	50.54	49.57	0.971***	50.07	50.1	-0.0310***
	(2.38)			(2.78)	(1.68)	(0.01)	(1.77)	(2.86)	(0.01)
Number of points of interests[f]	131.56	8	699.00	137.70	124.70	13.05***	147.2	116	31.21***
	(42.33)			(48.73)	(32.05)	(0.16)	(46.23)	(30.94)	(0.15)
Observations		270212			270040			270126	

NOTES This Table reports summary statistics of selected variables of 3142 counties using COVID-19 Impact Analysis Platform database available from the University of Maryland for the time period between March 19 to June 12. Nationwide sample size (county-by-day observations; 3,142 counties for 86 days) is 270,212. Two counties are excluded in the rural-urban sample due to missing classifications. Therefore, 1,472 counties are classified as urban and 1,668 are as rural, which makes a total of 270,040 counties in the urban-rural sub-sample. Social vulnerability index (SVI) was also missing for a county which is excluded in the SVI sub-sample analysis. Therefore, 3,141 counties are classified in two sub-samples of High and Low SVI based on the median of SVI 2018 distribution. [a] Low SVI: counties in the bottom half of the SVI distribution[b] High SVI: counties in the top 50% of the SVI distribution. [c] The Social Distancing Index (SDI) is defined by Maryland Transportation Institute (MTI) as follows: SDI = 0.8*[% stay-at-home + 0.01*(100 - %stay-at-home)*(0.1*% reduction all trips + 0.2*% reduction work trips + 0.4*% reduction non-work trips + 0.3*% reduction travel distance)] + 0.2*% reduction out-of-county trips. It is an integer between 0 (no social distancing at all) and 100 (100% of the residents in the county follow social distancing).[d] Percentage of all trips that cross state calculated by MTI. [e] It is calculated as the number of residents already exposed to COVID-19 per 1000 people. [f] Number of points of interests for crowd gathering per 1000 people calculated by MTI. Asterisks show that the difference between the urban vs. rural and high vs. low SVI is statistically significant i.e., * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Figure 2.3: OLS estimated impact of social distancing on the daily growth rate of COVID-19 infections in the US with no spatial dependence assumed



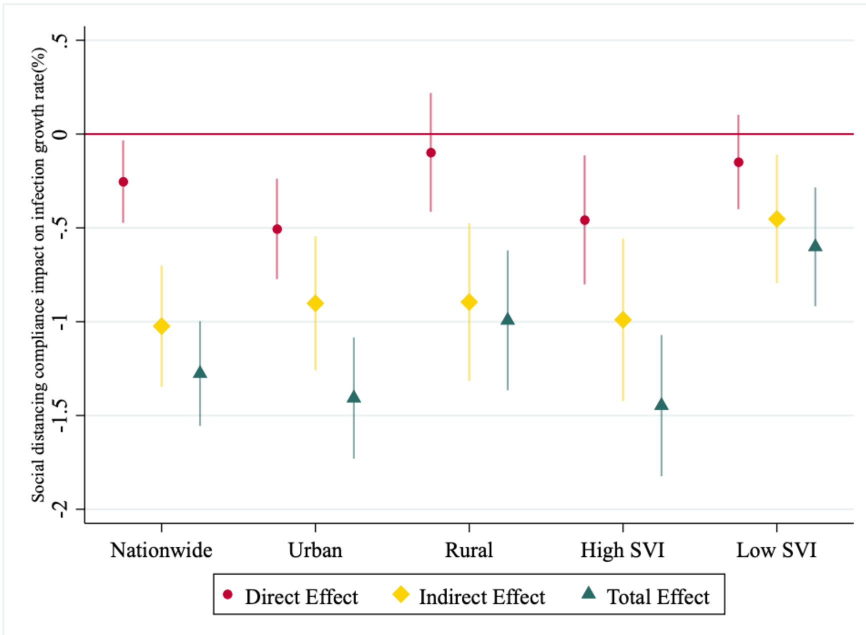
(a) OLS estimated impact of social distancing compliance on the daily growth rate of COVID-19 infections in the US.



(b) OLS estimated impact of different levels of social distancing compliance on the daily growth rate of COVID-19 infections as compared to non-compliant counties in the US.

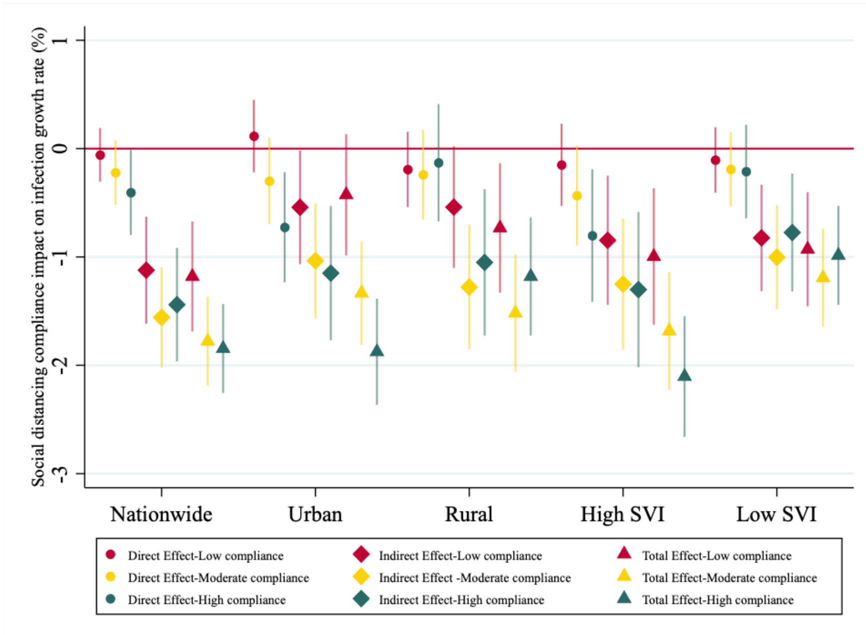
NOTES Figure(a) shows the impact with SDI measured in binary values and Figure(b) Counties in the bottom 25% of SDI distribution are called non-compliant counties which is the baseline in the estimations, those in the second quartile are low-compliance, third quartile are moderate-compliance, and the fourth quartile are the high-compliance counties. High SVI counties refer to those located in the top 50% of the SVI distribution and low SVI indicates counties in the bottom half of the SVI distribution. In all estimations, standard errors are heteroskedasticity-robust and clustered by counties. Bars present 95% confidence intervals.

Figure 2.4: Spatial estimation of social distancing compliance impact on daily growth rate of COVID-19 infections across US counties between March 19 to June 12, 2020



NOTES The figure shows direct, indirect and total effect of social distancing on daily growth rate of COVID-19 infections, for the binary SDI case. High SVI counties refer to those located in the top 50% of the SVI distribution and low SVI indicates counties in the bottom half of the SVI distribution. Direct effect refers to within county effect and Indirect effect indicates the effect of neighboring counties' on a county's infection growth rate. The total effect is the sum of direct and indirect effect. Standard errors are county cluster-robust. Bars present 95% confidence intervals.

Figure 2.5: Impact of various social distancing compliance levels on COVID-19 infection rate assuming spatial dependence between neighboring counties



NOTES Direct effect refers to within county effect and the Indirect effect refers to neighboring counties' effect on growth rate of infection. The total effect is the sum of direct and indirect effect. Counties in the bottom 25% of SDI distribution are called non-compliant counties which is the baseline in the estimations, those in the second quartile are low-compliance, third quartile are moderate-compliance, and the fourth quartile are the high-compliance counties. High SVI counties refer to the ones in the top 50% of the SVI distribution and low SVI in the bottom half of the SVI distribution. Standard errors are county cluster-robust. Bars present 95% confidence intervals.

Bibliography

Baker, S. R., Bloom, N., Davis, S. J., and Terry, S. J. (2020). COVID-19-induced economic uncertainty. Technical report, National Bureau of Economic Research.

Birge, J. R., Candogan, O., and Feng, Y. (2020). Controlling epidemic spread: Reducing economic losses with targeted closures. *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2020-57).

Bodas, M. and Peleg, K. (2020). Self-isolation compliance in the COVID-19 era influenced by compensation: Findings from a recent survey in Israel: Public attitudes toward the COVID-19 outbreak and self-isolation: a cross sectional study of the adult population of Israel. *Health Affairs*, 39(6):936–941.

Brockmann, D., David, V., and Gallardo, A. M. (2009). Human mobility and spatial disease dynamics. *Reviews of nonlinear dynamics and complexity*, 2:1–24.

Cajner, T., Crane, L. D., Decker, R., Hamins-Puertolas, A., and Kurz, C. J. (2020). Tracking labor market developments during the COVID-19 pandemic: A preliminary assessment.

Coibion, O., Gorodnichenko, Y., and Weber, M. (2020). Labor markets during the COVID-19 crisis: A preliminary view. Technical report, National Bureau of Economic Research.

Courtemanche, C., Garuccio, J., Le, A., Pinkston, J., and Yelowitz, A. (2020). Strong social distancing measures in the United States reduced the COVID-19 growth rate: Study evaluates the impact of social distancing measures on the growth rate of confirmed COVID-19 cases across the United States. *Health Affairs*, pages 10–1377.

Drago, R. and Miller, K. (2010). Sick at work: infected employees in the workplace during the h1n1 pandemic. *Institute for Women's Policy Research*, (B264).

Dunn, A., Hood, K., and Driessen, A. (2020). Measuring the effects of the COVID-19 pandemic on consumer spending using card transaction data. *US Bureau of Economic Analysis Working Paper WP2020-5*.

Guan, W.-j., Liang, W.-h., Zhao, Y., Liang, H.-r., Chen, Z.-s., Li, Y.-m., Liu, X.-q., Chen, R.-c., Tang, C.-l., Wang, T., et al. (2020). Comorbidity and its impact on 1590 patients with covid-19 in china: A nationwide analysis. *European Respiratory Journal*, 55(5).

Institute, M. T. (2020). University of maryland COVID-19 impact analysis platform.2020. *University of Maryland, https://data.covid.umd.edu (Accessed on June 16,2020)*.

Langford, C. (2002). The age pattern of mortality in the 1918-19 influenza pandemic: an attempted explanation based on data for England and wales. *Medical history*, 46(1):1–20.

Leclere, F. B., Jensen, L., and Biddlecom, A. E. (1994). Health care utilization, family context, and adaptation among immigrants to the United States. *Journal of health and social behavior*, pages 370–384.

LeSage, J. and Pace, R. (2009). Introduction to spatial econometrics. boca raton: Taylor and francishttp://dx. doi. org/10.1201/9781420064254.

Link, B. G. (2008). Epidemiological sociology and the social shaping of population health. *Journal of health and social behavior*, 49(4):367–384.

Lou, J., Shen, X., and Niemeier, D. (2020). Are stay-at-home orders more difficult to follow for low-income groups? Technical report, Working Paper.

Luo, S. and Tsang, K. P. (2020). The impact of stay-at-home orders on us output: a network perspective. *Available at SSRN 3571866*.

Marmot, M., Friel, S., Bell, R., Houweling, T. A., Taylor, S., on Social Determinants of Health, C., et al. (2008). Closing the gap in a generation: health equity through action on the social determinants of health. *The lancet*, 372(9650):1661–1669.

Nayak, A., Islam, S. J., Mehta, A., Ko, Y.-A., Patel, S. A., Goyal, A., Sullivan, S., Lewis, T. T., Vaccarino, V., Morris, A. A., et al. (2020). Impact of social vulnerability on COVID-19 incidence and outcomes in the United States. *medRxiv*.

Stock, J. H., Aspelund, K. M., Droste, M., and Walker, C. D. (2020). Estimates of the undetected rate among the sars-cov-2 infected using testing data from Iceland. *medRxiv*.

Wooldridge, J. M. (2016). *Introductory econometrics: A modern approach*. Nelson Education.

Wright, A. L., Sonin, K., Driscoll, J., and Wilson, J. (2020). Poverty and economic dislocation reduce compliance with COVID-19 shelter-in-place protocols. *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2020-40).

Zhang, L., Ghader, S., Pack, M. L., Xiong, C., Darzi, A., Yang, M., Sun, Q., Kabiri, A., and Hu, S. (2020). An interactive COVID-19 mobility impact and social distancing analysis platform. *medRxiv*.

Appendix B

(Second Chapter Appendix)

Table B.1: Ordinary Least Squares Estimation

	Dependent variable: Daily growth rate of confirmed COVID-19 cases (%)									
	Nationwide		Urban		Rural		High SVI[a]		Low SVI[b]	
14-day lag of SDI-compliance[c]	-0.43***		-0.65***		-0.28**		-0.71***		-0.1	
	(0.08)		(0.11)		(0.12)		(0.13)		(0.10)	
14-day lag of SDI low-compliance[d]		-0.13		0		-0.13		-0.21		-0.08
		(0.12)		(0.17)		(0.17)		(0.20)		(0.13)
14-day lag of SDI moderate-compliance[d]		-0.39***		-0.49***		-0.26		-0.61***		-0.14
		(0.12)		(0.15)		(0.17)		(0.19)		(0.13)
14-day lag of SDI high-compliance[d]		-0.64***		-0.84***		-0.48**		-1.11***		-0.16
		(0.12)		(0.17)		(0.19)		(0.20)		(0.14)
l14_out of state trips(%)[e]	0.01***	0.01***	0.04***	0.04***	0	0	0.03***	0.03***	0	0
	0.00	0.00	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
14-day lag of COVID-19 Exposure per 1000 people[f]	0	0	0.05***	0.06***	-0.03**	-0.03**	0	0	-0.02*	-0.02*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Number of Tests done per 1000 people	-0.09***	-0.09***	-0.12***	-0.12***	-0.07***	-0.07***	-0.11***	-0.11***	-0.07***	-0.07***
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
People older than 60(%)	-0.07***	-0.07***	-0.03***	-0.03***	-0.09***	-0.09***	-0.08***	-0.08***	-0.05***	-0.05***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Ln(median income)	0.13	0.19	-0.70***	-0.58***	0.88***	0.89***	0.58*	0.67**	0.86***	0.87***
	(0.17)	(0.17)	(0.20)	(0.20)	(0.33)	(0.33)	(0.31)	(0.32)	(0.29)	(0.29)
African-Americans(%)	0.03***	0.03***	0.01***	0.01***	0.04***	0.04***	0.03***	0.03***	0.01	0.01
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	(0.01)	(0.01)
Hispanic-Americans(%)	0	0	0.01**	0.01***	-0.01	-0.01	-0.01	0	-0.02***	-0.02***
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	(0.01)	(0.01)
Male	0.02	0.02	0.10***	0.10***	-0.02	-0.02	0.06*	0.06*	-0.12***	-0.12***
	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.02)	(0.02)
Number of points of interests[g]	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Constant	7.36***	6.83***	11.44***	10.44***	1.02	1	1.33	0.67	5.17	5.11
	(2.22)	(2.22)	(3.16)	(3.17)	(3.54)	(3.53)	(3.82)	(3.83)	(3.54)	(3.56)
Observations	226224		105984		120096		113112		113040	
R-squared	0.013	0.01	0.027	0.03	0.008	0.01	0.013	0.01	0.013	0.01
F-statistics	399.05	332.56	380.07	321.67	107.8	91.22	212.92	178.03	179.33	150.29

NOTES This Table reports OLS estimation results of social distancing compliance on growth rate of COVID-19 infections using COVID-19 Impact Analysis Platform data base provided by University of Maryland for the time period between March 19 to June 12. [a] High SVI: counties in the top 50% of the SVI distribution. [b] Low SVI: counties in the bottom half of the SVI distribution. [c] SDI stands for social distancing index. SDI-compliance is a dummy variable that takes value 1 for counties within the top 50% of social distancing distribution, and 0 otherwise (i.e. called non-compliant).[d] Counties in the bottom 25% of SDI distribution are called non-compliant counties which is the baseline in the estimations, those in the second quartile are low-compliance, third quartile are moderate-compliance, and the fourth quartile are the high-compliance counties. [e] Percentage of all trips that cross state calculated by MTI. [f] It is the number of residents already exposed to COVID-19 per 1,000 people. Robust county-clustered standard errors are presented in the parenthesis. [g] Number of points of interests for crowd gathering per 1,000 people calculated by MTI. Asterisks present statistical significance levels as follow, * p<0.10, ** p<0.05, *** p<0.01.

Table B.2: Spatial model estimation of the effect of social distancing on the daily growth rate of COVID-19 infections

	Dependent variable: Daily growth rate of confirmed COVID-19 cases				
	Nationwide	Urban	Rural	High SVI[a]	Low SVI[b]
Direct Effect[c]					
14-day lag of SDI-compliance[d]	-0.25**	-0.51***	-0.10	-0.46***	-0.15
	(0.0011)	(0.0014)	(0.0016)	(0.0018)	(0.0013)
14-day lag of out of state trips[e]	0.06***	0.16***	0.06**	0.09**	0.06***
	(0.0002)	(0.0004)	(0.0003)	(0.0004)	(0.0002)
14-day lag of COVID-19 Exposure per 1000 people[f]	0.14**	-0.01	-0.08	-0.36***	-0.04
	(0.0006)	(0.0005)	(0.0010)	(0.0008)	(0.0005)
14-day lag of Tests Done per 1000 people	0.02	-0.09***	-0.05**	-0.06***	-0.04***
	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0001)
Indirect Effect[g]					
14-day lag of SDI	-1.02***	-0.90***	-0.90***	-0.99***	-0.45***
	(0.0017)	(0.0018)	(0.0021)	(0.0022)	(0.0017)
14-day lag of out of state trips	0.45***	0.66***	0.14***	0.43***	0.25***
	(0.0005)	(0.0006)	(0.0005)	(0.0007)	(0.0004)
14-day lag of COVID-19 Exposure per 1000 people	-0.47***	-0.20***	-0.31***	0.01	-0.28***
	(0.0007)	(0.0005)	(0.0010)	(0.0008)	(0.0006)
14-day lag of Tests Done per 1000 people	-0.09***	-0.02	0.01	-0.03*	0.00
	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0001)
Total Effect[h]					
14-day lag of SDI	-1.28***	-1.41***	-0.99***	-1.45***	-0.60***
	(0.0014)	(0.0016)	(0.0019)	(0.0019)	(0.0016)
L14_out of state trips	0.52***	0.82***	0.20***	0.53***	0.31***
	(0.0005)	(0.0007)	(0.0005)	(0.0007)	(0.0004)
14-day lag of COVID-19 Exposure per 1000 people	-0.34***	-0.22***	-0.39***	-0.35***	-0.33***
	(0.0002)	(0.0002)	(0.0004)	(0.0003)	(0.0002)
14-day lag of Tests Done per 1000 people	-0.07***	-0.11***	-0.04***	-0.09***	-0.04***
	(0.0000)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Observations	226,224	105,984	120,096	113,112	113,040
County Fixed-effects	Yes	Yes	Yes	Yes	Yes
χ^2 statistics for spatial term	1860.51	1263.56	340.13	484.51	604.67
AIC	-145915.1	-118938.5	-42511.73	-43950.86	-111204.5

NOTES This Table reports spatial estimation results of social distancing compliance on growth rate of COVID-19 infections using COVID-19 Impact Analysis Platform database provided by the University of Maryland for the time period March 19 to June 12. [a] High SVI: counties in the top 50% of the SVI distribution. [b] Low SVI: counties in the bottom half of the SVI distribution.[c] Direct effect refers to within county's effect [d] SDI stands for social distancing index. SDI-compliance is a dummy variable that takes value 1 for counties within the top 50% of social distancing distribution, and 0 otherwise (i.e. called non-compliant).[e] Percentage of all trips that cross state as calculated by MTI. [f] It is calculated by MTI as the number of residents already exposed to COVID-19 per 1000 people. [g] Indirect effect shows neighboring counties social distancing effect on county's infection rate. [h] Total effect is the sum of the direct and indirect effects. Robust county-clustred standard errors are presented in the parenthesis. Asterisks present statistical significance levels as follow, * p<0.10, ** p<0.05, *** p<0.01.

Table B.3: Spatial estimation results of different social distancing thresholds on the daily growth rate of COVID-19 infections

	Dependent variable: Daily growth rate of confirmed COVID-19 cases (%)				
	Nationwide	Urban	Rural	High SVI[a]	Low SVI[b]
Direct Effect[c]					
14-day lag of SDI low-compliance[d]	-0.06	0.12	-0.19	-0.15	-0.10
	(0.0013)	(0.0017)	(0.0018)	(0.0019)	(0.0015)
14-day lag of SDI moderate-compliance[d]	-0.22	-0.30	-0.24	-0.43*	-0.19
	(0.0015)	(0.0020)	(0.0021)	(0.0024)	(0.0018)
14-day lag of SDI high-compliance[d]	-0.40**	-0.73***	-0.13	-0.80**	-0.21
	(0.0020)	(0.0026)	(0.0028)	(0.0031)	(0.0022)
14-day lag of out of state trips[e]	0.06***	0.15***	0.06**	0.09**	0.07***
	(0.0002)	(0.0004)	(0.0003)	(0.0004)	(0.0002)
14-day lag of COVID-19 Exposure per 1000 people[f]	0.13**	-0.01	-0.08	-0.36***	-0.04
	(0.0006)	(0.0005)	(0.0010)	(0.0008)	(0.0005)
2-day lag of Tests Done per 1000 people	0.02	-0.09***	-0.05**	-0.06***	-0.04***
	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0001)
Indirect Effect[g]					
14-day lag of SDI low-compliance	-1.12***	-0.54**	-0.54*	-0.85***	-0.82***
	(0.0025)	(0.0027)	(0.0029)	(0.0030)	(0.0025)
14-day lag of SDI moderate-compliance	-1.56***	-1.04***	-1.28***	-1.25***	-1.00***
	(0.0024)	(0.0027)	(0.0029)	(0.0031)	(0.0024)
14-day lag of SDI high-compliance	-1.44***	-1.15***	-1.05***	-1.30***	-0.77***
	(0.0027)	(0.0032)	(0.0035)	(0.0037)	(0.0028)
14-day lag of out of state trips	0.46***	0.65***	0.14***	0.43***	0.25***
	(0.0005)	(0.0006)	(0.0005)	(0.0007)	(0.0004)
14-day lag of COVID-19 Exposure per 1000 people	-0.46***	-0.21***	-0.30***	0.02	-0.28***
	(0.0007)	(0.0005)	(0.0010)	(0.0008)	(0.0006)
2-day lag of Tests Done per 1000 people	-0.09***	-0.02	0.01	-0.03*	0.00
	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0001)
Total Effect[h]					
14-day lag of SDI low-compliance	-1.18***	-0.43	-0.73**	-1.00***	-0.93***
	(0.0026)	(0.0029)	(0.0030)	(0.0032)	(0.0027)
14-day lag of SDI moderate-compliance	-1.78***	-1.33***	-1.52***	-1.68***	-1.19***
	(0.0021)	(0.0024)	(0.0027)	(0.0028)	(0.0023)
14-day lag of SDI high-compliance	-1.84***	-1.88***	-1.18***	-2.10***	-0.99***
	(0.0021)	(0.0025)	(0.0028)	(0.0028)	(0.0023)
14-day lag of out of state trips	0.52***	0.80***	0.21***	0.52***	0.32***
	(0.0005)	(0.0007)	(0.0005)	(0.0007)	(0.0004)
14-day lag of COVID-19 Exposure per 1000 people	-0.33***	-0.22***	-0.38***	-0.34***	-0.32***
	(0.0002)	(0.0002)	(0.0004)	(0.0003)	(0.0002)
2-day lag of Tests Done per 1000 people	-0.08***	-0.11***	-0.04***	-0.10***	-0.04***
	(0.0000)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Observations	226,224	105,984	120,096	113,112	113,040
County Fixed-effects	Yes	Yes	Yes	Yes	Yes
χ^2 statistics for spatial term	1863.45	1242.86	341.65	478.93	615.00
AIC	-145931.3	-118947.7	-42511.95	-43955.36	-111212.1

NOTES This Table reports spatial estimation results of different social distancing compliance levels on growth rate of COVID-19 infections using COVID-19 Impact Analysis Platform data base provided by University of Maryland for the time period between March 19 to June 12. [a] High SVI: counties in the top 50% of the SVI distribution. [b] Low SVI: counties in the bottom half of the SVI distribution.[c] Direct effect refers to within county effect. [d] SDI stands for social distancing index. Counties in the bottom 25% of SDI distribution are called non-compliant counties which is the baseline in the estimations, those in the second quartile are low-compliance, third quartile are moderate-compliance, and the fourth quartile are the high-compliance counties. [e] Percentage of all trips that cross state as calculated by MTI. [f] It is calculated by MTI as the number of residents already exposed to COVID-19 per 1000 people. [g] Indirect effect indicates the average impact of social distancing compliance among neighboring counties on a county's infection rate. [h] Total effect is the sum of the direct and indirect effects. Robust county-clustered standard errors are presented in the parenthesis. Asterisks present statistical significance levels as follow, * p<0.10, ** p<0.05, *** p<0.01.

Table B.4: Testing the non-linear combination of estimation coefficients in TableB.3

	Nationwide	Urban	Rural	High SVI	Low SVI
Direct Effect					
moderate-compliance / low-compliance	5.5251	-1.9740	1.1970	3.0822	1.9543
	(18.0659)	(3.8249)	(1.0469)	(3.6652)	(2.6946)
moderate-compliance / low-compliance	11.0077	-5.1377	0.6150	5.8676	2.2530
	(38.2794)	(8.0597)	(1.3337)	(7.6484)	(3.3541)
high-compliance / moderate-compliance	1.9923*	2.6027**	0.5138	1.9037**	1.1528
	(1.0882)	(1.3248)	(1.0200)	(0.7662)	(0.8981)
Indirect Effect					
moderate-compliance / low-compliance	1.3687***	1.7869***	2.3932**	1.4567***	1.2090***
	(0.2795)	(0.6698)	(1.1403)	(0.4627)	(0.3245)
high-compliance / low-compliance	1.2382***	1.8893***	1.9752**	1.4817***	0.9270***
	(0.2534)	(0.7175)	(0.9407)	(0.4866)	(0.3006)
high-compliance / moderate-compliance	0.9047***	1.0573***	0.8253***	1.0172***	0.7667***
	(0.1452)	(0.2096)	(0.2355)	(0.2479)	(0.2102)

NOTES This Table shows whether estimated coefficients (related to levels of compliance with social distancing) in TableB.3 are statistically different from each other. See notes for more details. Robust county-clustered standard errors are presented in the parenthesis. Asterisks present statistical significance levels as follow, * p<0.10, ** p<0.05, *** p<0.01.

Table B.5: Spatial model estimation of the effect of social distancing on the 3-day moving average growth rate of COVID-19 infections

	Nationwide	Urban	Rural	High SVI[a]	Low SVI [a]
Direct Effect[c]					
14-day lag of SDI-compliance [d]	-0.2096***	-0.2645***	-0.1550*	-0.3574***	-0.0752
	(0.0648)	(0.0782)	(0.0009)	(0.1012)	(0.0007)
14-day lag of out of state trips [e]	0.0622***	0.2024***	0.0463***	0.1121***	0.0538***
	(0.0126)	(0.0231)	(0.0160)	(0.0230)	(0.0134)
14-day lag of COVID-19 Exposure per 1000 people [f]	0.1308***	-0.0497*	-0.0426	-0.3400***	-0.0285
	(0.0364)	(0.0003)	(0.0006)	(0.0452)	(0.0003)
2-day lag of Tests Done per 1000 people	0.0161	-0.0814***	-0.0507***	-0.0548***	-0.0352***
	(0.0001)	(0.0073)	(0.0135)	(0.0103)	(0.0075)
Indirect Effect [g]					
14-day lag of SDI-compliance	-0.7742***	-0.8190***	-0.5933***	-0.7883***	-0.4140***
	(0.1005)	(0.1094)	(0.1273)	(0.1301)	(0.1062)
14-day lag of out of state trips	0.3834***	0.5588***	0.1259***	0.3552***	0.1888***
	(0.0287)	(0.0383)	(0.0292)	(0.0397)	(0.0247)
14-day lag of COVID-19 Exposure per 1000 people	-0.4555***	-0.1662***	-0.3297***	0.0133	-0.2888***
	(0.0399)	(0.0301)	(0.0611)	(0.0005)	(0.0336)
2-day lag of Tests Done per 1000 people	-0.0761***	-0.0153**	0.0193	-0.0244**	0.0036
	(0.0115)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Total Effect[h]					
14-day lag of SDI-compliance	-0.9838***	-1.0835***	-0.7483***	-1.1457***	-0.4892***
	(0.0900)	(0.1035)	(0.1156)	(0.1160)	(0.1022)
14-day lag of out of state trips	0.4456***	0.7612***	0.1722***	0.4673***	0.2426***
	(0.0293)	(0.0409)	(0.0305)	(0.0411)	(0.0267)
14-day lag of COVID-19 Exposure per 1000 people	-0.3246***	-0.2159***	-0.3723***	-0.3267***	-0.3173***
	(0.0138)	(0.0133)	(0.0239)	(0.0209)	(0.0146)
2-day lag of Tests Done per 1000 people	-0.0600***	-0.0968***	-0.0314***	-0.0792***	-0.0317***
	(0.0030)	(0.0033)	(0.0043)	(0.0041)	(0.0034)
Observations	226,224	105,984	120,096	113,112	113,040
County Fixed-Effects	YES	YES	YES	YES	YES
χ^2 statistics	4662.00	3271.80	857.46	1075.32	1854.68
AIC	-387375.6	-232931.7	-169819.9	-165431.7	-229297.9

NOTES This Table reports spatial estimation results of social distancing compliance on growth rate of COVID-19 infections using COVID-19 Impact Analysis Platform database provided by the University of Maryland for the time period March 19 to June 12. [a] High SVI: counties in the top 50% of the SVI distribution. [b] Low SVI: counties in the bottom half of the SVI distribution. [c] Direct effect refers to within county's effect [d] SDI stands for social distancing index. SDI-compliance is a dummy variable that takes value 1 for counties within the top 50% of social distancing distribution, and 0 otherwise (i.e. called non-compliant).[e] Percentage of all trips that cross state as calculated by MTI. [f] It is calculated by MTI as the number of residents already exposed to COVID-19 per 1000 people. [g] Indirect effect shows neighboring counties social distancing effect on county's infection rate. [h] Total effect is the sum of the direct and indirect effects. Robust county-clustered standard errors are presented in the parenthesis. Asterisks present statistical significance levels as follow, * p<0.10, ** p<0.05, *** p<0.01.

Table B.6: Spatial estimation results of different social distancing thresholds on 3-day moving average growth rate of COVID-19 infections

	Dependent variable: 3-day moving average growth rate of confirmed COVID-19 cases (%)				
	Nationwide	Urban	Rural	High SVI[a]	Low SVI[b]
Direct Effect [c]					
14-day lag of SDI low-compliance [d]	-0.0989	-0.0011	-0.1147	-0.1346	-0.1312
	(0.0007)	(0.0010)	(0.0010)	(0.0011)	(0.0009)
14-day lag of SDI moderate-compliance [d]	-0.2157**	-0.1707	-0.2109*	-0.3478**	-0.1378
	(0.0009)	(0.0012)	(0.0012)	(0.0014)	(0.0010)
14-day lag of SDI high-compliance [d]	-0.3897***	-0.5134***	-0.2169	-0.7898***	-0.1041
	(0.1153)	(0.1480)	(0.0016)	(0.1796)	(0.0013)
14-day lag of out of state trips [e]	0.0625***	0.1986***	0.0466***	0.1116***	0.0542***
	(0.0126)	(0.0231)	(0.0160)	(0.0231)	(0.0134)
14-day lag of COVID-19 Exposure per 1000 people [f]	0.1301***	-0.0496*	-0.0425	-0.3415***	-0.0283
	(0.0364)	(0.0003)	(0.0006)	(0.0452)	(0.0003)
2-day lag of Tests Done per 1000 people	0.0167	-0.0827***	-0.0509***	-0.0569***	-0.0351***
	(0.0001)	(0.0073)	(0.0135)	(0.0103)	(0.0075)
Indirect Effect[g]					
14-day lag of SDI low-compliance	-0.6484***	-0.4428***	-0.1727	-0.3323*	-0.5884***
	(0.1566)	(0.1636)	(0.0017)	(0.0018)	(0.1550)
14-day lag of SDI moderate-compliance	-0.9003***	-0.8830***	-0.6043***	-0.7095***	-0.7319***
	(0.1437)	(0.1616)	(0.1746)	(0.1818)	(0.1494)
14-day lag of SDI high-compliance	-1.0549***	-1.0193***	-0.7129***	-0.8156***	-0.7318***
	(0.1610)	(0.1874)	(0.2039)	(0.2142)	(0.1677)
14-day lag of out of state trips	0.3796***	0.5425***	0.1264***	0.3482***	0.1915***
	(0.0288)	(0.0385)	(0.0292)	(0.0398)	(0.0247)
14-day lag of COVID-19 Exposure per 1000 people	-0.4515***	-0.1663***	-0.3292***	0.0157	-0.2849***
	(0.0399)	(0.0302)	(0.0611)	(0.0005)	(0.0336)
2-day lag of Tests Done per 1000 people	-0.0794***	-0.0160**	0.0182	-0.0244**	0.0009
	(0.0115)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Total Effect[h]					
14-day lag of SDI low-compliance	-0.7473***	-0.4440**	-0.2874	-0.4670**	-0.7196***
	(0.1636)	(0.0018)	(0.0019)	(0.0019)	(0.1700)
14-day lag of SDI moderate-compliance	-1.1160***	-1.0537***	-0.8152***	-1.0573***	-0.8696***
	(0.1327)	(0.1528)	(0.1671)	(0.1675)	(0.1463)
14-day lag of SDI high-compliance	-1.4446***	-1.5327***	-0.9298***	-1.6054***	-0.8359***
	(0.1324)	(0.1570)	(0.1691)	(0.1718)	(0.1474)
14-day lag of out of state trips	0.4421***	0.7412***	0.1729***	0.4598***	0.2457***
	(0.0293)	(0.0411)	(0.0306)	(0.0412)	(0.0268)
14-day lag of COVID-19 Exposure per 1000 people f	-0.3214***	-0.2158***	-0.3717***	-0.3258***	-0.3132***
	(0.0139)	(0.0133)	(0.0239)	(0.0210)	(0.0147)
2-day lag of Tests Done per 1000 people	-0.0626***	-0.0987***	-0.0327***	-0.0813***	-0.0342***
	(0.0030)	(0.0034)	(0.0043)	(0.0042)	(0.0035)
Observations	226,224	105,984	120,096	113,112	113,040
County Fixed-Effects	YES	YES	YES	YES	YES
χ^2statistics	4635.61	3214.49	852.12	1048.33	1861.66
AIC	-387393.7	-232955.1	-169814.8	-165444.7	-229309.1

NOTES This Table reports spatial estimation results of different social distancing compliance levels on growth rate of COVID-19 infections using COVID-19 Impact Analysis Platform data base provided by University of Maryland for the time period between March 19 to June 12. [a] High SVI: counties in the top 50% of the SVI distribution. [b] Low SVI: counties in the bottom half of the SVI distribution.[c] Direct effect refers to within county effect. [d] SDI stands for social distancing index. Counties in the bottom 25% of SDI distribution are called non-compliant counties which is the baseline in the estimations, those in the second quartile are low-compliance, third quartile are moderate-compliance, and the fourth quartile are the high-compliance counties. [e] Percentage of all trips that cross state as calculated by MTI. [f] It is calculated by MTI as the number of residents already exposed to COVID-19 per 1000 people. [g] Indirect effect indicates the average impact of social distancing compliance among neighboring counties on a county's infection rate. [h] Total effect is the sum of the direct and indirect effects. Robust county-clustered standard errors are presented in the parenthesis. Asterisks present statistical significance levels as follow, * p<0.10, ** p<0.05, *** p<0.01.

Chapter 3

Black–White Disparities in COVID-19 Outcomes: The Relative Contribution of disparity in healthcare access

3.1 Introduction

COVID-19 has disproportionately affected disadvantaged groups and minorities, especially black communities in the US (Yancy, 2020; Millett et al., 2020). Louis-Jean et al. (2020) reports that COVID-19 mortality for black is almost four times the national rate. Another study shows that, in 14 states in the US, black individuals constitutes one-third of the hospitalization, while they only comprise 18% of the population in those states (Garg, 2020). Figure 3.1 illustrates the percentage of COVID-19 infections and mortality in black communities along with their share of the population for selective states during April 23, 2020 and July 22, 2020. As shown in the Figure, the percentage of COVID-19 infections and mortality in black communities is noticeably larger than their population portion, if not more than double in some states.

These disparities in COVID-19 outcomes refocus attention on the long-lasting racial disparity in health outcomes that are deep-rooted in systemic racism (Kullar et al., 2020). For a long time, existence of systemic racism triggers racial disparity in education, employment, income, housing, food insecurity, healthcare access, and other factors, which exposes the disadvantaged groups to higher risk of COVID-19 infections, hospitalization, and death (Williams and Cooper, 2019). Besides, exposure to socioeconomic and environmental inequalities increases the occurrence of health conditions such as diabetes, obesity, hypertension, etc. that are reported to further intensify the risk of sever COVID-19 infections (Garg, 2020).[1]

Finding the roots of disparity in COVID-19 outcomes helps to design policies to reduce racial disparity in COVID-19 pandemic or possible future pandemics. Lack of knowledge in the source of such disparities can lead to explanations grounded in biological differences or behavioral patterns, which gives misleading policy insights for reducing the health disparities (Chowkwanyun et al., 2020). Subsequently, this line of research has been recently known as immediate priority (Chowkwanyun et al., 2020; Kullar et al., 2020).

Among the aforementioned factors, healthcare access has been known for a long time as the focal point of eliminating disparity in health outcomes for socially disadvantaged groups (Andrulis, 1998). According to CDC, one of the most important determinants of healthcare access is the availability of health insurance.[2] Uninsured individuals are less likely to use healthcare services and have access to healthcare providers (Kilbourne, 2005). In particular, socially disadvantaged groups and minorities are found to be less likely to have health insurance (Berchick et al., 2018).

Despite the importance of inequality in healthcare access on health disparity, a limited number of studies have examined the impact of healthcare access on disparity in COVID-19 outbreak (i.e., infections and mortality). Chunara et al. (2020) finds that the probability of having healthcare access through telemedicine for black patients is only 60% of the prob-

[1] https:// www.cdc.gov/coronavirus/2019- ncov/need-extra-precautions/ people-at-higher-risk.html
[2] https://www.cdc.gov/nchs/data/factsheets/factsheet$_h$*iac.pdf*

ability for white individuals during the pandemic, although COVID-19 infection was more likely for these black telemedicine patients than white telemedicine individuals.

A big challenge to assess whether disparity in healthcare access contributes to racial disparity in COVID-19 outcomes is data limitations. Although most states report the COVID-19 infections and mortality by race, the detailed data is not at the individual level due to confidentiality rules. In this study, we examine whether disparity in access to healthcare widens the racial disparity in COVID-19 infections and mortality among black and white communities by linking multiple data sources at the state and individual-levels. In particular, we use "The COVID Tracking Project at The Atlantic" data to obtain state-level COVID-19 infections and mortality by race. Using Household Pulse Survey, a publicly available dataset by U.S. Census Bureau, we obtain the healthcare access and socioeconomic status of individuals nested in the states during the COVID-19 pandemic.

To examine how healthcare access impacts disparity in COVID-19 outcomes, we investigate (1) disparity in access to healthcare during the pandemic, (2) the disparity in COVID-19 outcomes for black and white sub-samples, and (3) how the disparity in healthcare access influences the disparity in COVID-19 outcomes, by employing decomposition methods, multi-level analysis as well as aggregate methods.

Our findings confirm that COVID-19 pandemic has significantly hit the black sub-sample more severely than whites. Black population comprise higher portion of COVID-19 infections and mortality than their share in the population. Our findings show that black individuals are not only less likely to have health insurance as compared to white individuals but also more likely to avoid medical care. Although a considerable portion of these differences between healthcare access of the two groups can be explained by their differences in observables (e.g., demographics, socioeconomic and health status), a significant portion cannot be explained by such observables.

The two-step analysis indicates that disparity in having health insurance, net of individual and state-level covariates, significantly decreases the disparity in COVID-19 mortality in black sub-sample while it has no effect on disparity in white sub-sample. We also find that having health insurance significantly reduces COVID-19 infections in white communities. Results of this paper highlight the importance of healthcare access such as health insurance in narrowing down the disparity in COVID-19 outcomes or other future pandemics.

3.2 Background

For many years, structural racism caused disparities in various health outcomes. Empirical studies indicate that black communities experience around 100,000 premature death annually due to racial disparity in health (Levine et al., 2016). Structural racism widens racial disparity in health outcomes through multiple interrelated channels: residential segregation, socioeconomic status, mass incarceration, and medical care (Bailey et al., 2020).

Residential segregation is the most deterministic channel through which structural racism influences health outcomes (Acevedo-Garcia et al., 2003). Black communities mostly reside in areas with dense population of minorities with low income and low socioeconomic status (Bailey et al., 2020). Residential segregation forms spatial distribution of resources such as quality of housing, quality of education, population density, employment opportunities, and access to health care (Bailey et al., 2020; Williams, 2001). These environmental factors shape the socioeconomic status of residents and thus increase susceptibility to infectious disease and COVID-19 (Gravlee, 2020). Besides, the residential environment inadequacy increases occurrence of chronic conditions- known as weathering hypothesis-, including hypertension, obesity, cardiovascular disease, heart disease that subsequently makes minorities more vulnerable to the complications of infectious disease such as COVID-19 (Bell et al., 2018; Maxwell, 2020). Wiemers et al. (2020) indicate that age, race, education, and income are the predictors of specific chronic medical conditions that raises the risk of COVID-19 infections and mortality.

In addition, employment condition and occupational inflexibility prevents minorities from complying adequately with social distancing and concludes in higher risk of infections. Studies found that black individuals are disproportionately work in the top nine essential industries which expose them to higher probability of getting infected by COVID-19 virus (Rogers et al., 2020). Among black individuals who were in risk of sever COVID-19 infections, 56.5 percent lived in household consists of at least one worker who could not work from home (Selden and Berdahl, 2020; Gould and Shierholz, 2020).

Literature has well-documented the racial inequality of US legal system, with black individuals being more incarcerated as compared to their white counterparts (Kutateladze et al., 2014; Knox et al., 2020). Disproportionate incarceration found to impact health of black individuals in different ways. Black individuals are in higher risk of mortality after release (Binswanger et al., 2007). Moreover, prisons are found to be major transmission site for infectious disease and COVID-19 in particular.[3]

Racial disparity in medical care access and usage is well-documented (Manuel, 2018). Minorities are having less access to healthcare as compared to whites. For instance, black individuals are less likely to have health insurance, which limits their access to medical care services (Ashton et al., 2003). Even with control for socioeconomic condition, demographics, and the level of medical care need, black individuals have been found to use healthcare services less than their white counterparts (Sealy-Jefferson et al., 2015). Racial discrimination also exists in healthcare providers, and thus the quality of treatment (Maina et al., 2018; Hall et al., 2015). In particular, Hall et al. (2015) list a vast range of implicit discrimina-

[3]https://covidprisonproject.com

tion in attitude of health professionals toward communities of color as compared to white counterparts; from longer waiting time to offering fewer treatment options.

3.3 Methodology

We investigate the impact of disparities in healthcare access on disparity in COVID-19 pandemic in three steps, (1) we assess whether there is a disparity in healthcare access during pandemic between black and white using decomposition methods, (2) we show there is disparity in COVID-19 outcomes (i.e., infections and mortality), (3) we examine how healthcare disparity impacts disparity in COVID-19 outcomes using aggregate method and a two-step approach that is an extension of Multilevel Analysis.

3.3.1 Racial disparity in healthcare access

We conduct a Blinder-Oaxaca decomposition analysis (BO) (Oaxaca, 1973) which was previously used to explain race and gender gaps in labor market, and recently implemented in health disparity research (Sen, 2014). The decomposition methods divide the differences in the means of an outcome variable between two groups (here differences in healthcare access between black and white individuals) into a portion that can be explained by group differences in observable explanatory variables and a portion that cannot be explained by the observables. That is, BO decomposition partitions the group differences to an "explained" part and an "unexplained" part. The "explained" part of this gap is the difference in the outcome induced by group differences in the mean levels of observables. In this study, the "explained" part represents the amount by which the black–white difference in healthcare access would be reduced if, ceteris paribus, black individuals experienced the same mean levels of observables: gender, education, marital status, health status, income, and employment status (i.e., lost job or not) exposures as whites. In contrast, the "unexplained" part is often interpreted as disparity and discrimination in the outcome between the two groups, which may reflect the group differences in unobservables as well.

The BO decomposition method estimates two separate regression models for black and white subgroups.[4] Conceptually, these regressions are the same as simple regression model with interaction terms between the race variable and each explanatory variable. The differences in the coefficients is known as disparity in the dependent variable between the subgroups. To account for the statistical dependence of health access in a household (i.e., clustering), we conducted the BO decomposition with household level clustering in order to capture the nesting feature of the data; individuals nested in household.

3.3.2 Racial disparity in COVID-19 pandemic

In the literature of health disparities, there is no unique definition for health disparity (Williams et al., 2006; Harper and Lynch, 2005). In this study, we define disparity in COVID-19 infections (mortality) for a racial group as difference between group's portion of COVID-19 infections(mortality) and portion of population consisting of that group (Moore et al., 2020).[5] Figure 3.3 presents the measurements of disparity for black and white by regions. As is depicted in Figure 3.3, black community experiences higher disparity in both COVID-19 infections and mortality in all regions with the largest gaps in Midwest and South regions. Although recent studies and reports show that COVID-19 impacts minorities and disadvantaged communities disproportionately (Louis-Jean et al., 2020; Garg, 2020; McLaren, 2020), we provide further evidence by applying statistical t-test to investigate if

[4] The details of the BO decomposition method are presented in the appendix.
[5] Moore et al. (2020) state that disparity exist if this difference is $\geq 5\%$.

the average disparity is significantly greater than 5% (Moore et al., 2020). More details are presented in the results section.

3.3.3 Contribution of disparity in healthcare access to disparity in COVID-19 outcomes

To examine how disparity in healthcare access affects the COVID-19 prevalence in black and white communities, we need to utilize individual-level (micro) characteristics to explain state-level outcome (macro). Since individuals are nested within the states, multilevel analysis (DiPrete and Forristal, 1994) is an ideal method of estimation. However, the structure of standard multilevel analysis is for macro-micro analysis, i.e., to explain micro-level outcomes through macro-level explanatory variables.

To address our micro-macro research question, we employ two different methods for comparison as well as for the purpose of robustness of our results. The first method involves aggregating the individual-level predictor to the state-level, i.e., using the group mean or other measures of central location (Becker et al., 2018; Lim et al., 2005). Eventually, a state-level analysis is performed in which the state-level outcome is regressed on the aggregated individual-level predictor. The second method is a two-step approach introduced by Griffin (1997) based on standard multilevel analysis in a hierarchical structure datasets (i.e., individuals are nested within states).

Aggregate method

Instead of using group means as in standard aggregate methods, we define the aggregate measurement of disparity in healthcare access similar to the definition of disparity in COVID-

19 outcomes.[6] In particular, healthcare disparity in black community is defined as the difference between portion of black individuals without health insurance and portion of black population in each week by states. Subsequently, we estimate the impact of disparity in healthcare access on disparity in COVID-19 infections and mortality using state Fixed-effect method as follows:

$$Disparity_{st} = \alpha_0 + \beta H_{st} + \alpha_1 \mathbf{X_{st}} + \mu_s + \epsilon_{st} \tag{3.1}$$

where $Disparity_{st}$ refers to black (white) disparity in COVID-19 outcome of state s at week t, H_{st} indicates the black (white) disparity in healthcare access in state s at week t (e.g., disparity in health insurance and medical care avoidance). $\mathbf{X_{st}}$ indicates a set of time-variant state-level covariates such as number of COVID-19 tests done per 1000 people, social distancing index in the state, unemployment rate and hospital bed utilization. μ_s denotes time-invariant state-specific characteristics such as geographic condition and population density.[7] ϵ_{st} indicates an i.i.d. error term.

We consider two different measures for the COVID-19 outcome variable ($Disparity_{st}$) and two values for disparities in healthcare access (H_{st}) for each racial sub-group: (1) black (white) disparity in COVID-19 infections as outcome and black (white) disparity in health insurance as the regressor, (2) black (white) disparity in COVID-19 infections as outcome and black (white) disparity in medical care avoidance as the regressor, (3) black (white) disparity in COVID-19 mortality as outcome and black (white) disparity in health insurance as the regressor, (4) black (white) disparity in COVID-19 mortality as outcome and black (white) disparity in medical care avoidance as the regressor.

[6] Croon and van Veldhoven (2007) show that arithmetic mean procedure gives valid estimates if the micro-level variance of the aggregated variable is zero

[7] Since the period of study is weekly, these variables are fixed along time for each state.

The two-step approach (micro-macro Analysis)

In hierarchical settings where individuals/households are nested within states, standard multilevel analysis is used to model the impacts of state-level predictors on individual-level outcomes. However, in this study, we seek to explain state-level outcome through individual-level predictors (i.e., micro-macro analysis). For this purpose, we implement the two-step approach developed by Griffin (1997). The two-step approach is an extension of standard multilevel analysis that enables us to examine the state-level outcomes by using individual-level predictors. In the first step, we regress the individual-level outcome (e.g, access to healthcare) on both individual and state-level predictors as follows:

$$H_{sht} = \beta_1 + \beta_\mathbf{2}\mathbf{X_{sh}} + \beta_\mathbf{3}\mathbf{Z_{sht}} + \beta_\mathbf{4}\mathbf{S}_s + \beta_\mathbf{5}\mathbf{K}_{st} + u_{0st} + \epsilon_{sht} \tag{3.2}$$

Where H_{sht} indicates healthcare access for individual in household h in the state s at time t (i.e., health insurance and medical care avoidance), $\mathbf{X_{sh}}$ is a vector of time-invariant individual-level variables such as gender, marital status, education, number of household members, and $\mathbf{Z_{sht}}$ indicates time-variant individual-level variables such as income and employment status, etc. $\mathbf{S_s}$ indicates the time-invariant state-level variables such as percentage of people older than 60 years old and median income in state, $\mathbf{K_{st}}$ indicates the time-variant state-level variables such as unemployment rate and number of COVID-19 tests done per 1000 people. ϵ_{sht} indicates an i.i.d. error term. We are interested in u_{0st}, the state-time level residual of the equation, which is interpreted as the aggregate effect of individual variable H_{sht}, net of both state and individual covariates. We obtain u_{0st} by including a random intercept for each combination of time and state in the estimation of equation (3.2) (Rabe-Hesketh and Skrondal, 2008).[8] In the second step, the state-level outcome (e.g., black disparity in the

[8] u_{0st} presents how much the healthcare access for state S at time t is deviated from the average healthcare

COVID-19 infections at the state-level) is regressed on the estimated residuals of individual's model, $\hat{u}_{0st}$, as follows:

$$Disparity_{st} = \gamma_0 + \gamma_1 \hat{u}_{0st} + \mu_s + \nu_{st} \tag{3.3}$$

where $Disparity_{st}$ is defined in previous section and reflects the disparity in COVID-19 infections (mortality), μ_s refers to time-invariant state covariates, and ν_{st} presents the error term. Griffin (1997) argues that $\hat{u}_{0st}$ provides a better estimation than the group mean aggregate. It is a model-based estimation of the state-level variance that is net of the individual-level discrepancies. Moreover, $\hat{u}_{0st}$ is adjusted for other explanatory variables at the both levels. This estimation may save degrees of freedom and avoid collinearity when we use $\hat{u}_{0st}$ as a predictor in a subsequent state-level regression. Furthermore, in multilevel analysis, the dependencies among individuals within a state are taken into account. For example, two residents in the same state might be affected by similar state policies as compared to two residents in different states.

The two-step approach is implemented separately on the black and white sub-samples. We consider four different measures for $Disparity_{st}$ in estimation of equation (3.3): (1) black disparity in COVID-19 infections, (2) black disparity in COVID-19 mortality, (3) white disparity in COVID-19 infections, (4) white disparity in COVID-19 mortality.

3.4 Data

To conduct the analysis, we link state and individual-level data from multiple resources. We obtain individual-level (micro) data from the Household Pulse Survey (HPS) available by

access.

U.S. Census Bureau. The HPS is conducted in partnership with other federal agencies to collect data on the social and economic impacts of COVID-19 in the U.S.[9] The HPS data is a weekly longitudinal data set of American households where phase 1 started April 23, 2020 and ended on July 21,2020 (Total of $986,153$ individuals). Accordingly, phase 1 of HPS includes 12 waves (weeks) of data. The first wave covers the period of April 23 until May 5 which is two weeks long, while the rest of the waves are for a week long. Figure 3.2 displays the timeline for each wave (week) of the HPS data. In the HPS, each household is followed up only three weeks and it is possible that not the same individual within the household was interviewed each time. HPS includes information about the respondent's demographics, employment, housing, food security, physical and mental health, access to healthcare, and educational disruption.

3.4.1 Healthcare access

We construct two measures for healthcare access using information available in HPS dataset (1) health insurance status, (2) medical care avoidance (delayed medical care; needed non-coronavirus medical care but did not get it) due to the coronavirus pandemic in the past 4 weeks (National Academies of Sciences, 2018; Murata and Kondo, 2020). In particular, respondents were asked "Are you currently covered by any types of health insurance or health coverage plans?" The possible response could be yes (1) or no (0).[10] Therefore, health insurance is a dummy variable that takes the value of 1 if an individual has health insurance coverage, 0 otherwise.

In regard to medical care avoidance, the respondents were asked: "At any time in the last 4 weeks, did you need medical care for something other than coronavirus, but DID NOT GET

[9] https://www.census.gov/householdpulsedata
[10] A total of 93,472 out of 986,153 observation are missing.

IT because of the coronavirus pandemic?" and "At any time in the last 4 weeks, did you DELAY getting medical care because of the coronavirus pandemic?" The respondents could answer yes (1), no (0), or choose not to answer the questions. We define a dummy variable for medical care avoidance as a combination of these two variables which takes value of 1 if the answer to either of these questions is yes and 0 otherwise.[11]

3.4.2 COVID-19 outcomes

To obtain COVID-19 infections and mortality by racial groups, we use data from the "The COVID Tracking Project at The Atlantic" that is publicly available.[12] This database provides state-level information on COVID-19 related infections and mortality by race and ethnicity every Sundays and Wednesdays starting April 12, 2020 across the U.S. plus Washington DC.

3.4.3 Other variables

In our models, we also include the following individual-level confounders using HPS data: age, gender (female = 1), marital status(coded as 5 categories), education (using 7 nominal categories), household size (total number of individuals in the household), income (coded as 8 categories), health status ("Would you say your health in general is excellent(1), very good(2), good(3), fair(4), poor(5)"), employment status ("Have you, or has anyone in your household experienced a loss of employment since March 13 = 1").

We also use data from the "COVID-19 Impact Analysis Platform" made publicly available by the University of Maryland (UMD) (Zhang et al., 2020, MTI). This database provides

[11] A total of 95,776 out of 986,153 observations are missing.

[12] https://covidtracking.com/about

information on COVID-19 related infections, deaths, social and economic indicators, social mobility, as well as COVID-19 testing information at state-level for the entire US on a daily basis. We use the UMD data to control for time-variant state-specifics during the pandemic such as unemployment rate, and number of COVID-19 tests done per 1000 people, social distancing index[13], percentage of hospital bed utilization. Further state-level information such as state's population race profile (i.e., black and white proportion) are obtained from US Census Bureau.

Therefore, we control in our models the following state-level variables: population density, population older than 60 years old, unemployment rate, number of COVID-19 tests done per 1000 people, social distancing index (with 2 weeks lag), percentage of hospital bed utilization.[14] Since expecting contemporaneous impact of healthcare access on disparities in COVID-19 outcomes may be unrealistic, we enter the corresponding variables with lags in our models. Moreover, due to the potential lags between COVID-19 infections and mortality (e.g., Testa et al. (2020) reports 2 to 6 weeks lags between COVID-19 infections and deaths), we enter healthcare access with one week lag in models with disparity in COVID-19 infections, and two weeks lag for models with disparity in COVID-19 mortality as the outcome. Table 3.1 displays a summary of variables used in the two-step approach.

3.4.4 Linking data sets

Figure 3.2 represents the overlap between time frame of HPS and racial tracker dataset. We accumulate the available racial tracker data within the weekly time frame of HPS. For instance, we add up the COVID-19 infections reported on April 26, April 29, May 3, and

[13]Social Distancing Index = 0.8*[% stay-at-home + 0.01*(100 - %stay-at-home)*(0.1*% reduction all trips + 0.2*% reduction work trips + 0.4*% reduction non-work trips + 0.3*% reduction travel distance)] + 0.2*% reduction out-of-county trips.

[14]we include social distancing with 2 weeks lag due to 14-day incubation period of COVID-19: https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html;

May 6 together and consider the summation as COVID-19 infections corresponding to the first week of HPS data.[15] We conduct a similar procedure for COVID-19 mortality.

To match the daily variables in UMD with HPS and racial tracker, we take the average of corresponding state-level variables (as mentioned earlier) in UMD within the time frame of HPS data.

Accordingly, we obtain the number of COVID-19 infections and mortality by racial groups correspond to the HPS weekly data. In this study, we focus on black versus white and limit the sample to sub-sample of black and whites.

Table 3.2 presents the summary statistics of the selected variables in HPS data by race. As Table 3.2 shows, black population in HPS dataset are significantly younger, more likely female, less educated, and have larger households on average. Also, on average, black individuals appear to have lower income and worse health status as compared to white individuals. As is shown in this table, blacks are more likely to lose their jobs than whites, on average. And on average, black individuals are not only less likely to have health insurance but also more likely to avoid medical care.

Table 3.3 reports average impact of COVID-19 infections and mortality on black and white as well as average percentage of population. As we see, black communities comprise on average 10.9% of the population, while they make up almost 28% of the COVID-19 infections and 25.1% of COVID-19 mortality. In contrast, whites compose 68% of population, 73.3% of the COVID-19 infections, and 74.9% of COVID-19 mortality on average.

[15] For example, number of infections on April 29 includes daily infections of April 27, 28, and 29.

3.5 Empirical Results

3.5.1 Racial Disparity in healthcare access-Decomposition Method

Table 3.4 shows the BO decomposition that divides the differential probability of access to healthcare between two race subgroups into a part that is "explained" by group differences in observables, such as gender, education, marital status, health status, income, and employment status, and a part that cannot be explained by the group differences in observables. The unexplained part is often interpreted as disparity, although it also involves the effects of racial group differences in unobservables.

According to this table, the mean of probability of having health insurance is 95% for whites and 90% for blacks, yielding a statistically significant gap of 5%. The decomposition part reflects that black individuals would have 3.7% higher probability of having health insurance if they had the same observable characteristics as whites. The largest portion of this difference is explained by income and probability of losing job during the pandemic. In particular, the mean increase in probability of having health insurance for black people would increase by 2.1% if they had the same employment status and income as whites. Also, it shows a black person with the same demographics (i.e., gender, age, education, marital status) as a white person would have 1.5% higher probability of having health insurance. The bottom part of Table 3.4 shows, on average, 22%$(\frac{0.011}{0.048})$ of the gap cannot be explained by observables, which interprets as disparity.

As is shown in Table 3.4, average probability of medical care avoidance is 29.6% for white individual, whereas it is 31% for black individuals. Black people experience 1.4%, on average, higher probability of medical care avoidance as compared to white people. This gap is mostly explained by differences in the observables. Black individuals would have 3.5%

lower probability of medical care avoidance if they had the same characteristics in terms of demographics, income, employment status, and health status as whites. Although most of the gap is explained by observable, there is a statistically significant 2.1% racial disparity in medical care avoidance that cannot be explained by the observable characteristics.

Higher medical care avoidance in black communities might be explained by discrimination in healthcare services. Recent studies show significant racial bias in healthcare providers and its implicit potential impact on healthcare demands (Hall et al., 2015; Johnson, 2020). In particular, Johnson (2020) finds that black individuals who had bad experience with healthcare providers and professionals are less likely to seek treatment.[16]

3.5.2 Racial Disparity in COVID-19 Infections and Mortality

Fig.(3.3) presents the black and white disparity in COVID-19 infections and mortality by regions in the U.S., respectively. The disparity gap between black and white is larger in Midwest and South regions than West and Northeast parts. Noticeable, the black disparity in both COVID-19 infections and mortality is positive, while the white disparity is negative. That is, portion of black infections (mortality) is larger than the portion of black population in each region.

Table 3.5 provides the summary statistics of disparity measurements for black and white over the state during the time period of study. The statistics show the black disparity varies between 0% and 40%. In particular, the portion of black COVID-19 infections goes up to 40% higher than their share of population, while the difference between portion of white infections and their share of population is at most 37%. On average, disparity in COVID-19 infections for blacks is more than double the disparity in COVID-19 infections for whites.

[16] Investigation of such mechanisms is out of the scope of this study.

Moore et al. (2020) defines disparity as difference of 5% between the minority portion of COVID-19 infections and their share of population. The one-sample t-test results of the comparison of mean disparity measurement and 5% are provided in the last two columns of Table 3.5. Findings confirm the significant disparity in COVID-19 infections and mortality for black sub-sample according to the definition proposed in Moore et al. (2020), indicating COVID-19 hit the black communities more severely than the whites.

3.5.3 Contribution of healthcare access to racial disparity in COVID-19 outcomes

Results based on aggregate method

Table 3.6 provides estimation results of healthcare access disparity on disparity in COVID-19 infections for black and white people using the aggregate model. The left panel of Table 3.6 shows that higher disparity in healthcare access in either form of health insurance or medical care avoidance has no significant impact on COVID-19 infections in black communities.[17] The right panel of 3.6 shows similar results for the white communities. Therefore, aggregate method estimations assert that there is no association between healthcare access disparity and disparity in COVID-19 infections.

Table 3.7 reports the estimation results of healthcare disparity on disparity in COVID-19 mortality for blacks and whites. It shows that disparity in health insurance significantly increases black disparity in COVID-19 mortality. In particular, 1% increase in health insurance disparity (i.e., higher portion of black individuals with no health insurance) is associated

[17] Table C.1 in the appendix reports the detailed estimation results corresponding to this table. The other state-level variables show the expected association with COVID-19 disparity in infections. More COVID-19 testing is associated with lower disparity, higher unemployment increases the disparity, more social distancing is associated with lower disparity, and hospital bed utilization plays no role in the disparity in COVID-19 infections among black community.

with 2.7% higher disparity in COVID-19 mortality. However, medical care avoidance has no impact on disparity in COVID-19 mortality.[18] For the white communities, health insurance disparity only marginally increases mortality disparity induced by COVID-19, while disparity in medical care avoidance plays no role.

Results based on two-step method

Table 3.8 provides the results from the two-step approach with disparity in COVID-19 infections as the outcome of interest.[19] The fist part shows that there is no significant relationship impact of health insurance (net of state and individual characteristics) on disparity in COVID-19 infections in black community. However, the medical care avoidance seems to reduce the disparity in COVID-19 infections. One explanation might be that the medical care avoidance may result in not being diagnosed with COVID-19 which decreases the number of positive cases and lead to lower disparity.[20] For the white communities, higher health insurance residuals significantly reduce the COVID-19 infections disparity, while medical care avoidance residuals have no impact.

Table 3.9 displays the two-step estimation results with disparities in COVID-19 mortality as the second step dependent variable.[21] According to this table, we find a significant reduction in disparity of COVID-19 mortality for black sub-sample as the net probability of having health insurance increases. While, health insurance seems to have no effect on COVID-19 mortality reduction in whites. For both black and white sub-samples, medical

[18] Table C.2 in the appendix reports the detailed estimation results corresponding to this table. The other observables show a similar impact on COVID-19 mortality disparity as they have on COVID-19 infections disparity except the hospital bed utilization which is significantly associated with reducing disparity in COVID-19 mortality.

[19] The detailed estimation of coefficients in step 1 (i.e., corresponding to equation (3.2)) are provided in Table C.3 in the appendix

[20] https://www.healthaffairs.org/do/10.1377/hblog20201023.55778/full/

[21] The detailed estimation of coefficients in step 1 (i.e., corresponding to equation (3.2)) are provided in Table C.3 in the appendix

care avoidance demonstrates no significant impact on mortality disparity reduction. The two-step analysis estimations not only confirm the findings from aggregate method but also reveal more potential relationships that could not be captured in the aggregate methods. This highlights the importance of adjustments for both individual and state-level explanatory variables in estimations. Findings in this study reflect the eminence of access to healthcare in COVID-19 disparity reduction in blacks as well as disparity reduction in COVID-19 infections in whites.

3.6 Discussion and Conclusion

In this paper, we investigate whether disparities in access to healthcare during the COVID-19 pandemic affects the widely-noted racial disparity in COVID-19 outcomes. We focus on black versus white subgroups, the two major racial groups comprise the U.S. population. To implement the analysis, we link multiple dataset at the state-level with individual-level HPS dataset in the period of April 23, 2020 and July 22, 2020.

We first explore the disparity in access to healthcare during the pandemic by employing BO decomposition method. Second, we explain the disparity in COVID-19 infections and mortality for black and white sub-samples, and finally, we explore how inequalities in healthcare access affect the disparity in COVID-19 infections and mortality by using two methods: aggregate method and the two-step approach. The two-step approach is more reliable than aggregate method, since it adjusts for the possible effects from the state and individual-level confounders.

Our results indicate that black communities are significantly less likely to have health insurance and more likely to avoid medical care as compared to their white sub-sample counterpart. Our findings exhibit that disparity in healthcare access in the form of health insurance has no impact on disparity in COVID-19 infections within black sub-sample, while it significantly lowers the disparity in COVID-19 infections for white sub-sample. We use medical care avoidance as second proxy for healthcare access and find that medical care avoidance significantly lowers the disparity of COVID-19 infections in black sub-sample, while it has no effect for the white sub-sample. With regard to COVID-19 mortality, we find that disparity in health insurance significantly reduces disparity in COVID-19 mortality in blacks. This results is robust across the two analysis methods. However, disparity in COVID-19 mortality for white sub-sample does not seem to be affected by disparities in health insurance. Medical care avoidance disparity has no impact on disparity in COVID-19 mortality of either blacks or whites.

These findings demonstrate how health insurance can ameliorate racial health disparities, especially in COVID-19 mortality in communities of color. The COVID-19 pandemic caused millions of people to lost their job and subsequently lost employer-based health insurance. A recent survey study shows that almost 40% of Americans who lost their job had employer-based health insurance (Collins et al., 2020). Even those who are employed may find drops in their health insurance since the employers seek to cut costs. (Cohen et al., 2020) predicts that around 40 million people will be uninsured before the pandemic ends. Lack of health insurance limits healthcare access, which leads to higher underlying health conditions and less well-controlled chronic conditions in minorities. Substantially, these conditions and illnesses further increase the risk of COVID-19 complications and mortality (Blumenthal et al., 2020). Therefore, it is important to make health insurance available to all, such as through expansions of the Affordable Care Act or similar plans that reduce the uninsured rate across all groups, especially black communities (Chaudry et al., 2019). This study provides supporting empirical evidence that eliminating healthcare access inequalities could be a potential mechanism for reducing racial disparity in COVID-19 outcomes or any other future pandemics.

This study has several limitations. First, some states with considerable numbers of infections and deaths, such as New York, Vermont, and North Dakota, do not break down the infections and deaths by race. And those states with reports on race do not reveal the detailed socioeconomic and healthcare access information of COVID-19 patients and deaths due to confidentiality rules. Lack of such information prevents us from precisely examining the underlying mechanism that healthcare access impacts the prevalence of COVID-19 in a certain sub-population. Second, it is well-known that the reported number of COVID-19 infections may not represent the actual number of COVID-19 infections due to lack of testing capacity or access to healthcare etc. (Stock et al., 2020). For example, it is highly likely that black individuals do not seek medical care and hence do not get counted among the infected cases (Czeisler et al., 2020). Third, in the HPS data, the respondents are followed only for three weeks which makes the panel structure not long enough to fully capture the trend of healthcare access during the COVID-19 pandemic. At the time this study is conducted, COVID-19 is surging more severely under its third wave. We plan to reinvestigate these issues as more detailed data become available by the end of 2021.

3.7 Table and Figures

Table 3.1: Summary of variables used in two-step approach

Regression	Type of variable	Variable	Level	Data source	Time-variant
Step 1	Dependent	Medical care access (health insurance[c], medical care avoidance[d])	HH[a]	HPS[b]	Yes
	Independent	gender, education[e], marital status[f], number of household members	HH	HPS	-[g]
		income[h], health status[i], employment status[j]	HH	HPS	Yes
		pop >60 years old, median income in state	state	Census	No
		unemployment rate, number of COVID-19 tests done per1000 people	state	UMD[k]	Yes
Step 2	Dependent	Disparity in COVID-19 (infections) mortality	state	Racial Tracker	Yes
	Independent	$\hat{u}_{0st}$	state	step 1 of equation	Yes

NOTES This Table presents summary of variables used in the two-step approach corresponding to equations (3.2) and (3.3). [a] HH stands for household. [b] HPS refers to Household Pulse Survey. [c] health insurance is a dummy variable which take value of 1 if individual has any health plan coverage, and 0 otherwise. [d] Medical care avoidance is a dummy variable for combination of two variables (1)delayed medical care (2) needed non-coronavirus medical care but did not get it due to the coronavirus pandemic in the past 4 weeks. If the answer to any of these questions is yes, the medical care avoidance takes value 1, and 0 otherwise. [e] Education is based on 7 categories that ranges from less than high schools to graduate degrees. [f] Marital status has 5 categories; ranges from now married until never married.[g] The demographics might change overtime if different members within a household takes the HPS survey in different waves. [h] Income includes 8 categories range from less than $25000 to more than $200,000. [i] Health status is self reported health and can be (1) Excellent, (2) Very good, (3) Good, (4) Fair, and (5) Poor. [j] Employment status is 1 if individual lost his/her job since march 13, 2020, and 0 otherwise. [k] UMD refers to University of Maryland Impact Analysis Platform.

Table 3.2: Summary statistics of HPS dataset

	White	Black	Total	Difference
Black	0	1	0.0832 (0.276)	
Age	52.49 (15.54)	47.72 (14.03)	52.1 (15.48)	4.777*** (0.06)
Female	0.587 (0.49)	0.7 (0.46)	0.597 (0.49)	-0.113*** (0.00)
Education [a]	5.351 (1.43)	4.962 (1.51)	5.319 (1.44)	0.389*** (0.01)
Marital Status [b]	2.103 (1.54)	2.964 (1.74)	2.174 (1.58)	-0.861*** (0.01)
Number of household members	2.766 (1.56)	3.055 (1.77)	2.79 (1.58)	-0.289*** (0.01)
Number of kids in household	0.619 (1.04)	0.874 (1.18)	0.64 (1.05)	-0.255*** (0.00)
Number of adults in household	2.147 (1.10)	2.182 (1.28)	2.15 (1.12)	-0.0345*** (0.00)
Income [c]	4.649 (2.06)	3.48 (2.00)	4.552 (2.08)	1.170*** (0.01)
Employment status [d]	0.381 (0.49)	0.495 (0.50)	0.391 (0.49)	0.114*** (0.00)
Health status [e]	2.336 (1.02)	2.689 (1.03)	2.366 (1.03)	-0.353*** (0.00)
Health Insurance [f]	0.945 (0.23)	0.896 (0.31)	0.941 (0.24)	0.0487*** (0.00)
Medical care avoidance [g]	0.297 (0.46)	0.311 (0.46)	0.298 -0.457	-0.0141*** (0.00)
Observations	780,555	70,803	851358	

NOTES This Table reports summary statistics of selected variables by race in Household Pulse Survey dataset for the time period between April 23, 2020 and July 22, 2020. The dataset covers all states in the US plus Washington D.C. [a] Education is based on 7 categories that ranges from less than high schools to graduate degrees. [b] Marital status has 5 categories; ranges from now married until never married. [c] Income includes 8 categories range from less than $25000 to more than $200,000.[d] Employment status is 1 if individual lost his/her job since march 13, 2020, and 0 otherwise. [e] Health status is self reported health and can be (1) Excellent, (2) Very good, (3) Good, (4) Fair, and (5) Poor. [f] health insurance is a dummy variable which take value of 1 if individual has any health plan coverage, and 0 otherwise. [g] Medical care avoidance is a dummy variable for combination of two variables (1)delayed medical care (2) needed non-coronavirus medical care but did not get it due to the coronavirus pandemic in the past 4 weeks. If the answer to any of these questions is yes, the medical care avoidance takes value 1, and 0 otherwise. Asterisks show that the difference between the black vs. white is statistically significant i.e., * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 3.3: Summary statistics of COVID-19 infections and mortality by Race

	Mean	Minimum	Maximum
COVID-19 Outcomes			
Percentage of blacks in the population	10.9 (10.40)	0.34	45
Percentage of whites in population	67.9 (16.18)	21	93
Percentage of blacks in the COVID-19 infections	26.7 (17.90)	0.4	75.6
Percentage of whites in the COVID-19 infections	73.3 (17.90)	24.4	99.6
Percentage of blacks in the COVID-19 mortality	25.1 (19.20)	0	88.3
Percentage of whites in the COVID-19 mortality [a]	74.9 (19.20)	11.7	100
State Variables			
Number of tests done per 1000 people	72.568 (47.24)	10.166	298.793
Unemployment rate	10.603 (3.35)	4.443	26.75
Social distancing Index [b]	31.911 (9.13)	18	71
Hospital bed utilization	52.764 (11.97)	28.791	102.724
Observations [c]	612		

NOTES This Table reports summary statistics of population decomposition and COVID-19 outcomes for black and white. The population data is obtained from US Census Bureau. The data source for the COVID-19 outcomes is "The COVID Tracking Project at The Atlantic". The state variables are obtained from "COVID-19 Impact Analysis Platform" database available from the University of Maryland. [a] The maximum of percentage of whites in the COVID-19 mortality is 100 because for some states there is no COVID-19 mortality record for black communities such as Vermont. [b] The Social Distancing Index (SDI) is defined by Maryland Transportation Institute (MTI) as follows: SDI = 0.8*[% stay-at-home + 0.01*(100 - %stay-at-home)*(0.1*% reduction all trips + 0.2*% reduction work trips + 0.4*% reduction non-work trips + 0.3*% reduction travel distance)] + 0.2*% reduction out-of-county trips. It is an integer between 0 (no social distancing at all) and 100 (100% of the residents in the state follow social distancing). [c] we have total of 51 states and 12 weeks as our time variables that covers the period between April 23 and July 22, 2020. There fore, we have 612 observations as combination of states and weeks.

Table 3.4: Blinder-Oaxaca decomposition of healthcare access for black and white

	Health Insurance [a]	Medical Care Avoidance [b]
Differential		
Prediction_White	0.946*** (0.00)	0.296*** (0.00)
Prediction_Black	0.898*** (0.00)	0.310*** (0.00)
Difference	0.048*** (0.00)	-0.014*** (0.00)
Explained [c]		
Demographics [d]	0.015*** (0.00)	0.006*** (0.00)
Employment status[e] and Income	0.021*** (0.00)	-0.007*** (0.00)
Health Status [f]	0.000*** (0.00)	-0.034*** (0.00)
Total	0.037*** (0.00)	-0.035*** (0.00)
Unexplained		
Demographics	-0.009** (0.00)	-0.019*** (0.01)
Employment status and Income	0.004*** (0.00)	0.001 (0.00)
Health Status	-0.001 (0.00)	-0.002 (0.00)
Constant	0.018*** (0.01)	0.041*** (0.01)
Total [g]	0.011*** (0.00)	0.021*** (0.00)
Observations	888636	886304

NOTES: This table displays the Blinder-Oaxaca decomposition results of health insurance and medical care avoidance by race. [a] health insurance is a dummy variable which take value of 1 if individual has any health plan coverage, and 0 otherwise. [b] Medical care avoidance is a dummy variable for combination of two variables (1)delayed medical care (2) needed non-coronavirus medical care but did not get it due to the coronavirus pandemic in the past 4 weeks. If the answer to any of these questions is yes, the medical care avoidance takes value 1, and 0 otherwise. [c] Explained means portion of the difference between the outcome for black and white that can be explained by observables list in this table. [d] Demographics include: Female, age, education, Marital status. [e] Employment status is 1 if individual lost his/her job since march 13, 2020, and 0 otherwise. [f] Health status is self reported health and can be (1) Excellent, (2) Very good, (3) Good, (4) Fair, and (5) Poor. [g] Total refers to the total amount of difference that cannot be explained by the observables and is interpreted as racial disparity. Asterisks present statistical significance levels as follow, * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 3.5: Summary statistics of Disparity variable for black and white

	Observations	Mean [a]	sd [b]	Min[c]	Max[d]	TTest [e] t-statistics	p-value
Black disparity in the COVID-19 infections	568	0.16	0.10	0.005	0.40	39.65	0
White disparity in the COVID-19 infections	568	0.07	0.19	-0.23	0.37	9	0
Black disparity in the COVID-19 mortality	518	0.13	0.11	-0.03	0.46	27.7	0
White disparity in the COVID-19 mortality	518	-0.10	0.17	-0.23	0.05	12.95	0

NOTES: This table reports the summary statistics of disparity in COVID-19 infections and mortality for black and white communities, separately. [a] Mean refers to the average of a variable over all states and weeks. [b] sd refers to the standard deviation of a variable. [c] Min refers to the minimum of a variable over states and weeks. [d] Max refers to the maximum of the variable over states and weeks. *e*The TTest conducted for comparing the mean disparity with 5% according to Moore et al. (2020).

Table 3.6: Aggregate method estimation of disparity in healthcare access on disparity in COVID-19 infection

	Disparity in COVID-19 Infections [a]			
	black		white	
Lag of health insurance disparity [a]	0.0398 (0.03)		0.0366 (0.03)	
Lag of disparity medical care avoidance [b]		-0.0202 (0.09)		-0.026 (0.09)
Observations	417	463	477	477

NOTES: This table reports the aggregate method estimation results of healthcare access disparity on disparity in COVID-19 infections. We have controlled for the state-fixed effects. More details of these tables are presented in the Table C.1 in the appendix. [a] Disparity in COVID-19 infections for a specific racial group is defined as difference between group's portion of COVID-19 infections and portion of population consisting of that group. [b] health insurance disparity for a specific racial group is defined as difference between portion of people without health insurance and their share in the population. [c] disparity in medical care avoidance for a specific racial group refers to the difference between potion of people who avoid getting medical care and their share in the population. Asterisks present statistical significance levels as follow, * p<0.10, ** p<0.05, *** p<0.01.

Table 3.7: Aggregate method estimation of disparity in healthcare access on disparity in COVID-19 mortality

	Disparity in COVID-19 mortality [a]			
	black		white	
Lag 2 of health insurance disparity [b]	0.0269** (0.01)		0.0242* (0.01)	
Lag 2 of disparity in medical care avoidance [c]		0.0468 (0.05)		0.0456 (0.05)
Observations	414	434	440	440

NOTES: This table reports the aggregate method estimation results of healthcare access disparity on disparity in COVID-19 mortality. We have controlled for the state-fixed effects. More details of these tables are presented in the Table C.2 in the appendix. [a] Disparity in COVID-19 mortality for a specific racial group is defined as difference between group's portion of COVID-19 mortality and portion of population consisting of that group. [b] health insurance disparity for a specific racial group is defined as difference between portion of people without health insurance and their share in the population. [c] disparity in medical care avoidance for a specific racial group refers to the difference between potion of people who avoid getting medical care and their share in the population. Asterisks present statistical significance levels as follow, * p<0.10, ** p<0.05, *** p<0.01.

Table 3.8: Two-step estimation of healthcare access impact on disparity in COVID-19 infections

	Disparity in COVID-19 Infections [a]			
	black		white	
Lag of health insurance residuals [b]	-0.02 (0.16)		-0.45** (0.23)	
Lag of medical care avoidance [c]		-1.51** (0.64)		0.02 (0.10)
Observations	522	522	523	523

NOTES: This table reports the aggregate method estimation results of healthcare access disparity on disparity in COVID-19 infections. We have controlled for state variables such as population density, population of older than 60 years old, unemployment rate, hospital bed utilization. Estimation results of step 1 are presented in the Table C.3 in the appendix. [a] Disparity in COVID-19 infections for a specific racial group is defined as difference between group's portion of COVID-19 infections and portion of population consisting of that group. [b] health insurance residuals refers to estimation of state-week random intercept (u_{0st}) in equation(3.2) with health insurance as dependent variable. [c] Medical care avoidance residual refers to estimation of state-week random intercept (u_{0st}) in equation(3.2) with medical care avoidance as dependent variable. Asterisks present statistical significance levels as follow, * p<0.10, ** p<0.05, *** p<0.01.

Table 3.9: Two-step estimation of healthcare access impact on disparity in COVID-19 mortality

	Disparity in COVID-19 Mortality [a]			
	black		white	
Lag 2 of health insurance residuals [b]	-0.17** (0.08)		0.1 (0.12)	
Lag 2 of medical care avoidance [c]		-0.2 (0.33)		0.09 (0.06)
Observation	440	440	440	440

NOTES: This table reports the aggregate method estimation results of healthcare access disparity on disparity in COVID-19 mortality. We have controlled for state variables such as population density, population of older than 60 years old, unemployment rate, hospital bed utilization. Estimation results of step 1 are presented in the Table C.3 in the appendix. [a] Disparity in COVID-19 infections for a specific racial group is defined as difference between group's portion of COVID-19 infections and portion of population consisting of that group. [b] health insurance residuals refers to estimation of state-week random intercept (u_{0st}) in equation (3.2) with health insurance as dependent variable. [c] Medical care avoidance residual refers to estimation of state-week random intercept (u_{0st}) in equation (3.2) with medical care avoidance as dependent variable. Asterisks present statistical significance levels as follow, * p<0.10, ** p<0.05, *** p<0.01.

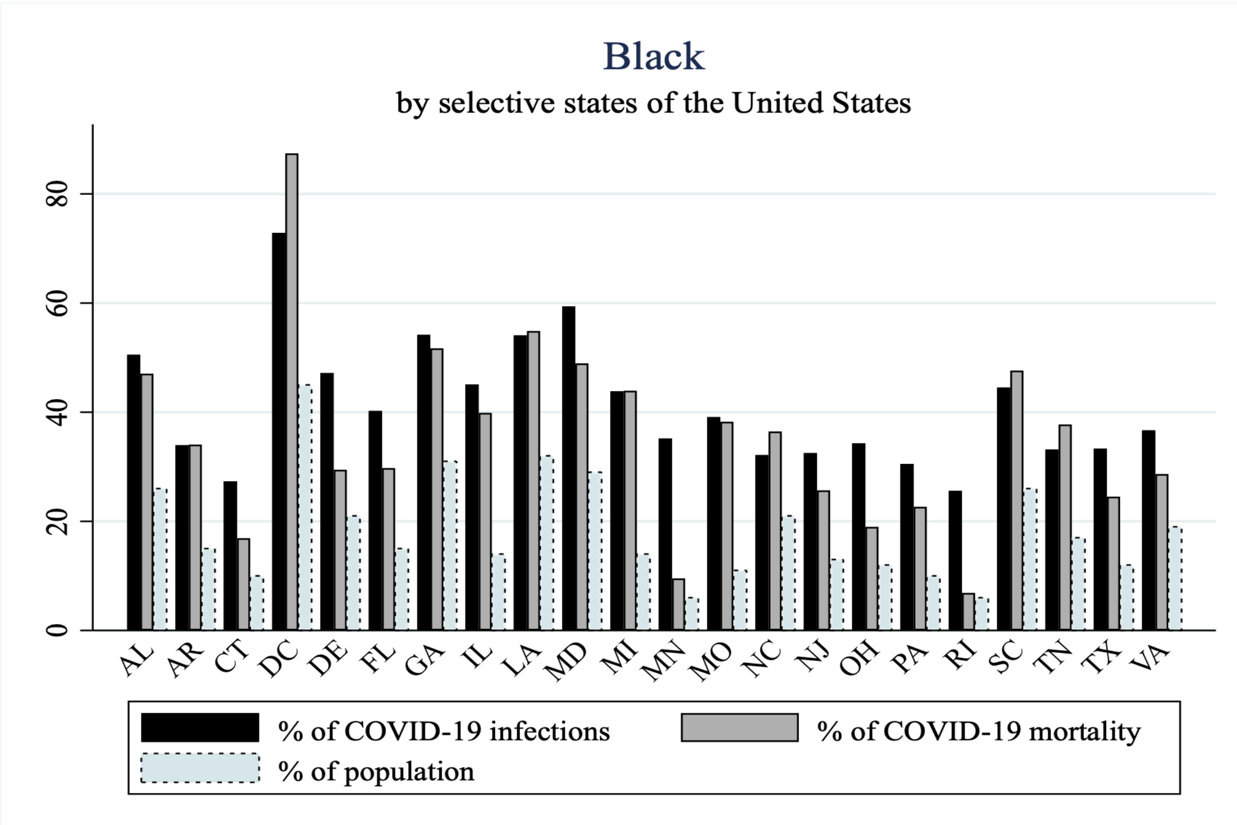


Figure 3.1: Black portion of COVID-19 infections and mortality compared with their population share. Black and grey bars display the black share in COVID-19 infections and mortality, respectively. Light blue bars with dashed outline represent percentage of blacks in the state.

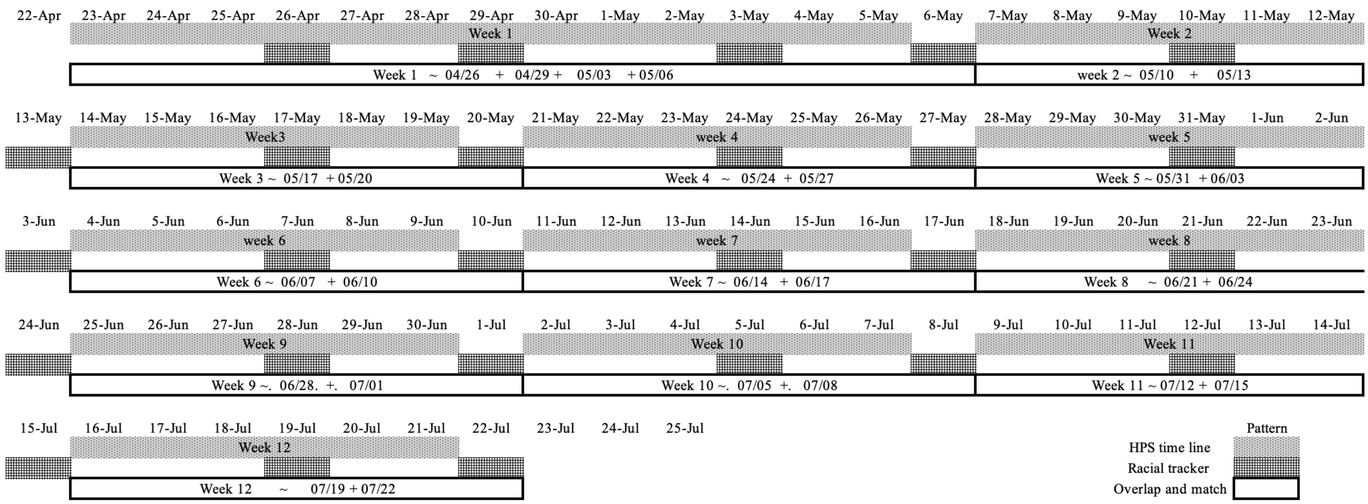


Figure 3.2: Matching the time line of data in HPS, racial tracker

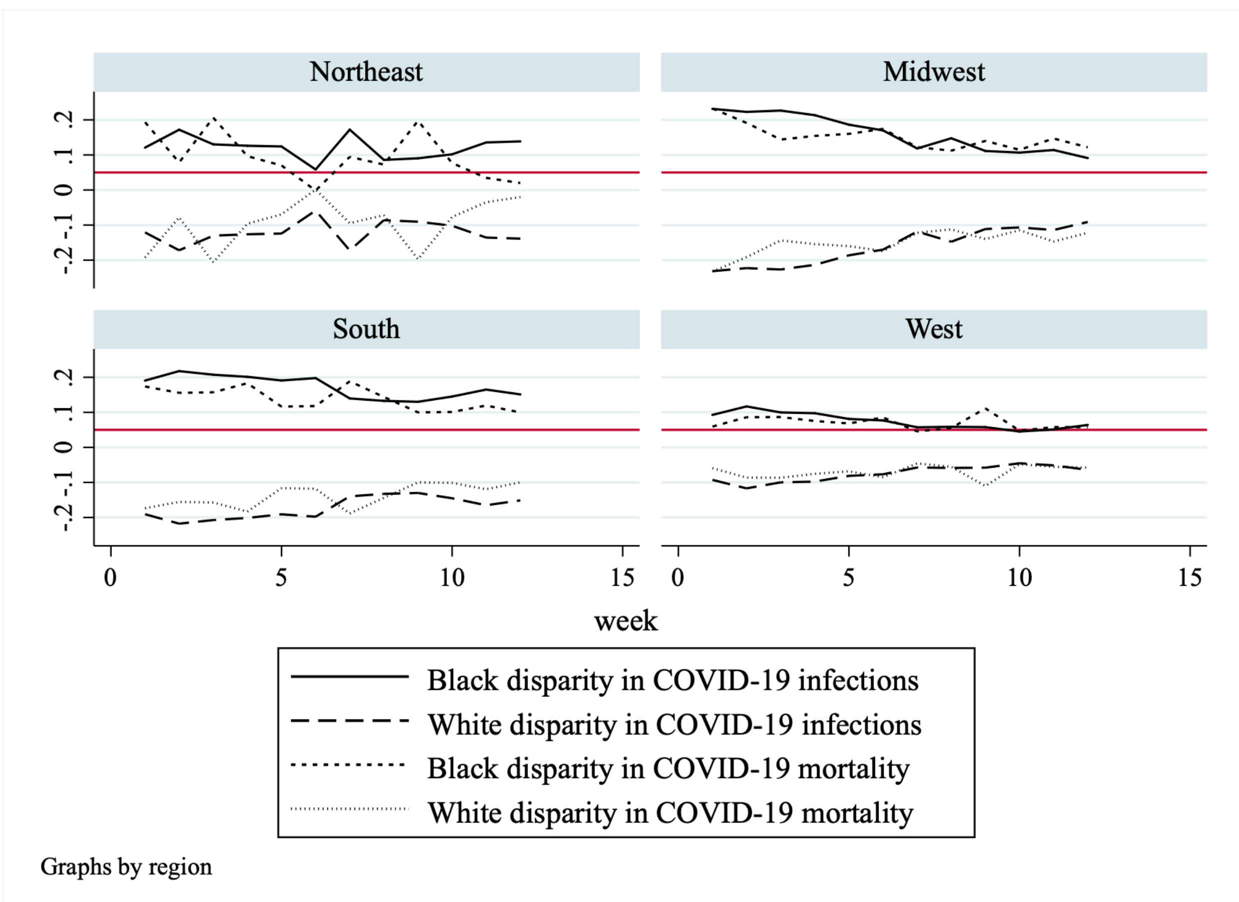


Figure 3.3: Infections and mortality disparity for black and whites by regions.

Note: The red horizontal line is the 0.05 line for comparison of disparities according to Moore et al. (2020)

Bibliography

Acevedo-Garcia, D., Lochner, K. A., Osypuk, T. L., and Subramanian, S. V. (2003). Future directions in residential segregation and health research: a multilevel approach. *American journal of public health*, 93(2):215–221.

Andrulis, D. P. (1998). Access to care is the centerpiece in the elimination of socioeconomic disparities in health.

Ashton, C. M., Haidet, P., Paterniti, D. A., Collins, T. C., Gordon, H. S., O'Malley, K., Petersen, L. A., Sharf, B. F., Suarez-Almazor, M. E., Wray, N. P., et al. (2003). Racial and ethnic disparities in the use of health services. *Journal of general internal medicine*, 18(2):146–152.

Bailey, Z. D., Feldman, J. M., and Bassett, M. T. (2020). How structural racism works—racist policies as a root cause of us racial health inequities.

Becker, D., Breustedt, W., and Zuber, C. I. (2018). Surpassing simple aggregation: Advanced strategies for analyzing contextual-level outcomes in multilevel models. *Methods, data, analyses: a journal for quantitative methods and survey methodology (mda)*, 12(2):233–263.

Bell, C. N., Thorpe Jr, R. J., Bowie, J. V., and LaVeist, T. A. (2018). Race disparities in cardiovascular disease risk factors within socioeconomic status strata. *Annals of epidemiology*, 28(3):147–152.

Berchick, E. R., Hood, E., and Barnett, J. C. (2018). Health insurance coverage in the united states: 2017. *Current population reports. Washington DC: US Government Printing Office.*

Binswanger, I. A., Stern, M. F., Deyo, R. A., Heagerty, P. J., Cheadle, A., Elmore, J. G., and Koepsell, T. D. (2007). Release from prison—a high risk of death for former inmates. *New England Journal of Medicine*, 356(2):157–165.

Blumenthal, D., Fowler, E. J., Abrams, M., and Collins, S. R. (2020). Covid-19—implications for the health care system.

Chaudry, A., Jackson, A., and Glied, S. A. (2019). Did the affordable care act reduce racial and ethnic disparities in health insurance coverage? *New York: Commonwealth Fund, August*, 21.

Chowkwanyun, M., Reed, J., and Adolph, L. (2020). Racial health disparities and covid-19—caution and context. *New England Journal of Medicine*.

Chunara, R., Zhao, Y., Chen, J., Lawrence, K., Testa, P. A., Nov, O., and Mann, D. M. (2020). Telemedicine and healthcare disparities: A cohort study in a large healthcare system in new york city during covid-19. *Journal of the American Medical Informatics Association*.

Cohen, R., Terlizzi, E., Martinez, M., and Cha, A. (2020). Health insurance coverage: early release of estimates from the national health interview survey, january–june 2019 [internet]. hyattsville (md): National center for health statistics; 2020 may [cited 2020 nov 9].

Collins, S. R., Gunja, M. Z., Aboulafia, G. N., Czyzewicz, E., Kline, C. H., Rapoport, R., and Glancey, S. (2020). An early look at the potential implications of the covid-19 pandemic for health insurance coverage. *Rama on healthcare. In https://ramaonhealthcare. com/an-early-look-at-the-potential-implications-of-the-covid-19-pandemic-for-health-insurance-coverage*.

Croon, M. A. and van Veldhoven, M. J. (2007). Predicting group-level outcome variables from variables measured at the individual level: a latent variable multilevel model. *Psychological methods*, 12(1):45.

Czeisler, M. É., Marynak, K., Clarke, K. E., Salah, Z., Shakya, I., Thierry, J. M., Ali, N., McMillan, H., Wiley, J. F., Weaver, M. D., et al. (2020). Delay or avoidance of medical care because of covid-19–related concerns—united states, june 2020. *Morbidity and mortality weekly report*, 69(36):1250.

DiPrete, T. A. and Forristal, J. D. (1994). Multilevel models: methods and substance. *Annual Review of Sociology*, 20(1):331–357.

Garg, S. (2020). Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019—covid-net, 14 states, march 1–30, 2020. *MMWR. Morbidity and mortality weekly report*, 69.

Gould, E. and Shierholz, H. (2020). Not everybody can work from home: Black and hispanic workers are much less likely to be able to telework. *Economic Policy Institute. https://www. epi. org/blog/black-and-hispanic-workers-are-much-less-likely-to-be-able-towork-from-home/. Accessed April*, 16.

Gravlee, C. C. (2020). Systemic racism, chronic health inequities, and covid-19: A syndemic in the making? *American Journal of Human Biology*.

Griffin, M. A. (1997). Interaction between individuals and situations: Using hlm procedures to estimate reciprocal relationships. *Journal of Management*, 23(6):759–773.

Hall, W. J., Chapman, M. V., Lee, K. M., Merino, Y. M., Thomas, T. W., Payne, B. K., Eng, E., Day, S. H., and Coyne-Beasley, T. (2015). Implicit racial/ethnic bias among health care professionals and its influence on health care outcomes: a systematic review. *American journal of public health*, 105(12):e60–e76.

Harper, S. and Lynch, J. (2005). Methods for measuring cancer disparities: using data relevant to healthy people 2010 cancer-related objectives. Technical report.

Johnson, A. (2020). On the minds of black lives matter protesters: A racist health system. *ProPublica. org, June*, 5.

Kilbourne, A. M. (2005). Care without coverage: too little, too late. *Journal of the National Medical Association*, 97(11):1578.

Knox, D., Lowe, W., and Mummolo, J. (2020). Administrative records mask racially biased policing. *American Political Science Review*, 114(3):619–637.

Kullar, R., Marcelin, J. R., Swartz, T. H., Piggott, D. A., Macias Gil, R., Mathew, T. A., and Tan, T. (2020). Racial disparity of coronavirus disease 2019 in african american communities. *The Journal of infectious diseases*, 222(6):890–893.

Kutateladze, B. L., Andiloro, N. R., Johnson, B. D., and Spohn, C. C. (2014). Cumulative disadvantage: Examining racial and ethnic disparity in prosecution and sentencing. *Criminology*, 52(3):514–551.

Levine, R. S., Foster, J. E., Fullilove, R. E., Fullilove, M. T., Briggs, N. C., Hull, P. C., Husaini, B. A., and Hennekens, C. H. (2016). Black-white inequalities in mortality and life expectancy, 1933–1999: implications for healthy people 2010. *Public health reports*.

Lim, F., Bond, M. H., and Bond, M. K. (2005). Linking societal and psychological factors to homicide rates across nations. *Journal of Cross-Cultural Psychology*, 36(5):515–536.

Louis-Jean, J., Cenat, K., Njoku, C. V., Angelo, J., and Sanon, D. (2020). Coronavirus (covid-19) and racial disparities: a perspective analysis. *Journal of racial and ethnic health disparities*, 7(6):1039–1045.

Maina, I. W., Belton, T. D., Ginzberg, S., Singh, A., and Johnson, T. J. (2018). A decade of studying implicit racial/ethnic bias in healthcare providers using the implicit association test. *Social Science & Medicine*, 199:219–229.

Manuel, J. I. (2018). Racial/ethnic and gender disparities in health care use and access. *Health services research*, 53(3):1407–1429.

Maxwell, C. (2020). Coronavirus compounds inequality and endangers communities of color. *Center for American Progress*.

McLaren, J. (2020). Racial disparity in covid-19 deaths: Seeking economic roots with census data. *NBER working paper*, (w27407).

Millett, G. A., Jones, A. T., Benkeser, D., Baral, S., Mercer, L., Beyrer, C., Honermann, B., Lankiewicz, E., Mena, L., Crowley, J. S., et al. (2020). Assessing differential impacts of covid-19 on black communities. *Annals of Epidemiology*.

Moore, J. T., Ricaldi, J. N., Rose, C. E., Fuld, J., Parise, M., Kang, G. J., Driscoll, A. K., Norris, T., Wilson, N., Rainisch, G., et al. (2020). Disparities in incidence of covid-19 among underrepresented racial/ethnic groups in counties identified as hotspots during june 5–18, 2020—22 states, february–june 2020. *Morbidity and Mortality Weekly Report*, 69(33):1122.

Murata, C. and Kondo, K. (2020). Access to healthcare and health disparities. In *Social Determinants of Health in Non-communicable Diseases*, pages 199–206. Springer, Singapore.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International economic review*, pages 693–709.

Rabe-Hesketh, S. and Skrondal, A. (2008). *Multilevel and longitudinal modeling using Stata*. STATA press.

Rogers, T. N., Rogers, C. R., VanSant-Webb, E., Gu, L. Y., Yan, B., and Qeadan, F. (2020). Racial disparities in covid-19 mortality among essential workers in the united states. *World medical & health policy*, 12(3):311–327.

Sealy-Jefferson, S., Vickers, J., Elam, A., and Wilson, M. R. (2015). Racial and ethnic health disparities and the affordable care act: a status update. *Journal of racial and ethnic health disparities*, 2(4):583–588.

Selden, T. M. and Berdahl, T. A. (2020). Covid-19 and racial/ethnic disparities in health risk, employment, and household composition: Study examines potential explanations for racial-ethnic disparities in covid-19 hospitalizations and mortality. *Health Affairs*, 39(9):1624–1632.

Sen, B. (2014). Using the oaxaca–blinder decomposition as an empirical tool to analyze racial disparities in obesity. *Obesity*, 22(7):1750–1755.

Stock, J. H., Aspelund, K. M., Droste, M., and Walker, C. D. (2020). Estimates of the undetected rate among the sars-cov-2 infected using testing data from iceland. *medRxiv*.

Testa, C. C., Krieger, N., Chen, J. T., and Hanage, W. P. (2020). Visualizing the lagged connection between covid-19 cases and deaths in the united states: an animation using per capita state-level data (january 22, 2020-july 8, 2020). *Harvard Center for Population and Development Studies Working Paper*, 19(4).

Wiemers, E. E., Abrahams, S., AlFakhri, M., Hotz, V. J., Schoeni, R. F., and Seltzer, J. A. (2020). Disparities in vulnerability to severe complications from covid-19 in the united states. Technical report, National Bureau of Economic Research.

Williams, D. R. (2001). Wednesday, october 24, 2001-9: 10 am abstract# 32866 racial residential segregation: A fundamental cause of racial disparities in health. In *The 129th Annual Meeting of APHA*.

Williams, D. R. and Cooper, L. A. (2019). Reducing racial inequities in health: Using what we already know to take action. *International journal of environmental research and public health*, 16(4):606.

Williams, M. B., Mitchell, F., Thomson, G. E., et al. (2006). *Examining the health disparities research plan of the National Institutes of Health: unfinished business*. National Academies Press.

Yancy, C. W. (2020). Covid-19 and african americans. *Jama*.

Zhang, L., Ghader, S., Pack, M. L., Xiong, C., Darzi, A., Yang, M., Sun, Q., Kabiri, A., and Hu, S. (2020). An interactive COVID-19 mobility impact and social distancing analysis platform. *medRxiv*.

Appendix C

(Third Chapter Appendix)

C.1 Blinder-Oaxaca Decomposition

The Blinder-Oaxaca decomposition partitions the group differences into an "explained" portion and an "unexplained" portion. The "explained" portion of this gap is the difference in the outcome attributable to the group differences in the mean levels of observables, while the "unexplained" part is often interpreted as disparity and discrimination in the outcome between the two groups, which may include the group differences in unobservables as well. Suppose there are two groups, black (B) and whites (W); H is the outcome variable (healthcare access); and a X is set of explanatory variables such as demographics, socioeconomic status, and etc. BO decomposition explains how much of the mean difference in healthcare access $(E(H_W) - E(H_B))$ is accounted for group differences in the explanatory variables. Given the base linear model:

$$\mathbf{H}_g = \mathbf{X}'_g \beta_g + \epsilon_g, \quad E(\epsilon_g) = 0, \quad g \in (W, B) \tag{C.1}$$

where $\mathbf{X}$ is a vector of predictors and a constant, β contains the slope parameters and the intercept, and ϵ is the error term, the mean outcome difference can be expressed as the difference in the linear prediction at the group-specific means of the explanatory variables as follows:

$$E(\mathbf{H}_W) - E(\mathbf{H}_B) = E(\mathbf{X}'_W)\beta_W - E(\mathbf{X}'_B)\beta_B \tag{C.2}$$

We can further decompose the difference between the two expected outcomes as follows:

$$E(\mathbf{H}_W) - E(\mathbf{H}_B) = E(\mathbf{X}'_W)\beta_W - E(\mathbf{X}'_B)\beta_B = E(\mathbf{X}'_W)(\beta_W - \beta_B) + (E(\mathbf{X}'_W) - E(\mathbf{X}'_B))'\beta_B \tag{C.3}$$

In the RHS of equation (C.3), the first expression is the coefficient effect that interprets as disparity and the second term presents the part that is explained by group differences in explanatory variables called endowment effects.

C.2 Supplementary Tables

Table C.1: Aggregate method estimation of disparity in healthcare access on disparity in COVID-19 infection

	Disparity in COVID-19 Infections [a]			
	black		white	
Lag of health insurance disparity [a]	0.0398 (0.03)		0.0366 (0.03)	
Lag of disparity medical care avoidance [b]		-0.0202 (0.09)		-0.026 (0.09)
Number of tests done per 1000 people	-0.0001*** (0.00)	-0.0001*** (0.00)	-0.0001 (0.00)	-0.0001 (0.00)
Unemployment rate	0.0045*** (0.00)	0.0036*** (0.00)	0.0013 (0.00)	0.0011 (0.00)
Lag 2 of Social distancing Index	-0.0003 (0.00)	-0.0005 (0.00)	-0.0003 (0.00)	-0.0002 (0.00)
Hospital bed utilization	-0.0018*** (0.00)	-0.0014*** (0.00)	-0.0038*** (0.00)	-0.0039*** (0.00)
Constant	0.1006*** (0.03)	0.0945*** (0.03)	-0.0914 (0.06)	-0.0884 (0.06)
Observations	417	463	477	477

NOTES: This table reports the aggregate method estimation results of healthcare access disparity on disparity in COVID-19 infections. We have controlled for the state-fixed effects. [a] Disparity in COVID-19 infections for a specific racial group is defined as difference between group's portion of COVID-19 infections and portion of population consisting of that group. [b] health insurance disparity for a specific racial group is defined as difference between portion of people without health insurance and their share in the population. [c] disparity in medical care avoidance for a specific racial group refers to the difference between potion of people who avoid getting medical care and their share in the population. Asterisks present statistical significance levels as follow, * p<0.10, ** p<0.05, *** p<0.01.

Table C.2: Aggregate method estimation of disparity in healthcare access on disparity in COVID-19 mortality

	Disparity in COVID-19 mortality [a]			
	black		white	
Lag 2 of health insurance disparity [b]	0.0269**		0.0242*	
	(0.01)		(0.01)	
Lag 2 of disparity in medical care avoidance [c]		0.0468		0.0456
		(0.05)		(0.05)
Number of tests done per 1000 people	-0.0001***	-0.0001***	0.0005***	0.0005***
	(0.00)	(0.00)	(0.00)	(0.00)
Unemployment rate	0.0014**	0.0013**	0.0109**	0.0109**
	(0.00)	(0.00)	(0.00)	(0.00)
Lag 2 of Social distancing Index	-0.0005***	-0.0005***	-0.0062***	-0.0061***
	(0.00)	(0.00)	(0.00)	(0.00)
Hospital bed utilization	-0.0006*	-0.0005	0.0043*	0.0041*
	(0.00)	(0.00)	(0.00)	(0.00)
constant	0.1140***	0.1043***	-0.3112**	-0.3160**
	(0.02)	(0.02)	(0.13)	(0.13)
Observations	414	434	440	440

NOTES: This table reports the aggregate method estimation results of healthcare access disparity on disparity in COVID-19 mortality. We have controlled for the state-fixed effects. [a] Disparity in COVID-19 mortality for a specific racial group is defined as difference between group's portion of COVID-19 mortality and portion of population consisting of that group. [b] health insurance disparity for a specific racial group is defined as difference between portion of people without health insurance and their share in the population. [c] disparity in medical care avoidance for a specific racial group refers to the difference between potion of people who avoid getting medical care and their share in the population. Asterisks present statistical significance levels as follow, * p<0.10, ** p<0.05, *** p<0.01.

Table C.3: Step 1 estimation results in the two-step approach

	Black sub-sample		White sub-sample	
	Health Insurance [a]	Medical Care Avoidance [b]	Health Insurance	Medical Care Avoidance
Individual-level variables				
Female	0.0475***	0.0670***	0.0170***	0.0461***
	(0.00)	(0.00)	(0.00)	(0.00)
Age	0.0026***	-0.0005***	0.0017***	-0.0004***
	(0.00)	(0.00)	0.00	0.00
Number of household members	0.0061***	0.0094***	0.0013***	0.0098***
	(0.00)	(0.00)	(0.00)	(0.00)
Education				
2. Some high school	0.0664***	-0.0108	0.0398***	-0.0084
	(0.01)	(0.02)	(0.00)	(0.01)
3. High school graduate or equivalent (for example GED)	0.0836***	0.0079	0.1002***	0.0428***
	(0.01)	(0.02)	(0.00)	(0.01)
4. Some college, but degree not received or is in progress	0.1176***	0.0943***	0.1250***	0.1156***
	(0.01)	(0.02)	(0.00)	(0.01)
5. Associate's degree (for example AA, AS)	0.1240***	0.0990***	0.1332***	0.1166***
	(0.01)	(0.02)	(0.00)	(0.01)
6. Bachelor's degree (for example BA, BS, AB)	0.1269***	0.1201***	0.1503***	0.1185***
	(0.01)	(0.02)	(0.00)	(0.01)
7. Graduate degree (for example master's, professional, doctorate)	0.1346***	0.1493***	0.1489***	0.1461***
	(0.01)	(0.02)	(0.00)	(0.01)
Marital Status				
2. Widowed	0.0004	-0.0082	0.0053***	-0.0178***
	(0.01)	(0.01)	(0.00)	(0.00)
3. Divorced	-0.0171***	0.0150***	-0.0195***	0.0174***
	(0.00)	(0.01)	(0.00)	(0.00)
4. Separated	-0.0215***	0.0179**	-0.0373***	0.0233***
	(0.01)	(0.01)	(0.00)	(0.00)
5. Never married	-0.0211***	-0.0220***	-0.0199***	-0.0416***
	(0.00)	(0.00)	(0.00)	(0.00)
Income				
2. $25,000 - $34,999	0.0234***	0.0013	0.0144***	-0.0137***
	(0.00)	(0.01)	(0.00)	(0.00)
3. $35,000 - $49,999	0.0567***	0.0092	0.0353***	-0.0038*
	(0.00)	(0.01)	(0.00)	(0.00)
4. $50,000 - $74,999	0.0723***	0.0163***	0.0628***	-0.0001
	(0.00)	(0.01)	(0.00)	(0.00)
5. $75,000 - $99,999	0.0892***	0.0143**	0.0804***	0.0025
	(0.00)	(0.01)	(0.00)	(0.00)
6. $100,000 - $149,999	0.1032***	0.0261***	0.0925***	0.0026
	(0.00)	(0.01)	(0.00)	(0.00)
7.$150,000 - $199,999	0.1065***	0.0158*	0.0959***	-0.0055**
	(0.01)	(0.01)	(0.00)	(0.00)
8.$200,000 and above	0.1050***	0.0224**	0.0963***	-0.0023
	(0.01)	(0.01)	(0.00)	(0.00)

Continued to the next page

Continued from previous page

	Black sub-sample		White sub-sample	
	Health Insurance [a]	Medical Care Avoidance [b]	Health Insurance	Medical Care Avoidance
Individual-level variables				
Health status				
2.Very good	0.0044	0.0692***	0.0005	0.0724***
	(0.00)	(0.01)	(0.00)	(0.00)
3.Good	0.0047	0.1429***	-0.0062***	0.1701***
	(0.00)	(0.01)	(0.00)	(0.00)
4.Fair	0.0127***	0.2641***	-0.0064***	0.3002***
	(0.00)	(0.01)	(0.00)	(0.00)
5.Poor	0.0099	0.3698***	0.0093***	0.4057***
	(0.01)	(0.01)	(0.00)	(0.00)
State-level variables				
Employment Status	0.0862***	-0.0599***	0.0556***	-0.0783***
	(0.00)	(0.00)	(0.00)	(0.00)
Population density	0.000	0.000	0 .000	0.0000***
	(0.00)	(0.00)	(0.00)	(0.00)
State-median Income	0.0000***	0.0000	0.0000***	0.0000***
	(0.00)	(0.00)	(0.00)	(0.00)
Number of tests done per 1000 people	0.0001***	-0.0001***	0.0001***	-0.0001***
	(0.00)	(0.00)	(0.00)	(0.00)
Unemployment rate	0.0038***	0.0018***	0.0009***	0.0032***
	(0.00)	(0.00)	(0.00)	(0.00)
Percentage of people older than 60 years old	0.0038***	-0.0017*	0.0022***	0.0010**
	(0.00)	(0.00)	(0.00)	(0.00)
Constant	0.2435***	0.1205***	0.4676***	0.0857***
	(0.03)	(0.03)	(0.01)	(0.02)
Variance Component[c]	.017	.012	.008	.019
Observations	71118	71216	782749	783298

NOTES: This table reports estimation results of step 1 in two-step approach. [a] health insurance is a dummy variable which take value of 1 if individual has any health plan coverage, and 0 otherwise. [b] Medical care avoidance is a dummy variable for combination of two variables (1)delayed medical care (2) needed non-coronavirus medical care but did not get it due to the coronavirus pandemic in the past 4 weeks. If the answer to any of these questions is yes, the medical care avoidance takes value 1, and 0 otherwise.[c] It refers to the standard deviation of random effect component. Asterisks present statistical significance levels as follow, * p<0.10, ** p<0.05, *** p<0.01.