Title
Essays in Applied Microeconomics

Author
Sheng, Yanying

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Essays in Applied Microeconomics

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

in

Economics

by

Yanying Sheng

Committee in charge:

Professor Julie Cullen, Chair
Professor Eli Berman
Professor Gordon Dahl
Professor Itzik Fadlon
Professor Ruixue Jia

2022

The dissertation of Yanying Sheng is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

TABLE OF CONTENTS

Dissertation Approval Page iii

Table of Contents iv

List of Figures vi

List of Tables vii

Acknowledgements x

Vita xi

Abstract of the Dissertation xii

Chapter 1 How Racial Animus Forms and Spreads: Evidence from the Coronavirus Pandemic 1
- 1.1 Introduction 1
- 1.2 Measures of Racial Animus 6
  - 1.2.1 Google and Twitter Proxies 6
  - 1.2.2 Relationship between Racial Animus, Hate Crimes, and Restaurant Visits 8
- 1.3 Evolution of Racial Animus in US amid the Pandemic 10
- 1.4 Evidence from DID Event Study 11
  - 1.4.1 Data and Empirical Strategy 11
  - 1.4.2 Effects of the First Local Case on Racial Animus 14
  - 1.4.3 What Motivates Racial Animus and Who Responds the Most 18
- 1.5 Factors Fueling Racial Animus 21
- 1.6 Conclusion 25
- 1.7 Conclusion 26

Chapter 2 Doctor Who — Can physicians from low socio-economic status families reduce the socio-economic gradient in health? 38
- 2.1 Institutional Settings and Data 42
  - 2.1.1 Data 44
  - 2.1.2 Descriptive Statistics 49
- 2.2 Identification Strategy 51
  - 2.2.1 Estimation Equation 53
  - 2.2.2 Identifying Assumption and Validity 54
- 2.3 Effects of physician-patient SES concordance 54
  - 2.3.1 SES Concordance Effects on Mortality 55
  - 2.3.2 SES Concordance Effects on Health behaviors 58
  - 2.3.3 Mechanisms 63

2.3.4 Robustness Checks 67
2.4 Conclusion 69
2.5 Acknowledgement 70

Chapter 3 Employer Reputation and the Online Labor Market: Evidence from Glassdoor.com and Dice.com 84
3.1 Introduction 84
3.2 Institutional Background and Data 88
3.2.1 Reputation data from Glassdoor.com 88
3.2.2 Job posting and applications data from Dice.com 89
3.2.3 Constructing analysis sample 90
3.3 Identification Strategy 92
3.4 Validity of RD Design 93
3.5 Results 95
3.6 Robustness 99
3.6.1 Job level analysis 100
3.6.2 Concerns related to outlier 101
3.7 Conclusion 101
3.8 Acknowledgement 102

Appendix A Appendix For Chapter One 123
A.1 Rescaled Google Search Index 123
A.2 Additional Figures & Tables 125

Appendix B Appendix For Chapter Two 142
B.1 Additional Figures & Tables 142

References 172

LIST OF FIGURES

Figure 1.1: The Effect of the First Local COVID-19 Diagnosis on Racial Animus . . . 27
Figure 1.2: The Effect of the First Local COVID-19 Diagnosis on Racially Charged Tweets COVID-Related vs Non-COVID-Related Tweets 28
Figure 1.3: The Effect of the First Local COVID-19 Diagnosis on Racially Charged Tweets First-time vs Existing Ch-word Users 29
Figure 1.4: Predictors of being First-time Ch-word Users 30
Figure 1.5: Relationship between Racially Charged Tweets Nationwide and Trump Tweets 31
Figure 1.6: The Effect of the First Local COVID-19 Diagnosis on Racial Animus by the Negative Economic Impact of COVID-19 32

Figure 2.1: Physician SES and Mortality by Patient Education 71
Figure 2.2: Health-SES Gradient by Outcomes of Interest 72
Figure 2.3: The Effect of Physician-patient SES Concordance on Mortality 73
Figure 2.4: The Effect of Physician-patient SES Concordance on Health Behaviors . . 74
Figure 2.5: The Effect of Physician-patient SES concordance on Health Behaviors Related to Chronic Conditions 75

Figure 3.1: Validity of RD 103
Figure 3.2: Displayed Reputation Summary Statistics 104
Figure 3.3: Example Employer Page and Job Posting 105
Figure 3.4: The Effect of Displayed Reputation on Job Seeker Behavior 106
Figure 3.5: The Effect of Displayed Reputation on Employer Behavior 107
Figure 3.6: Heterogeneous Effect of Displayed Reputation by Stickiness 108
Figure 3.7: Heterogeneous Effect of Displayed Reputation by Firm Type 109
Figure 3.8: Heterogeneous Effect of Displayed Reputation by Firm Size 110
Figure 3.9: Heterogeneous Effect of Displayed Reputation by Cumulative Evaluations 111

Figure A.1: Location of Media Markets and Counties with Data 125
Figure A.2: Timeline of COVID-19 Developments and Evolution of Racial Animus . . 126
Figure A.3: The Effect of the First Local COVID-19 Diagnosis on Racial Animus *Rescaled* Google Search Index 127
Figure A.4: The Effect of the First Local COVID-19 Diagnosis on *Daily* Racial Animus 128
Figure A.5: Self-Reported Hate Incidents in the United States 129
Figure A.6: The Effect of the First Local COVID-19 Diagnosis on Racial Animus against *Non-Asian* Minorities 130
Figure A.7: The Effect of the First Local COVID-19 Diagnosis on Racial Animus Placebo Test 131
Figure A.8: Number of President Trump's Tweets about China or COVID-19 132

Figure B.1: Summary Statistics on the Total Population and Physicians (PCP) by Birth Cohort 143
Figure B.2: The correlations between Physician-patient SES Concordance and Mortality 144

LIST OF TABLES

Table 1.1: Relationship between Racial Animus, Hate Crimes, and Chinese Restaurant Visits . . . 33
Table 1.2: The Effect of the First Local COVID-19 Diagnosis on Racial Animus Google Search Index . . . 34
Table 1.3: The Effect of the First Local COVID-19 Diagnosis on Racial Animus Twitter Post Index . . . 35
Table 1.4: Predictors of Tweeting Ch-word among First-Time Ch-word Users after the First Local COVID-19 Diagnosis . . . 36
Table 1.5: Relationship between Trump Tweets and Racial Animus Nationwide . . . 37

Table 2.1: Summary Statistics - Clinics . . . 76
Table 2.2: Summary Statistics - Physicians and Clinics . . . 77
Table 2.3: Summary Statistics - Patients . . . 78
Table 2.4: Test for Selection in Patient-Physician Reassignment . . . 79
Table 2.5: The Effect of Physician(PCP)-patient SES Concordance on Mortality . . . 80
Table 2.6: The Effect of Physician(PCP)-patient SES Concordance on Mortality Caused by Chronic Conditions . . . 81
Table 2.7: The Effect of Physician(PCP)-patient SES Concordance on Healthcare Utilization . . . 82
Table 2.8: The Effect of Physician-patient SES Concordance on Health Behaviors Related to Chronic Conditions . . . 83

Table 3.1: Summary Statistics - Supply Side . . . 112
Table 3.2: Summary Statistics - Demand Side . . . 113
Table 3.3: Validity of RD - Discontinuity of Firm Characteristics . . . 114
Table 3.4: The Effect of Displayed Reputation on Firm Behavior . . . 115
Table 3.5: The Effect of Displayed Reputation on Job Seeker Behavior . . . 116
Table 3.6: The Effect of Displayed Reputation by Firm Type . . . 117
Table 3.7: The Effect of Displayed Reputation by Firm Size . . . 118
Table 3.8: The Effect of Displayed Reputation by Stickiness . . . 119
Table 3.9: The Effect of Changes in Displayed Reputation . . . 120
Table 3.10: Job-level Effect of Displayed Reputation . . . 121
Table 3.11: RDD with Alternative Bandwidths . . . 122

Table A.1: Sample Selection - Media Markets and Counties with Google and Twitter data 133
Table A.2: Summary Statistics . . . 134
Table A.3: Timing of the First Local COVID-19 Diagnosis - Weeks from Jan 19, 2020 . 135
Table A.4: Number of Media Markets and Counties by Week of the First COVID-19 Local Diagnoses . . . 136
Table A.5: Characteristics of First-Time Ch-word Users and Control Users . . . 137
Table A.6: Predictors of First-time Ch-word Users - Twitter Activity . . . 138

Table A.7: Predictors of First-time Ch-word Users - User Profile Keywords 139
Table A.8: Examples of President Trump's Tweets about China or COVID-19 140
Table A.9: Relationship between Racial Animus Nationwide and Tweets from Politicians and National News Outlets 141

Table B.1: ICD-10 Codes Used to Identify Cause of Death 144
Table B.2: ATC Codes Used to Identify Health Behaviors Related to Chronic Conditions 145
Table B.3: Test for Selection in Patient-Physician Reassignment by patient characteristics 146
Table B.4: Test for Selection in Patient-Physician Reassignment by patient pre-closure conditions 147
Table B.5: The Effect of Having a Low-SES Physician (PCP) for High-SES patients by event time 148
Table B.6: The Effect of Having a Low-SES Physician (PCP) for Low-SES patients by event time 149
Table B.7: The Effect of Physician-patient SES Concordance on Mortality by Population Demographics 150
Table B.8: The Effect of Physician(PCP)-patient SES Concordance on Mortality Caused by Chronic Conditions and Gender 151
Table B.9: The Effect of Physician(PCP)-patient SES Concordance on Mortality Caused by Chronic Conditions by Birth Cohort 152
Table B.10: The Effect of Physician-patient SES Concordance on Cancer Mortality . . . 153
Table B.11: The Effect of Physician(PCP)-patient SES Concordance on Primary Care Reimbursement 154
Table B.12: The Effect of Physician-patient SES Concordance on Healthcare Utilization by Gender 155
Table B.13: The Effect of Physician-patient SES Concordance on Healthcare Utilization by Birth Cohort 156
Table B.14: The Effect of Physician-patient SES Concordance on Primary Care Reimbursement by Gender 157
Table B.15: The Effect of Physician-patient SES Concordance on Primary Care Reimbursement by Birth Cohort 158
Table B.16: The Effect of Physician-patient SES Concordance on Health Behaviors Related to Chronic Conditions by Gender 159
Table B.17: The Effect of Physician-patient SES Concordance on Health Behaviors Related to Chronic Conditions by Birth Cohort 160
Table B.18: Robustness Check: Adherence (A) vs. detection (D) effects 161
Table B.19: Internal Validity: The Role of Other Physician Characteristics in Reducing the SES-gradient in Mortality 162
Table B.20: Internal Validity: The Role of Other Physician Characteristics in Reducing the SES-gradient in Mortality 163
Table B.21: Robustness Check: The Effect of Physician Academic Performance 164
Table B.22: Mechanisms: The Effect of Physicians' Parents' Illness on The SES Gradient in Mortality 165

Table B.23: Robustness Check: Define Physician SES Using Parents' Educational Rank 166
Table B.24: Robustness Check: Define SES Using Other Education Levels 167
Table B.25: Robustness Check: Restricted Choice Sample 168
Table B.26: Robustness check: Aggregating Physician SES to the Clinic Level 169
Table B.27: Robustness check: Subsample of Physicians with Non-missing SES 170
Table B.28: Robustness Check: Excluding Non-ethnic Danish patients 171

ACKNOWLEDGEMENTS

Chapter 1, in full, is a reprint of the material as it appears in Journal of Economic Behavior and Organization, forthcoming. Runjing Lu; Sophie Yanying Sheng., Elsevier, 2022. The dissertation author was a primary investigator and author of this paper.

Chapter 2, in full, is currently being prepared for submission for publication of the material. Ida Lykke Kristiansen; Yanying Sheng. The dissertation author was a primary investigator and author of this paper.

Chapter 3, in full, is currently being prepared for submission for publication of the material. Ke Ma; Yanying Sheng. The dissertation author was a primary investigator and author of this paper.

VITA

2016	B. S. in Mathematics, Economics *summa cum laude*, Bates College
2016-2017	Analyst, Analysis Group
2017-2022	Graduate Teaching Assistant, University of California, San Diego
2022	Ph. D. in Economics, University of California San Diego
2022-	Economist, LinkedIn

PUBLICATIONS

Urvi Desai, Noam Y Kirson, Nori Mehta, Jody Wen, **Yanying Sheng**, Wenyu Ye, Jeffrey Scott Andrews, "Trends in Health Service Use and Potentially Avoidable Hospitalizations Prior to Alzheimer's Disease Diagnosis: A Matched, Retrospective Study of US Medicare Beneficiaries", *Alzheimer's Dementia*, 13, 2017

ABSTRACT OF THE DISSERTATION

Essays in Applied Microeconomics

by

Yanying Sheng

Doctor of Philosophy in Economics

University of California San Diego, 2022

Professor Julie Cullen, Chair

This dissertation contains three essays studying topics in applied microeconomics. The first chapter studies the formation and the spread of crisis-driven racial animus during the coronavirus pandemic. Exploiting plausibly exogenous variation in the timing of the first COVID-19 diagnosis across US areas, we find that the first local case leads to an immediate increase in local anti-Asian animus, as measured by Google searches and Twitter posts that include a commonly used derogatory racial epithet. This rise in animus specifically targets Asians and mainly comes from users who use the epithet for the first time. These first-time ch-word users are more likely to have expressed animosity against non-Asian minorities in the past, and their interaction with other anti-Asian individuals predicts the timing of their first ch-word tweets. Moreover,

online animosity and offline hate incidents against Asians both increase with the salience of the connection between China and COVID-19; while the increase in racial animus is not associated with the local economic impact of the pandemic. Finally, the pandemic-driven racial animus we documented may persist beyond the duration of the pandemic, as most racist tweets do not explicitly mention the virus.

The second chapter investigate if primary care physician (physician henceforth) and patient concordance in terms of socio-economic status (SES) reduce the SES inequality in health. We exploit variations in SES concordance between physicians and patients that are induced by plausibly exogenous clinic closures. We find that SES concordance lowers low-SES patients' mortality while high-SES patients' mortality does *not* depend on their physicians' SES. Together, these effects translate to a 23% reduction in the SES-mortality gradient. Mortality reductions related to cardiovascular conditions are especially pronounced. We study patients' health behavior and physicians' treatment choices to explain how SES concordance reduces patient mortality. Low-SES patients with low-SES physicians receive more care at the intensive margin; making more office visits per year and receiving more services per visit. In addition, they are more likely to be prescribed Statins, adhere to diabetes check-up visits, and are less likely to have avoidable hospitalizations due to COPD, relative to comparison groups.

The third chapter asks: how does employer reputation affect the online labor market? We investigate this question using a novel dataset combining reviews from Glassdoor.com and job applications data from Dice.com. Labor market institutions such as Glassdoor.com crowd-sources information about employers to alleviate information problems faced by workers when choosing an employer. Raw crowd-sourced employer ratings are rounded when displayed to job seekers. By exploiting the rounding threshold, we identify the causal impact of Glassdoor ratings using a regression discontinuity framework. We document effects from both labor demand and supply sides at equilibrium. We find that displayed employer reputation affects employer's ability to attract workers, especially when the displayed rating is sticky. Employers respond to the rounding threshold by posting more new positions and re-activating more job postings. The effects are the strongest for firms that are private, smaller, and less established, suggesting that online reputation is a substitute for other types of reputation.

Chapter 1

How Racial Animus Forms and Spreads: Evidence from the Coronavirus Pandemic

1.1 Introduction

Racial animus can affect welfare in measurable ways, as economists have noted since the seminal work of (10). Recent papers have shown that racial animus can hinder economic development, affect political institutions, and induce social unrest (e.g., 16, 14, 91, 47). To curb racial animus at the outset and to mitigate its consequences, a crucial first step is to understand how it forms and grows.

In this paper, we shed light on what factors motivate racial animus, which individuals are more susceptible to such factors, and how racial animus spreads, using the coronavirus (COVID-19) pandemic as a natural experiment. The Centers for Disease Control and Prevention (CDC) has emphasized that people of Asian descent are at no greater risk of spreading the virus than other Americans. Nonetheless, since the outbreak of the virus, news reports of hate crimes against Asian Americans have increased (66).[1] The unexpected nature and regional variation of

[1]For example, see https://www.nbcnews.com/news/asian-america/video-shows-passenger-defending-asian-woman-facing-racism-new-york-n1162296NBC News, https://nyti.ms/3ccvHzONew York Times, and

the pandemic provide a valuable opportunity to study the rise and spread of racial animus – in this case, against Asians.

To proxy for an area's racial animus against Asians, we use the percentage of Google searches and Twitter posts (tweets) that include the words "chink" or "chinks" (hereafter, the ch-word).[2] Google searches can capture *private* racial animus given others cannot view one's searches. Past papers have documented a clear relationship between Google searches of racial slurs and racial animus against minorities (2, 91). Furthermore, as we will show below, an area's monthly Google searches for the epithet is positively correlated with monthly anti-Asian hate crimes and is negatively correlated with monthly visits to Chinese restaurants. Our second proxy is based on tweets, which has been used to measure *public* displays of racial animus (67). These two proxies are valuable alternatives to more traditional measures, such as offline hate crimes which may only capture the most extreme hatred and may not fully reflect the levels of racial animus due to blanket stay-at-home orders during the pandemic. In addition, use of racial slurs online is an important outcome in and of itself, as researchers have shown a strong relationship between exposure to racial discrimination online and depression and anxiety measured offline (101).

To motivate, we exploit the timeline of COVID-19 developments in the United States to understand the general evolution of anti-Asian animus during the pandemic. We find little increase in the national racial animus upon the first US COVID-19 case and only a small uptick in the week when the World Health Organization (WHO) declared COVID-19 a pandemic. In contrast, we observe a clear jump in the week when President Trump tweeted "Chinese virus."

In order to causally identify the effects of COVID-19 on racial animus against Asians, we use a difference-in-differences (DID) event study design exploiting the variation in the timing of

---

https://www.usatoday.com/story/news/politics/2020/05/20/coronavirus-hate-crimes-against-asian-americans-continue-rise/5212123002/USA Today.

[2]We focus on the use of the ch-word because it is the most salient and unambiguously pejorative racial slur against Asians. According to the Philadelphia Bar Association, the epithet "is now widely used throughout the United States as a racial slur against people of Asian descent" (6). Importantly, it has not been reclaimed by the Asian American community (3).

the *first local* COVID-19 diagnosis across areas. Specifically, we compare the change in racial animus following the first diagnosis in an area to the change in other areas during the same period. First local diagnoses are likely to increase the salience of the virus, and the salience of diseases has been shown to induce xenophobia in lab experiments (40). The identifying assumption is that the *precise* timing of the first diagnosis in an area is plausibly exogenous; whether an area has its first diagnosis this week (day) or the next is largely unpredictable and unlikely to correlate with other factors that simultaneously change local racial animus.[3]

Our DID event study reveals that, in the week after the first local COVID-19 diagnosis, an area's Google search rate of the ch-word increases by 22.6 percent of the area's maximum search rate during the sample period, and an area's Twitter post rate of the epithet increases by 118.6 percent of the average post rate across all areas during the sample period. These effects persist for six weeks after the first local case. Given the correlation based on historical data, the increase in Google search rate of the ch-word would be associated with a 6.5 percent increase in anti-Asian hate crimes holding everything else constant. The results, where applicable, are quantitatively unchanged under a *dynamic* event study design which allows for varying treatment effects across event periods (93). Our results are also robust to using alternative racial animus measures based on tweets which include other anti-Asian slurs and are not counter-hate; to excluding early- and hard-hit states; and to controlling for severity of local infection, existence of stay-at-home orders, general local attention to Asians, and area and year-month fixed effects.

When we examine the content of ch-word tweets, we find that the share showing emotions of anger and disgust increases from 23.3 to 40.8 percent after the first local diagnosis. This shift in sentiment suggests that the increase in racially charged tweets represents a real change in attitude towards Asians. Moreover, the increase in racial animus is directed *only* at Asians and not at other minority groups. The singling out of Asians implies that the increase is likely not due to an overall rise of ethnic distrust or tensions from general uncertainty about

[3]Papers like (37) have noted that areas with larger population sizes or better medical systems tend to have first diagnoses earlier. We include area fixed effects to control for these time-invariant characteristics.

cross-group differences in health status or risk-taking behavior. Rather, it is targeted at a specific group associated with the geographical origin of the virus. In addition, 75 percent of ch-word tweets posted following the first local case do not explicitly mention COVID-19, implying that the pandemic-induced racial animus towards Asians extends to broader topics and may persist beyond the duration of the pandemic.

We also leverage the rich information in historical tweets and Twitter user network to study *which individuals* are more likely to start expressing hate because of the pandemic. We find that the surge in ch-word tweets is driven primarily by the extensive margin (i.e., existing Twitter users who post the term for the first time) rather than the intensive margin (i.e., increase in tweets from users who have previously used the term). These first-time ch-word users are 40 percent of the mean more likely than never users to have tweeted racial slurs against non-Asian minorities in the past, implying that the pandemic may have redirected their anti-minority sentiments towards Asians. They are also 58 and 28 percent of the mean more likely to list "Trump" and "politics" in their user profiles.

Finally, we turn our attention to the factors fueling the spread of racial animus among individuals. Exposure to anti-Asian users is one such factor. We find that interacting with anti-Asian users in a day predicts a 22 percent higher likelihood (relative to the mean) of tweeting the ch-word the next day. The salience of the connection between COVID-19 and the Asian population is another factor. We proxy for this salience by using the number of President Trump's tweets that mention China and COVID-19 simultaneously. We find that one additional such China-and-COVID tweet in a day corresponds to an eight percent increase in anti-Asian hate incidents and an increase in national ch-word tweets on the same day, equivalent to 14 percent of the daily average. An event study using *hourly* tweet data also reveals an immediate increase in ch-word tweets following the president's China-and-COVID tweets but not before. In contrast, we find little evidence that negative economic impacts from the pandemic motivates the initial rise of racial animus. Areas with a more severe economic damage from the pandemic do not exhibit a higher increase in racial animus than areas with a less severe impact.

This paper contributes to the literature studying the causes of animus toward minorities. This body of work has shown that negative shocks such as terrorist attacks and deterioration of economic conditions induce animus against racial or religious minorities. For example, (53), (45), and (50) document that 9/11 and jihadi terror attacks lead to increases in anti-Muslim hate crimes. (4) and (2) find that the Great Recession and negative shocks to agricultural income in historical Europe contribute to animus against minorities. In addition, desire to avoid health threats has also been postulated to motivate racial bias (82). Lab experiments have shown that exposing subjects to disease-related primes leads to increased xenophobia (40, 73, 9). However, the causal evidence on whether infectious diseases lead to racial animus in the field is still lacking.[4] An exception is (51), which documents that the black death caused an increase in anti-Jewish pogroms in medieval Europe.

Our contribution is to provide causal evidence on how negative shocks, such as pandemics, trigger racial animus, and shed light on who are more susceptible to such shocks and how racial animus spreads. Our findings have implications for mitigating animus amid future crises. We find that the rise in racial animus is specific to Asians who are associated with the geographical origin of the virus and that the salience of this association amplifies animus against the group. Therefore, careful naming of a disease (e.g., COVID-19 and Delta variant as opposed to Chinese virus and Indian variant) and debunking claims of a purported connection between a disease and a group could be helpful in curbing animus. Additionally, our findings reveal that the extensive margin and social media play an important role in spreading racial animus, suggesting that moderating racist individuals and their interaction with others on social media could help constrain racial animus in the future.

Finally, our paper speaks to the literature on political rhetoric. Political rhetoric has been shown to influence public opinion and behavior, such as presidential approval (36), public

[4]More recent papers on the prevalence of hate during the COVID-19 pandemic are mostly descriptive (e.g., 83, 106, 60, 20) or take a structural approach (e.g., (33)).

perception of a foreign country (86), and anti-minority hate crimes (65). We add to this literature by providing another example of how the rhetoric of political figures regarding a public crisis influences racial animus at the national level. On the flip side, harnessing these public figures' opinion-shaping power could be useful in curbing animus.

1.2 Measures of Racial Animus

1.2.1 Google and Twitter Proxies

We use two measures to proxy for an area's racial animus against Asians: the percentage of Google searches and the percentage of tweets that include the words "chink" or "chinks." The ch-word is not uncommon in Google searches or tweets. Between June 2019 and June 2020, this racial epithet was included in more than a quarter million searches and 60,000 tweets.[5] Google searches and tweets that include the epithet are mostly negative. For instance, "chink eye" and "chink virus" are common terms in such Google queries and Twitter posts. People may search the epithet to look for jokes or memes about Asians or to look for like-minded others with whom they can share anti-Asian sentiments.

We use Google Trends to obtain weekly Google search data for the ch-word at the media market level between July 2019 and April 2020. The data are not the raw number of searches but the weekly percentage of searches that include the term (*search rate*), taken from a random sample of total searches representative at the media market and time levels and scaled by the highest weekly search rate in the same market during the entire extraction period – in our case, between July 2019 and April 2020. In particular, the racially charged Google search index for

[5] The number of Google searches is an approximation from https://searchvolume.io/. The data are only available for the 12-month period before our query on June 8, 2020.

media market m at time t extracted over period T is

$$Googlesearchindex_{mt,T} = 100 \times \frac{\frac{Searchesincluding``chink(s)''_{mt}}{Totalsearches_{mt}}}{\max\limits_{t \in T}\{\frac{Searchesincluding``chink(s)''_{mt}}{Totalsearches_{mt}}\}} \tag{1.1}$$

Note that Google returns a zero value when the racially charged search index for a given area and time falls below an unreported threshold. We thus focus on media markets that have a valid racially charged Google search index in the baseline period (2014-2018). This leaves us with 60 of 210 media markets, covering approximately 74 percent of the US population and 78 percent of the US GDP in 2019 across 33 states. Compared to other media markets, the ones in our sample tend to have a larger population, higher percentage of Asians, slightly lower baseline anti-Asian hate crime rate, and more enplanements of international airports (Table A.1 column (1)). Shaded areas in Figure A.1 panel A indicate the media markets in our sample.

The above metric can capture the timing but not the level of a change in an area's search index. As an alternative, we rescale the Google search index so that the search rate in different media markets is normalized by *one* base search rate. We try three different bases: Huntsville-Decatur (Florence)'s search rate on March 15, 2020; Wilkes Barre-Scranton on March 29, 2020; and Buffalo on April 5, 2020. We choose these bases to obtain rescaled indexes for as many media markets as possible, i.e., 35, 29, and 29, respectively. As detailed in A.1, rescaling drops many media markets whose search rate is zero on the date when the base search rate occurs (*benchmark date*). For this reason, we only use the rescaled version as a robustness check.

We obtain Twitter data from Crimson Hexagon, which houses all public tweets through a direct partnership with Twitter. We download all geo-located tweets that include the ch-word between November 1, 2019, and May 2, 2020. Crimson Hexagon does not provide the total number of tweets posted in a given area and time. We thus extract the number of all public tweets that include the word "the," the most common word on Twitter, in a given area and time as a substitute. Assuming that the proportion of tweets that include "the" is stable across areas, the number of tweets that include "the" can approximate Twitter activity. We define the racially charged Twitter post index for a given area and time as the number of tweets including the ch-word per 100,000 tweets including the word "the."

We calculate the Twitter post index for 658 counties across 50 states and Washington D.C., encompassing 60 percent of the US population and nearly 70 percent of the US GDP in 2019. Counties are included if their residents ever posted "the" tweets between 2014 and 2018.[6] Counties with Twitter data tend to have a larger population, higher support for the Democratic Party, and higher enplanements of international airports, but show no difference in baseline anti-Asian hate crime rate compared to other counties (Table A.1 column (2)). Shaded areas in Figure A.1 panel B are counties with Twitter data. Analyses using Twitter data are conducted at the county level unless noted otherwise.

The fact that Google and Twitter data do not cover the full of the US should not affect internal validity of our study, but it could pose a threat to external validity. Therefore, we use *both* data sources, which could alleviate concerns about the external validity of our findings.

1.2.2 Relationship between Racial Animus, Hate Crimes, and Restaurant Visits

For the racially charged Google search index and Twitter post index to be meaningful proxies for racial animus, the only assumption we need is that an increase in racial animus makes a person more likely to use the ch-word. Under this assumption, higher racial animus results in a higher percentage of Google searches and tweets that include the racial epithet. Existing papers that use a similar proxy for racial animus suggest that the assumption is likely to hold (2, 34, 91).

To better understand the above proxies, we check how they predict anti-Asian hate crimes and visits to Chinese restaurants. Hate crime data come from the FBI Uniform Crime Reports (UCR) and are available up to 2018. A majority of these hate crimes are simple or aggravated

[6] About half of the tweets in the sample lack geo-identifiers and hence cannot be associated with a certain county.

assault (30 percent) and in-person intimidation (34 percent). Table 1.1, panel A, columns (1) through (4) report the media market-level correlation between the monthly Google search index and the monthly number of anti-Asian hate crimes between January 2014 and December 2018, controlling for local population size, unemployment rate, year-month fixed effects, and media market fixed effects. On average, a one-standard-deviation increase in the Google search index corresponds to an increase in the anti-Asian hate crimes in the same month, amounting to 8.9 percent of the monthly average.[7] The correlation is robust to controlling for the Google search index for "Asian(s)," which is related to the ch-word but neutral in connotation, as shown in column (2). In columns (3) and (4), we include both the index in the current month and the index in the prior month. The relationship between the Google search index and hate crimes is mainly contemporaneous.

Next, we change the dependent variable to monthly visits to Chinese restaurants in each media market between January 2018 and December 2019, additionally controlling for the monthly visits to all local restaurants. The visit data are from Safegraph and are available starting in 2018.[8] Table 1.1, panel A, columns (5) and (6) show that a one-standard-deviation increase in the Google search index is linked to 484 fewer monthly visits to Chinese restaurants, equaling 0.5 percent of the monthly average. The relationship between the Google search index and visit rate is also contemporaneous.

Finally, we replicate the above correlations using Twitter data in Table 1.1, panel B. We aggregate hate crimes to the media market level due to their low occurrences at the county level. To maintain consistency, we also aggregate restaurant visits to the media market level. Overall, the Twitter post index does not correlate with anti-Asian hate crimes or visits to Chinese restaurants. One potential explanation is that Twitter data represent public displays of racial animus and undergo more social censoring. We may only see a change on Twitter when the shift

[7] The percent increase is calculated by multiplying the standard deviation of the index (23.07) with the coefficient and dividing the product with the outcome mean.

[8] Safegraph provides data on foot traffic to roughly 4.1 million points of interest in the United States.

in racial animus is substantially large.

1.3 Evolution of Racial Animus in US amid the Pandemic

To motivate, we study the general evolution of anti-Asian animus as the pandemic develops. An ideal experiment would be to contrast rates of racially charged Twitter posts and Google searches in the U.S. during the pandemic to counterfactual rates absent the pandemic. However, a perfect counterfactual does not exist because all individuals and areas were more or less impacted by the pandemic. For this reason, we use racially charged Twitter posts and Google searches in 2019 as controls. The assumption is that racially charged Twitter posts and Google searches in 2020 would have been the same as in 2019 absent the pandemic.[9]

We first compare an individual's weekly likelihood of tweeting the ch-word during the first 16 full weeks in 2020 and the *same* person's likelihood of doing so in the corresponding weeks in 2019. An advantage of this analysis is that it does not require geo-identifiers, so we can include all 26,065 Twitter users who ever tweeted the ch-word between 2014 and 2018.[10] We use the following specification:

$$Y_{iyw} = \sum_{w=2}^{16} \beta_w \times 1\{y = 2020\} + \alpha_i + \alpha_w + \varepsilon_{iyw} \tag{1.2}$$

where Y_{iyw} is a binary variable which equals one if individual i tweets the ch-word in week w of year y. We use $w = 1$, the first full week of a year, as the comparison period. Our treatment variable is $1\{y = 2020\}$, which equals one if the year is 2020, and 0 if the year is 2019. We include person fixed effects α_i and week-of-year fixed effects α_w to absorb individuals' baseline propensity to tweet the racial epithet and the seasonality in such tweets. We cluster standard

[9]This assumption could be violated if there are other contemporaneous shocks affecting racial animus. Our strategy in the next section avoids this issue.

[10]We cannot look at the universe of Twitter users because Crimson Hexagon only allows tweet extraction based on keywords.

errors by individual.

The individual-level analysis reveals that the likelihood of tweeting the ch-word co-moves with important developments of COVID-19. In Figure A.2 panel A, we plot β_w from equation 1.2. While we find little to no increase in the likelihood of tweeting the term following the first US COVID-19 case or declarations of health emergency and only a small uptick in the week when the WHO declared COVID-19 a pandemic, we observe a clear jump in the week when President Trump first tweeted “Chinese virus.”

Lacking individual-level search data, we compare media market-level weekly racially charged Google search index in 2020 and the index in the corresponding markets and weeks in 2019. Specification is the same as equation 1.2, except that Y_{iyw} is now the Google search index in media market i in week w of year y. We plot β_w in Figure A.2 panel B. While we also see a spike in the Google search index in the week when President Trump tweeted “Chinese virus,” we cannot draw definitive conclusions for other weeks.

1.4 Evidence from DID Event Study

We now turn to our main empirical strategy, a DID event study design exploiting the variation in the precise timing of the first COVID-19 diagnoses across the United States. We compare the changes in racially charged Google search index (Twitter post index) in the weeks before and after the first local case to the changes in other media markets (counties) during the same period. This design allows us to avoid concerns about contemporaneous shocks that influence racial animus at the same time as the pandemic develops.

1.4.1 Data and Empirical Strategy

We download the data on US COVID-19 cases and deaths between January 21 and May 2, 2020, from the Johns Hopkins University Coronavirus Resource Center. We match the date of the first case in each media market and county to those with valid Google and Twitter data. Table A.4 displays the number of media markets and counties by the timing of their first local diagnoses. All media markets have their first diagnoses in the sample period and have Google data for at least six weeks after the diagnosis. These media markets make up the Google sample. Seventeen counties with Twitter data are excluded because they did not have diagnoses in the sample period.[11] The remaining 641 counties make up the Twitter sample and have data for at least one week after the first local diagnosis; the number of counties decreases to 636, 629, 613, 555, 416 in weeks two to six.[12] Therefore, the Google (Twitter) sample is a panel of media markets (counties) from six weeks before to six weeks after the first local diagnosis. Table A.2 reports summary statistics for each of the samples.

To understand predictors of the diagnosis timing, we regress the week of first local diagnosis on a battery of local characteristics in Table A.3. The analysis reveals that a larger population size predicts earlier diagnoses for both the Google and the Twitter samples, while enplanements of international airports predict slightly later diagnoses for the Google sample. However, the proportion of Asians does not have predictive power for the timing, consistent with the CDC's statement that Asians are at no greater risk of spreading the virus. More importantly, pre-pandemic anti-Asian hate crime rate does not predict the timing, suggesting that the treatment timing is orthogonal to baseline racial animus.

We then estimate the following regression:

$$Y_{it} = \sum_{t=-6, t\neq -1}^{6} \beta_t + \gamma' X_{it} + \alpha_i + \alpha_{ym(t)} + \varepsilon_{it} \tag{1.3}$$

where Y_{it} is the racially charged Google search index (Twitter post index) in media market (county) i in event time t, which is the number of time periods relative to the first local diagnosis. β_t represents event dummies for six weeks before to six weeks after the first local

[11] Results are unaffected when we include these counties.

[12] Crimson Hexagon was discontinued in July, 2020, so we cannot extend Twitter sample.

diagnosis, excluding our comparison period $t = -1$. X_{it} is a vector of area-specific time-varying characteristics such as the local number of COVID-19 diagnoses or deaths, an indicator for a state-level stay-at-home order, and the Google search index or Twitter post index for "Asian(s)". We include county or media market fixed effects α_i and year-month fixed effects $\alpha_{ym(t)}$ to control for an area's baseline racial animus and national trends in racial animus.[13] We cluster standard errors by media market for Google data and by county for Twitter data. We also estimate equation 3.3 at the daily level, where we include event dummies from 14 days before to 21 days after the first local diagnosis while omitting the dummy for day -4 and additionally control for day-of-week fixed effects.

If the trends of racially charged Google search index (or Twitter post index) across media markets (or counties) are parallel in the absence of local COVID-19 cases, and the treatment effect of the first local case does not vary across event times, β_t identifies the weighted average treatment effect across treatment areas in time t on local searches or posts of the ch-word. Testing for parallel pre-trends can shed light on the first identifying assumption. As we will show, this assumption appears to hold. The second assumption is harder to test, and its violation could bias the estimates in unknown directions. For example, if earlier treated areas experience an increasing (or decreasing) treatment effect over time due to evolving local pandemic situations, using these areas as controls for later treated places could bias the average treatment effect downward (upward).

To alleviate concerns about time-varying treatment effects, we use a *dynamic* DID event study comparing areas with a first case before and after the case, using areas that *have not had* any cases as controls. To implement the dynamic event study, we follow (68) and stack our data as a series of 2×2 matrices (treated/not-yet-treated × pre/post). We define areas which have their first cases in calendar week g as cohort g, and cohort-specific event time in calendar month

[13] Although the Google search index is a normalized search rate so that the maximum search rate in a media market is equivalent to a search index of 100, there is still considerable variation in the sample mean of this normalized index varying between eight and 50. For example, 50 means that the average search rate in a media market is half of its maximum search rate during the extraction period.

m as $e_g = m - g$. The treatment effect on cohort g in event time e_g is labeled as β_{e_g}. Following (93), we define the average treatment effect for event time e among all cohorts G as:

$$\beta_e = \sum_{g \in G} \beta_{e_g} \times w_g \tag{1.4}$$

where the aggregation weight w_g is the population in areas belonging to cohort g. We calculate clustered standard errors at the area level for β_e via the delta method.

A limitation of the dynamic event study is that it requires enough not-yet-treated areas in event time e_g to estimate β_{e_g}. Since counties are smaller than media markets, there is more variation in the timing of first local diagnoses at the county level. As a result, most counties in the Twitter sample have control counties for multiple post periods while most media markets in the Google sample have none after event 0. Therefore, we only apply the dynamic event study to Twitter data and use this approach as a robustness check.

1.4.2 Effects of the First Local Case on Racial Animus

Main Findings

We start by examining how an area's Google searches for the ch-word respond to the first COVID-19 case in the local area. Figure 1.1, panel A plots β_w from equation 3.3 using an area's racially charged Google search index as the outcome. The Google search index jumps markedly in the week after the first local case and persists at high levels in the following weeks. The pre-trends are flat and statistically insignificant, suggesting that the parallel trend assumption is likely to hold. Regression results corresponding to this figure are found in Table 1.2, column (1). For example, consider the $+1w$ coefficient: compared to the week before the first local case, in the first week, an area's racially charged search rate increases by 22.6 percent of the area's maximum search rate over the sample period. The treatment effects remain mostly above 17 percent for the following five weeks. Given our findings of the correlation between the Google search index and hate crimes, the increase in the index in the month after the first local diagnosis translates to an increase of 0.0095 anti-Asian hate crimes or 6.5 percent of the average monthly anti-Asian hate crimes between 2014 and 2018.[14]

Figure A.3 shows the event time plot when we replace the original Google search index with the indexes rescaled using three different bases. The patterns are qualitatively similar to those using the original index, although the magnitude of the estimates is now roughly half the size. This is because base search rates for the rescaled indexes are higher than search rates in most media markets. The standard errors of the estimates also become much larger because rescalling forces us to drop nearly half of the media markets (see A.1 for detail). Because of this, we only present results using the original Google search index in the rest of the paper.

We next turn to Twitter to understand how the first local case affects *public* use of the ch-word. In Figure 1.1, panel B, we plot the effect of the first local case on the racially charged Twitter post index. Similar to the Google search index, the Twitter post index also jumps in the week after the first case. Specifically, relative to the week before the case, the Twitter post index increases by 0.7 per 100,000 "the" tweets in the week after, amounting to 118.6 percent of the weekly average during the sample period. The effects remain high in weeks 2-6. Table 1.3, column (1) reports the regression results.

To confirm that our results are not driven by the functional form of the Twitter post index or the specific racial epithet we choose, we use alternative functional forms and other ways of identifying anti-Asian content. Raw number of ch-word tweets and number of ch-word tweets per million population reveal similar patterns as the original Twitter post index, as shown in columns (2) and (3). Additionally, we construct a new index using COVID-related tweets posted between January 15 and April 17, 2020 that are classified as anti-Asian via machine learning.[15]

[14] We obtain the number by multiplying the average of $+1w$ through $+4w$ coefficients in Table 1.2 with the number of monthly anti-Asian hate crimes per unit increase in the Google search index in Table 1.1.

[15] We thank (106) for providing the data. These anti-Asian tweets include phrases like "Chinese Virus" and "Wuhan Virus" and exclude counter-hate tweets that may have racist keywords in them. Only counties that had their first diagnoses between February 16 and March 22, 2020 are included in this analysis.

Column (4) shows that the effects estimated with this new index share a similar pattern to the ones in column (1) but are seven times as large. The original Twitter post index is thus likely a conservative measure of racial animus.

An evolving local pandemic situation may produce time-varying treatment effects, which could bias results of a regular DID event study. To alleviate this concern, in column (5), we re-estimate the effect using a dynamic DID event study (93). The estimates are quantitatively similar to those in column (1), implying that time-varying treatment effect is likely not an issue here.

Figure A.4 presents results using indexes at the *daily* frequency. Both indexes start to rise two to three days after the first local case, suggesting that residents react to the news of the first local COVID-19 case fairly quickly.

Discussion and Robustness Check

In this subsection we discuss alternative explanations for the rise in the ch-word use in an area after the first local COVID-19 case and explore the robustness of our main findings.

Increased ch-word usage may result from a general rise in online activities due to blanket stay-at-home orders rather than a change in racial animus. However, our search index and post index already account for an overall change in online activities because they are normalized by the total searches and tweets in a given area and time. In addition, when we include an indicator for state-level stay-at-home order in Table 1.2 column (2) and Table 1.3 column (6), results are quantitatively similar to those from our main specification.

Alternatively, an increase in general attention to Chinese or Asians may lead to higher ch-word usage. In Table 1.2 column (3) and Table 1.3 column (7), we control for searches or tweets of terms that capture such general attention but are neutral in connotation, i.e., "Asian(s)." Results are unaffected.

Our results are also robust to excluding early- and hard-hit states like New York, Washington, and California, as shown in Table 1.2 column (4) and Table 1.3 column (8). Our findings thus represent a general phenomenon across the United States rather than only in a few states particularly impacted by the pandemic.

One may also worry that "Twitter bots" rather than actual users are responsible for the rise in ch-word use on Twitter. However, only 10.4 percent of users who post anti-Asian tweets between January 2020 and April 2020 are potential bots (106). Moreover, our results are quantitatively unchanged when we exclude users who are more likely to be bots, i.e., those who tweeted the ch-word more than five times (99 percentile in our sample) during the sample period, in Table 1.3 column (9).

The increase in searches and tweets including the ch-word could also come from the seasonality in ch-word use and may exist absent the pandemic. To test this possibility, we generate a placebo diagnosis date for each area using the same calendar day and month of its actual diagnosis date but changing the year from 2020 to 2019. We reestimate equation 3.3 using the placebo dates and plot the effects in Figure A.7. Reassuringly, the Google search index and Twitter post index do not change around the placebo dates, suggesting that seasonality cannot explain our findings.

Finally, the increase in ch-word use on Twitter could reflect a change in the social cost of publicly expressing racial animus rather than a shift in attitudes towards Asians. However, this would not explain the increase in racist Google searches, which are done in private. Several other pieces of evidence also support a shift in attitudes. First, the proportion of ch-word tweets showing emotions of anger and disgust increases from 23.3 percent between November 2020 and the first local diagnosis to 40.8 percent in the six weeks following the first diagnosis.[16] Second, data on self-reported hate incidents from Asian Pacific Policy and Planning Council (AP3CON) Stop Hate Reporting System show that the daily average of anti-Asian hate incidents nationwide was alarmingly 70 in late March 2020 and 13 between April and May 2020 (Figure A.5). Third,

[16]Crimson Hexagon assigns each tweet emotion tag(s) generated via a natural language processing algorithm. Please refer to https://www.brandwatch.com/blog/understanding-sentiment-analysis for more details.

Pew Research Center's Global Attitudes Survey, conducted in June to August 2020, shows that unfavorable views of China have reached historic high (104). Taken together, the rise in ch-word usage likely represents a real change in animus against Asians and not just a lower cost of publicly expressing it.

1.4.3 What Motivates Racial Animus and Who Responds the Most

Thus far, we have provided evidence that animus against Asians, as measured by Google searches and Twitter posts including the ch-word, surges immediately following the first diagnosis in an area. We next explore *what* motivates individuals to increase animus in response to the pandemic and *who* respond the most.

As a first step, we test whether the rise is specific to Asians. If the racial animus is motivated by an overall increase in ethnic distrust or tensions from general uncertainty about cross-group differences in health status or risk-taking behavior, we expect to see an increase in animus against non-Asian minorities too. By contrast, if the racial animus is targeted *only* at Asians, it is more likely to be motivated by the association between Asians and the geographical origin of the virus.

To proxy for racial animus against other minorities, we construct Google search and Twitter post indexes for common racial epithets against major minority groups in the United States, such as the n-word (both singular and plural) against African Americans, "wetback(s)" against Hispanics, and "kike(s)" against the Jewish population.[17] We estimate equation 3.3 using racially charged searches and tweets against these minorities as outcomes.[18] The coefficients on the event dummies are plotted in Figure A.6. None of the examined racial epithets experience

[17]We do not use "spic(s)" as the epithet against Hispanics because the cleaner brand "Spic and Span" experienced growing interest during the pandemic. We do not include "redskin(s)" because the corresponding queries and tweets are about an American football team formerly called "Washington Redskins".

[18]When using the n-word as the outcome, we include an indicator for the week of January 26, 2020 because the word's use spiked due to an MSNBC anchor using the n-word when broadcasting Kobe Bryant's death. When using the k-word as the outcome, we include an indicator for the week of February 23, 2020 because Los Angeles Dodgers player Enrique ("Kiké") Hernandez led to a spike in the word's use.

an increase in Google searches following the first local diagnosis. A similar pattern is found for tweets using the w-word and the k-word.[19] The lack of response in the use of racial epithets against other minorities suggests that the pandemic-induced racial animus is mainly driven by the connection between Asians and the geographical origin of the virus.

Although the anti-Asian animus is motivated by the potential geographical origin of the virus, racially charged tweets extend to broader topics than just the virus. Figure 1.2 demonstrates that the increase in ch-word tweets mostly comes from those that *do not* explicitly mention COVID-19, i.e., no mention of "virus", "COVID", "pandemic" or "epidemic". As a result, pandemic-induced racial animus may persist beyond the duration of the pandemic.

We next study which individuals are more susceptible to the pandemic shock. We begin by examining whether the increase in ch-word usage comes from users who only start to harbor animus against Asians after the pandemic hits or from existing anti-Asian users who step up their animosity. We define *existing* ch-word users as individuals who tweeted the ch-word at least once between 2014 and the sixth week before the first local COVID-19 diagnosis. We define *first-time* ch-word users as individuals who never tweeted the ch-word between 2014 and the sixth week before the first local diagnosis and who posted at least 10 tweets before their first ch-word tweet. This definition avoids counting newly registered Twitter users as first-time ch-word users.

Figure 1.3 plots the breakdown in effects by the first-time versus existing ch-word user status. The increase in ch-word tweets from first-time users is roughly 4.5 times of that from existing users in the first two weeks after the first local diagnosis. This breakdown suggests that the extensive margin plays an more important role than the intensive margin in driving racial animus during the pandemic. After the first local diagnosis, 4,621 Twitter users started to use

[19] We present the result for tweets using the n-word in Figure A.6 panel C. N-word tweets may not be a valid proxy for racial animus against African Americans on Twitter because of Black Lives Matter protests, Black History Month in February, and seasonality which is evident when comparing the n-word usage between 2019 and 2020 in panel D. Note that we include an indicator for the week of February 9, 2020 in panel A because a viral n-word tweet unrelated to COVID-19 contributed to 95 percent of the n-word tweets on that day.

the racial epithet, potentially exposing their combined 13 million followers to racially charged content and creating a multiplier effect on racial animus.

To better understand the type of individuals whose anti-Asian sentiment is easily influenced by the pandemic, we analyze user profiles and historical tweets of first-time ch-word users.[20] To form a comparison group, we extract the same information for 3,000 randomly selected Twitter users who registered before July 2019 and never tweeted the ch-word by the end of our sample period (hereafter, control users).

Table A.5 reports the summary statistics for first-time ch-word users and control users. Both groups of users are seasoned Twitter users: their average account age is roughly six years, and their average number of followers is well over 1,000. Compared to control users, first-time ch-word users are more likely to tweet racial epithets against other minorities and have interacted with anti-Asian users before the pandemic. They also appear to pay more attention to politics and news, as evident by their much higher interaction with twitter accounts of prominent politicians and major news outlets. Interestingly, very few ch-word and control users ever tweeted COVID-related conspiracies.

To formally characterize users that are more susceptible to the pandemic-driven racial animus, we run two user-level regressions. The first one regresses an indicator for being a first-time ch-word user on the user's pre- and mid-pandemic Twitter activity, and the second one on user profile keywords; both regressions control for account age, log number of followers, and log number of followings. Regression results are plotted in Figure 1.4 and reported in appendix Tables A.6 and A.7. Figure 1.4 panel A presents the relationship between being a first-time ch-word user and user activity on Twitter. Users who interacted with anti-Asian users before the pandemic are over twice the mean more likely than others to tweet the ch-word for the first time upon the pandemic. As we will show in the next section, this interaction plays a key role in spreading animus against Asians. In addition, users who tweeted racial epithets

[20] We downloaded historical tweets and user profiles for 3,033 of these users in August 2021. We cannot download the rest because their accounts are private, suspended, or deactivated.

against non-Asian minorities before the pandemic are 40 percent the mean more likely to be first-time ch-word users. This finding implies that the crisis may have redirected pre-existing anti-minority sentiments towards Asians. Interestingly, paying attention to major politicians and news outlets also predicts a slightly higher chance of being a first-time ch-word user. Finally, tweeting COVID-related conspiracies has a precisely estimated zero effect on tweeting the ch-word, suggesting that such conspiracies is not the main cause of racial animus among users in our sample.

Panel B plots the relationship between being a first-time ch-word user and user profile keywords.[21] Consistent with results in panel A, keywords indicating attention to politics have the largest positive predictive power. Users who list "Trump" and "politics" in their profiles have a 58 and 28 percent higher chance (relative to the mean) of tweeting the ch-word for the first-time after the pandemic shock, respectively. As we will show in the next section, opinions of public figures, such as those of President Trump, likely play a crucial role in inciting anti-Asian sentiment during the pandemic. In contrast, keywords related to profession and family life, such as "artist," "wife," and "husband," predict a significantly lower propensity to tweeting the ch-word upon the pandemic.

1.5 Factors Fueling Racial Animus

In this section, we explore factors that may have helped propagate anti-Asian animus during the pandemic. Understanding these factors is crucial to stopping the spread of animus from the outset amid future crises.

We know from the previous section that first-time ch-word users are the main driving force behind the rise of ch-word usage on Twitter during the pandemic. In Table 1.4, we zoom

[21] For ease of presentation, we only include the 25 most common user profile words used by first-time ch-word users and those by control users. Since there is an overlap between the two sets of words, the number of words included in the regression is less than 50.

in on these users and their Twitter activity between the date of the first local COVID-19 case and the end of sample (May 2, 2020) to understand what prompts their first ch-word tweets. We regress a user's likelihood of tweeting the ch-word in a given day on a series of indicators for whom they interacted with and what they tweeted about in the day before. We control for user characteristics as well as county, year-of-week, and day-of-week fixed effects to absorb the average propensity to tweet the ch-word in a county and the national trend of such tweets.

Exposure to anti-Asian individuals Table 1.4 column (1) shows that interaction with anti-Asian users (i.e. users who have previously used the ch-word) in a given day is associated with a 0.28 percentage point increase in the likelihood of tweeting the epithet in the following day, amounting to 22 percent of the sample mean. This finding highlights the importance of social media in spreading racial animus and is consistent with papers which document how social media influence real outcomes like voting behaviors (e.g., 41). Our finding suggests that moderating racist individuals and their interaction with others on social media could constrain the spread of animus.

Opinions of public figures. The only other positive predictor in column (1) is a user's interaction with President Trump. (Re)tweeting, replying, or mentioning the president in a day is associated with a 0.33 percentage point increase in the likelihood of tweeting the ch-word the next day, or 26 percent of the sample mean. This finding is consistent with (65) which shows that President Trump's tweets affect public behavior such as hate crimes. In contrast, mentioning other prominent politicians of either parties or national news accounts has little to no predictive power, or even predicts a lower likelihood of tweeting the epithet. When we additionally control for the number of new COVID-19 cases or deaths in the local area in column (2), the results remain similar. Taken together, certain public figures play a key role in shaping public opinions of a subject matter. Harnessing their opinion-shaping power could be useful in curbing animus in the future.

Salience of Asian-COVID connection One potential factor mediating the relationship between ch-word use and interaction with President Trump is the salience of the connection between COVID-19 and the Asian population. It is possible that President Trump's tweets that mention COVID-19 and China simultaneously (hereafter, China-and-COVID tweet) may increase the salience of the connection and influence racial animus. We categorize all President Trump's tweets between January 1, 2020 and May 2, 2020 that contain any of the words "china", "chinese", "huawei", "xi", "covid", "covid-19", "corona", "coronavirus", "virus", "epidemic", or "pandemic" into three categories: those mentioning only China (China-only), only COVID-19 (COVID-only), and both China and COVID-19 (China-and-COVID). Table A.8 presents examples of President Trump's tweets. Figure A.8 plots the daily frequency of his tweets.

In Table 1.5, we regress the daily racially charged Twitter post index at the national level on the number of the president's tweets in each of the three categories while controlling for year-week and day-of-week fixed effects. Column (1) shows that one additional China-and-COVID tweet of President Trump in a day corresponds to roughly five more racially charged tweets per million "the" tweets nationwide on the same day. This increase is non-trivial and is equivalent to 14 percent of the national daily average. Importantly, the Twitter post index does not co-move with the president's China-only or COVID-only tweets, highlighting that the *connection* between China and COVID-19 is what matters. Results remain similar when we control for the daily number of new COVID-19 cases and deaths nationwide in column (2).

The time-series correlation may be confounded by contemporaneous shocks unrelated to the president's tweets. To alleviate this concern, we conduct an event study comparing nationwide racially charged Twitter post index in the *hours* before and after President Trump's China-and-COVID tweets, using the index during the same hours-of-day on days without such tweets as controls. Figure 1.5 shows that the index in the four hours leading up to the China-and-COVID tweets is no different from other times, but it jumps in the first hour after such tweets and continues to grow. The immediacy of the change upon the president's tweets suggests a causal interpretation of the relationship between the salience of the China-and-COVID connection and the anti-Asian sentiment at the national level.

In addition, we study whether the salience of the connection has translated into hate incidents against Asians. We obtain self-reported anti-Asian hate incidents from AP3CON Stop AAPI Hate Reporting Center, a hate incident self-reporting website that went online on March 17, 2020. This is the best hate-tracking organization specialized in anti-Asian hate incidents in the United States https://www.cbsnews.com/news/coronavirus-pandemic-anti-asian-hate-crimes-tracking/(CBS News, 2020). In Table 1.5, column (3), we regress the log of daily hate incidents at the national level on the number of the president's tweets in each of the aforementioned categories while controlling for year-week and day-of-week fixed effects. We find that one additional China-and-COVID tweet from the president in a day corresponds to a roughly eight percent increase in self-reported hate incidents against Asians nationwide on the same day.[22] When we control for the daily number of new COVID-19 cases and deaths nationwide in column (4), results are unchanged.[23]

In contrast to the clear relationship between anti-Asian sentiments and the president's tweets, we find little evidence that the sentiment co-moves with tweets from other prominent politicians or national news outlets (Table A.9). The difference is likely due to the large difference in the number of Twitter followers between the president and the others. President Trump amassed 88.7 million followers before Twitter suspended his account in January 2021, while the follower number as of October, 2021 for the prominent politicians and national news outlets are mostly below 10 millions with only Fox and CNN reaching 20.2 and 54.7 millions, respectively.

Economic downturn The COVID-19 pandemic poses risks on both lives and livelihoods. Existing work has documented that a deterioration of economic conditions can fuel animus towards

[22] We conduct the analysis at the daily level because the exact hour of the incidents is not available. We cannot estimate equation 3.3 with AP3CON data due to the lack of pre-periods given the late start date of the data.

[23] We find little relationship between the racially charged Google search index and President Trump's tweets. Results are available upon request.

minorities (4, 2, 85). We thus study the heterogeneity in the change in racial animus by the level of the pandemic's negative impact on the local economy. We partition the main regression samples by whether the proportion of an area's annual average employment in "leisure and hospitality" and "education and health services," the two hardest-hit industries in employment according to the Bureau of Labor Statistics (BLS), is above or below the sample median (32 percent in Google data and 35 percent in Twitter data). We also partition the samples by whether the percent change in net revenue between January and March, 2020 among local small businesses is above or below the sample median (-39 percent in the Google sample and -37 percent in the Twitter sample) using data built by (18). Figure 1.6 shows that the areas that experience high versus low negative economic impact respond similarly to the first local COVID-19 diagnosis. In other words, the negative economic impact of the disease appears to play a relatively weaker role in motivating the initial rise of racial animus. One potential reason is that the long-term impact was not well understood at the beginning of the pandemic.

1.6 Conclusion

Growing racial tension is a serious challenge facing society. Understanding how racial animus forms and spreads is a critical step in addressing the issue. Using evidence from the COVID-19 pandemic, our paper sheds light on how and why negative shocks incite racial animus, types of individuals susceptible to such shocks, and factors that help spread the animus.

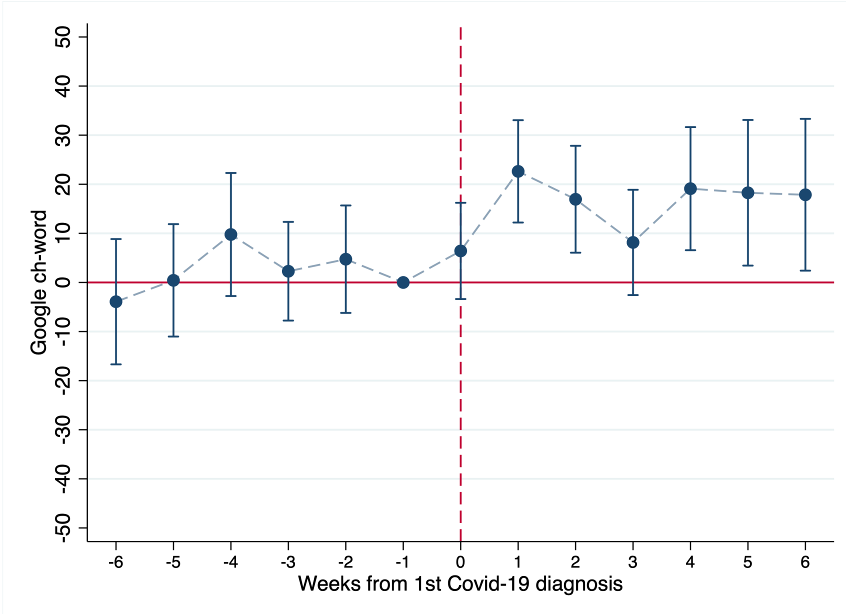
We exploit variation in the timing of the first COVID-19 diagnosis across US areas and find that the first local case leads to an immediate increase in local racial animus. This rise in animus specifically targets Asians, implying that the association between this group and the potential geographical origin of the virus likely motivates the animosity. The majority of racist tweets come from users who post the epithet for the first time; these first-time ch-word users are more likely to have expressed animosity against non-Asian minorities in the past, and their interaction with anti-Asian individuals predicts the timing of their first ch-word tweets. These findings suggest that preconceived notions about minorities and social media network both help in the formation and the spread of racial hatred amid crisis. Moreover, users who list "Trump" in their profiles are more susceptible to the pandemic shock; online animosity and offline hate incidents against Asians both increase when President Trump more frequently links China and COVID-19 in his tweets. These findings underscore the crucial role of public figures in influencing public opinions of a subject matter. Finally, the pandemic-driven racial animus we documented may persist beyond the duration of the pandemic, as most racist tweets do not explicitly mention the virus.

Our findings have practical implications. Careful naming of a shock, debunking claims of any alleged connection between a shock and a group, moderating racist individuals and their interaction with others on social media, and harnessing public figures' opinion-shaping power could all be helpful in curbing animus amid future crises.

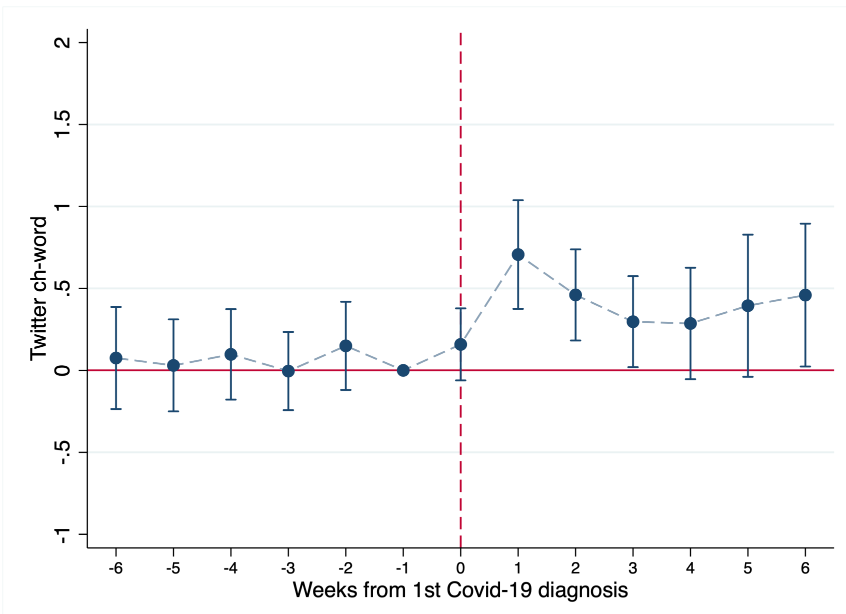
This paper also opens up several avenues for future research. While we estimate the effect of pandemics on racial animus, it would be interesting to know the downstream consequences of such crisis-driven animus, for example, on labor market, geographical sorting, and immigration. We characterize the users who are more susceptible to pandemic-induced animus against Asians, and it would be useful to characterize the users who express animosity against minorities in general so as to predict such behaviors and proactively curb the spread of racist content online.

1.7 Conclusion

Chapter 1, in full, is a reprint of the material as it appears in Journal of Economic Behavior and Organization, forthcoming. Runjing Lu; Sophie Yanying Sheng., Elsevier, 2022. The dissertation author was a primary investigator and author of this paper.



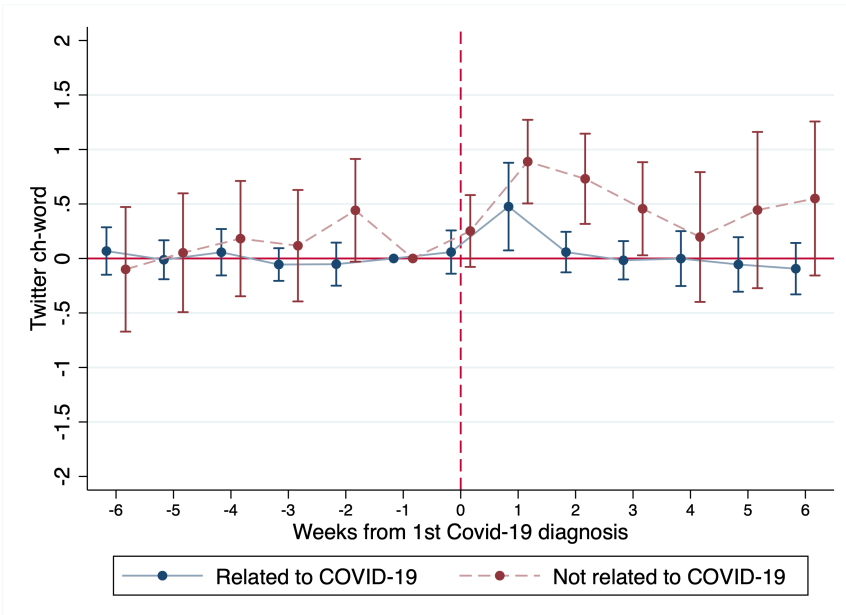
A: Google search index



B: Twitter post index

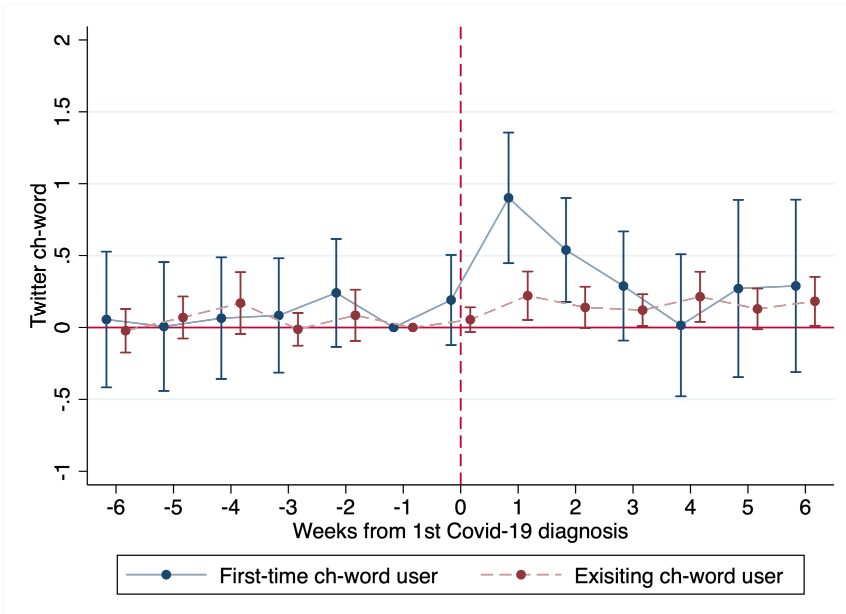
Note: The figure presents the effect of the first local COVID-19 diagnosis on the racially charged Google search index and Twitter post index. See section 1.2.1 for the definitions of the indexes. Panels A and B plot the estimates and 95 percent confidence intervals of the coefficients on the event dummies in equation 3.3 using the Google search index and the Twitter post index as the outcome, respectively. The estimates in panels A and B correspond to column (1) of Table 1.2 and column (1) of Table 1.3. Regressions control for year-month fixed effects and media market (panel A) or county (panel B) fixed effects. Standard errors are clustered by media market (panel A) or by county (panel B).

Figure 1.1: The Effect of the First Local COVID-19 Diagnosis on Racial Animus



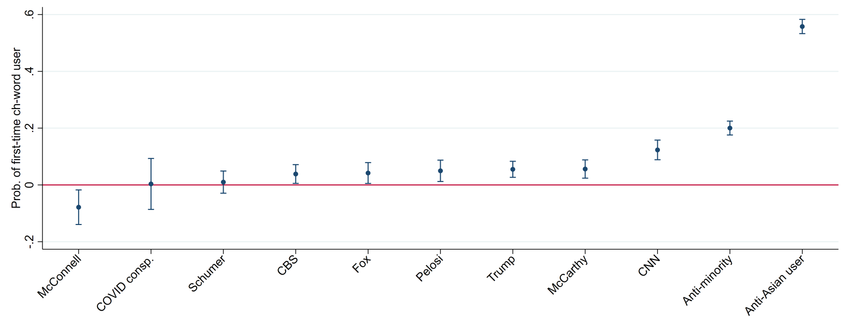
Note: The figure presents the effect of the first local COVID-19 diagnosis on the racially charged Twitter post index by whether or not the tweets are related to COVID-19. COVID-related racially charged Twitter post index are defined as the number of ch-word tweets including keywords: "COVID-19", "COVID", "virus", "pandemic", or "epidemic", per 100,000 "the" tweets. The solid blue (dashed red) line plots the estimates and 95 percent confidence intervals of the coefficients on the event dummies in equation 3.3 using the (non-) COVID-related Twitter post index as the outcome. All regressions control for year-month fixed effects and county fixed effects. Standard errors are clustered by county.

Figure 1.2: The Effect of the First Local COVID-19 Diagnosis on Racially Charged Tweets COVID-Related vs Non-COVID-Related Tweets

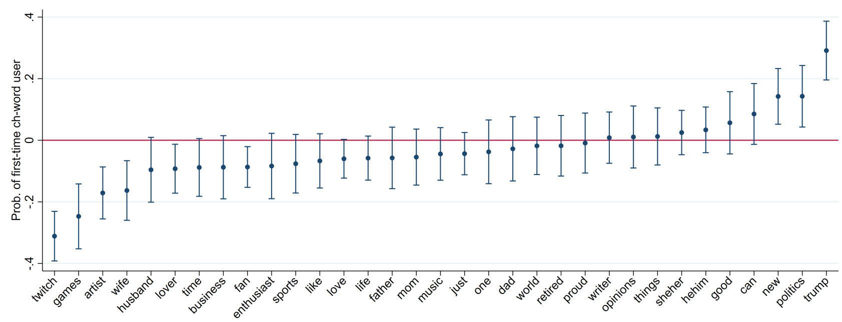


Note: The figure presents the effect of the first local COVID-19 diagnosis on the racially charged Twitter post index by whether the posting user is a first-time or an existing ch-word user. See section 1.4.3 for definitions of first-time and existing ch-word users. The solid blue (dashed red) line plots the estimates and 95 percent confidence intervals of the coefficients on the event dummies in equation 3.3 using the racially charged Twitter post index based on first-time (existing) ch-word users as the outcome. All regressions control for year-month fixed effects and county fixed effects. Standard errors are clustered by county.

Figure 1.3: The Effect of the First Local COVID-19 Diagnosis on Racially Charged Tweets First-time vs Existing Ch-word Users



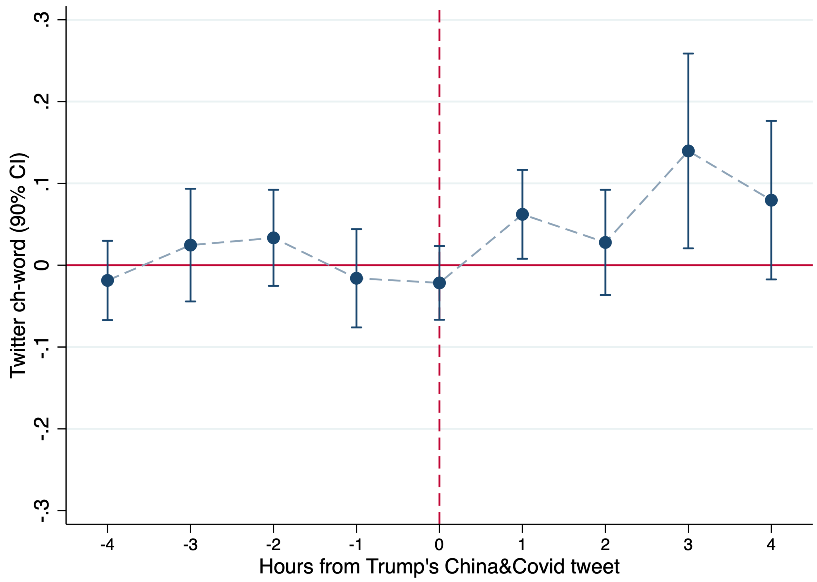
A: Twitter activity



B: Twitter user profile keywords

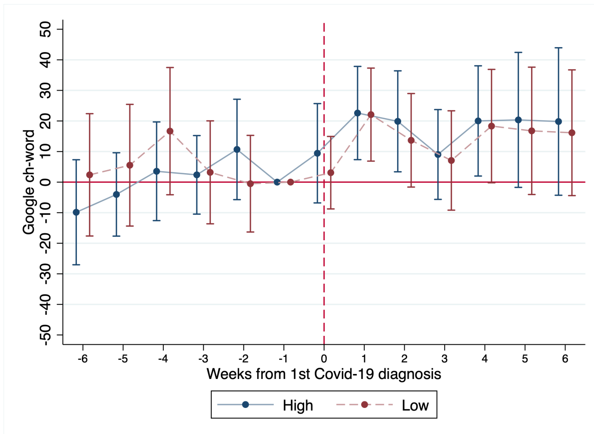
Note: This figure presents the relationship between being a first-time ch-word user and one's Twitter activity and user profile keywords. Panels A and B plot the coefficients and 95 percent confidence intervals from regressing an indicator for being a first-time ch-word user on user's pre- and mid-pandemic Twitter activity, and user profile keywords, respectively. Both regressions control for account age, log number of followers, and log number of followings. Regressors in panel A are defined as follows: "Anti-Asian user" is one if an user has interacted with other ch-word users before the pandemic; "Anti-minority" is one if an user has tweeted racial epithets against non-Asian minorities (the n-word, w-word, and k-word) before the pandemic; "Trump" is one if an user has ever mentioned #trump or @realDonaldTrump before the pandemic; "McCarthy", "McConnell", "Pelosi", "Schumer", "Fox", "CNN", and "CBS" are similarly defined using @kevinomccarthy, @McConnellPress (or @LeaderMcConnell), @SpeakerPelosi, @SenSchumer, @cnn, @foxnews, @cnn, and @cbsnews as keywords, respectively; "COVID consp." is one if an user has ever tweeted keywords related to COVID-19 conspiracies (i.e., plandemic, fakepandemic, scamdemic, film your hospital, 5gcoronavirus, or coronavirustruth) by the end of our sample period. Regressors in panel B are the 25 most common user profile words used by first-time ch-word users and the 25 most common user profile words by control users. There is an overlap between the two sets of words, so the number of words included in the regression is less than 50. Standard errors are heteroscedasticity-consistent. Regression results are reported in Tables A.6 and A.6.

Figure 1.4: Predictors of being First-time Ch-word Users

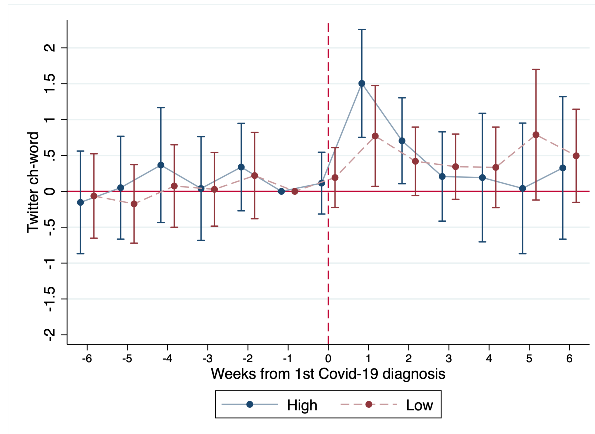


Note: The figure presents the relationship between the number of President Trump's tweets that mention both Covid-19 and China (China-and-COVID tweets) in an hour and the number of ch-word tweets per 100,000 "the" tweets nationwide in the four hours before and the four hours after the president's tweets. The figure plots the estimates and 90 percent confidence intervals of the coefficients on the interactions between hourly event dummies and the number of Trump's China-and-COVID tweets at hour zero. Event dummy for the hours outside of those being plotted are omitted. The regression controls for year-week fixed effects, day of week fixed effects, and hour fixed effects. Standard errors are clustered by date.

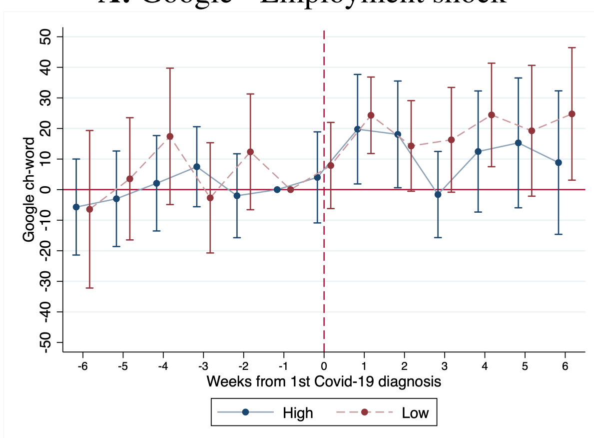
Figure 1.5: Relationship between Racially Charged Tweets Nationwide and Trump Tweets



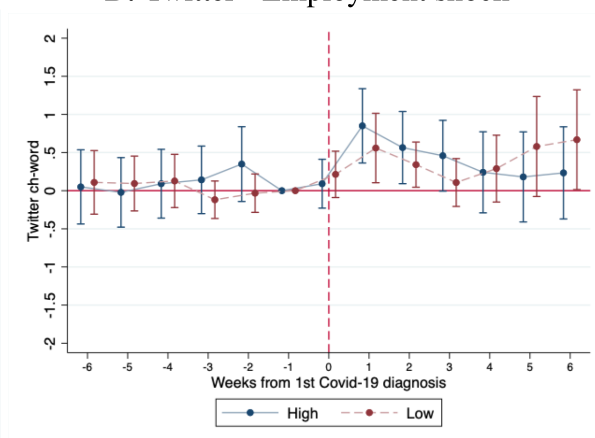
A: Google - Employment shock



B: Twitter - Employment shock



C: Google - Revenue shock



D: Twitter - Revenue shock

Note: The figures present the heterogeneous effect of the first local COVID-19 diagnosis on the racially charged Google search index and Twitter post index by the negative economic impact of the pandemic. Panels A and B partition the regression sample by whether the proportion of an area's annual average employment in "leisure and hospitality" and "education and health services" is above or below the sample median (i.e., 32 percent in the Google sample and 35 percent in the Twitter sample). Panels C and D partition the regression sample by whether the percent change in net revenue among local small businesses between January and March is above or below the sample median (i.e., -39 percent in the Google sample and -37 percent in the Twitter sample). Panels A and C (B and D) plot the estimates and 95 percent confidence intervals of the coefficients on the event dummies in equation 3.3 using the racially charged Google search index (Twitter post index) as the outcome. Specifications in panels A and C mirror those in column (1) of Table 1.2, and specifications in panels B and D mirror those in column (1) of Table 1.3.

Figure 1.6: The Effect of the First Local COVID-19 Diagnosis on Racial Animus by the Negative Economic Impact of COVID-19

Table 1.1: Relationship between Racial Animus, Hate Crimes, and Chinese Restaurant Visits

VARIABLES	(1) Incidents	(2) Incidents	(3) Incidents	(4) Incidents	(5) Visits	(6) Visits	(7) Visits	(8) Visits
Panel A: Google search index								
Google ch-word(t)	0.00057*	0.00057*	0.00057*	0.00057*	-21.069*	-22.017*	-23.489**	-24.381**
	(0.00031)	(0.00031)	(0.00031)	(0.00031)	(12.629)	(12.704)	(11.632)	(11.697)
Google ch-word(t-1)			-0.00019	-0.00018			-13.934	-14.900
			(0.00035)	(0.00035)			(11.467)	(11.594)
Total visits					0.052***	0.052***	0.048***	0.048***
					(0.005)	(0.005)	(0.006)	(0.006)
Population(m)	0.51926***	0.51907***	0.52042***	0.52287***				
	(0.19939)	(0.19942)	(0.19936)	(0.19991)				
Google Asian(s)(t)		-0.00005		-0.00018		-138.636**		-120.800**
		(0.00109)		(0.00111)		(60.169)		(54.317)
Google Asian(s)(t-1)				0.00075				10.816
				(0.00163)				(59.669)
Observations	3,600	3,600	3,600	3,600	1,440	1,440	1,380	1,380
R-squared	0.309	0.309	0.309	0.309	0.996	0.996	0.997	0.997
Outcome mean	.147	.147	.147	.147	104962.736	104962.736	104962.736	104962.736
Panel B: Twitter post index								
Twitter ch-word	-0.00041	-0.00037	-0.00048	-0.00047	-60.249	-61.190	-30.883	-30.178
	(0.00099)	(0.00096)	(0.00099)	(0.00100)	(49.798)	(50.150)	(42.641)	(42.737)
Twitter ch-word (t-1)			-0.00063	-0.00064			-10.216	-10.089
			(0.00065)	(0.00065)			(36.060)	(36.286)
Total visits					0.047***	0.047***	0.044***	0.044***
					(0.005)	(0.005)	(0.006)	(0.006)
Population(m)	0.56488***	0.56517***	0.58097***	0.58136***				
	(0.17387)	(0.17391)	(0.17623)	(0.17630)				
Twitter Asian(s)(t)		-0.00003		-0.00002		0.294		-0.279
		(0.00003)		(0.00002)		(0.682)		(0.646)
Twitter Asian(s)(t-1)				-0.00001				-0.266
				(0.00002)				(0.693)
Observations	11,116	11,116	10,921	10,921	4,493	4,493	4,300	4,300
R-squared	0.220	0.220	0.220	0.230	0.996	0.996	0.997	0.997
Outcome mean	.057	.057	.057	.057	41065.784	41065.784	41065.784	41065.784

Notes: The table presents the relationship between the racially charged Google search index and the Twitter post index, anti-Asian hate crimes, and visits to Chinese restaurants. Hate crime data are from the FBI UCR, visit data are from Safegraph, and all data are at the media market×year-month level. Outcome variables are the monthly number of anti-Asian hate crimes between January 2014 and December 2018 (columns (1)-(4)) and the monthly number of visits to Chinese restaurants between January 2018 and December 2019 (columns (5)-(8)). *Google Asian(s)* is the Google search index for the word “Asian(s).” *Twitter Asian(s)* is the number of tweets including “Asian(s)” per 100,000 “the” tweets. All regressions control for local unemployment rate, year-month fixed effects, and media market fixed effects.
** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 1.2: The Effect of the First Local COVID-19 Diagnosis on Racial Animus Google Search Index

VARIABLES	(1) Ch-word index	(2) Severity control	(3) Asian control	(4) Exclude states
-6w	-3.920	-2.694	-4.265	-8.979
	(6.379)	(6.620)	(6.404)	(8.341)
-5w	0.431	1.100	-0.198	-2.575
	(5.722)	(5.820)	(5.699)	(7.083)
-4w	9.764	10.088	9.419	9.205
	(6.263)	(6.316)	(6.233)	(7.649)
-3w	2.282	2.503	2.247	2.458
	(5.023)	(5.085)	(5.020)	(5.912)
-2w	4.739	4.899	4.771	2.564
	(5.469)	(5.535)	(5.467)	(6.150)
+0w	6.421	6.326	6.274	6.574
	(4.898)	(4.911)	(4.864)	(5.127)
+1w	22.628***	22.442***	22.030***	22.771***
	(5.210)	(5.246)	(5.280)	(5.721)
+2w	16.945***	15.936***	16.727***	18.104***
	(5.439)	(5.443)	(5.407)	(5.621)
+3w	8.155	5.702	7.894	8.614
	(5.359)	(5.907)	(5.403)	(5.829)
+4w	19.106***	15.972**	18.873***	19.527**
	(6.265)	(6.999)	(6.253)	(7.461)
+5w	18.263**	15.375*	18.041**	14.709*
	(7.411)	(8.113)	(7.428)	(8.679)
+6w	17.861**	15.002*	18.125**	18.017*
	(7.726)	(8.046)	(7.751)	(9.267)
Observations	780	780	780	663
R-squared	0.190	0.192	0.193	0.180
Outcome mean	30.03	30.03	30.03	30.03

Notes: The table presents the effect of the first local COVID-19 diagnosis on the racially charged Google search index. All columns report the estimates of coefficients on the event dummies in equation 3.3. Column (1) corresponds to Figure A.4, panel A. Column (2) controls for the number of COVID-related new cases and deaths and whether the state has any stay-at-home orders in place. Column (3) controls for the Google search index for "Asian(s)." Column (4) excludes Washington, New York, and California. All regressions control for media market fixed effects and year-month fixed effects. Standard errors are clustered by media market.
** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 1.3: The Effect of the First Local COVID-19 Diagnosis on Racial Animus Twitter Post Index

VARIABLES	(1) Ch-word index	(2) Ch-word level	(3) Ch-word per capita	(4) Exclude counter-hate	(5) Dynamic DID	(6) Severity control	(7) Asian control	(8) Exclude states	(9) Exclude bots
-6w	0.075	-0.022	0.061		-0.037	0.070	0.053	0.127	0.098
	(0.159)	(0.276)	(0.307)		(0.097)	(0.159)	(0.255)	(0.165)	(0.151)
-5w	0.030	-0.069	-0.036	-0.801	-0.085	0.027	0.091	0.056	0.039
	(0.143)	(0.158)	(0.267)	(1.252)	(0.091)	(0.143)	(0.242)	(0.142)	(0.153)
-4w	0.098	-0.128	-0.117	-0.328	-0.025	0.095	0.248	0.113	0.075
	(0.140)	(0.165)	(0.240)	(1.181)	(0.107)	(0.140)	(0.239)	(0.141)	(0.144)
-3w	-0.004	0.024	-0.100	0.450	-0.082	-0.006	0.095	0.018	0.014
	(0.121)	(0.091)	(0.195)	(1.152)	(0.081)	(0.121)	(0.213)	(0.129)	(0.138)
-2w	0.150	0.065	0.412	-0.361	0.120	0.149	0.331	0.136	0.242
	(0.137)	(0.050)	(0.308)	(0.967)	(0.094)	(0.137)	(0.212)	(0.146)	(0.180)
+0w	0.158	0.012	0.390**	5.154***	0.120***	0.163	0.168	0.169	0.203
	(0.112)	(0.069)	(0.170)	(1.005)	(0.094)	(0.159)	(0.171)	(0.122)	(0.142)
+1w	0.707***	0.227**	1.037***	5.075***	0.689***	0.718***	1.077***	0.572***	0.952***
	(0.169)	(0.105)	(0.197)	(1.046)	(0.159)	(0.166)	(0.238)	(0.162)	(0.228)
+2w	0.460***	0.348***	1.140***	2.855***	0.428***	0.478***	0.763***	0.389**	0.538***
	(0.142)	(0.109)	(0.252)	(1.039)	(0.111)	(0.145)	(0.199)	(0.151)	(0.173)
+3w	0.297**	0.631***	1.331***	2.688***	0.181*	0.315**	0.526**	0.300*	0.255*
	(0.141)	(0.193)	(0.396)	(0.842)	(0.095)	(0.152)	(0.204)	(0.154)	(0.137)
+4w	0.286*	0.789**	1.947**	1.521	0.122	0.307*	0.361	0.273	0.132
	(0.173)	(0.310)	(0.771)	(1.257)	(0.103)	(0.184)	(0.269)	(0.187)	(0.157)
+5w	0.394*	0.683***	1.650***	1.158	0.240	0.421*	0.535*	0.385	0.144
	(0.221)	(0.201)	(0.466)	(1.396)	(0.154)	(0.248)	(0.323)	(0.240)	(0.178)
+6w	0.459**	0.696***	1.664***	2.264	0.340**	0.489*	0.533*	0.479**	0.373*
	(0.222)	(0.223)	(0.469)	(1.566)	(0.150)	(0.252)	(0.315)	(0.243)	(0.198)
Observations	7,930	7,976	7,976	3,141	103,694	7,930	5,578	7,188	11,811
R-squared	0.121	0.809	0.270	0.611	0.112	0.121	0.142	0.123	0.060
Outcome mean	0.591	0.681	1.075	6.779	0.591	0.591	0.591	0.591	0.569

Notes: The table presents the effect of the first local COVID-19 diagnosis on the prevalence of ch-word tweets in an area. All columns report the estimates of coefficients on the event dummies in equation 3.3, except for column (5). Column (1) corresponds to Figure A.4, panel B. The outcome variable in column (2) is the number of ch-word tweets, and the regression controls for the number of "the" tweets. The outcome variable in column (3) is the number of ch-word tweets per one million county population. Column (4) uses an alternative Twitter post index, which removes counter-hate tweets (see section 1.4.2). Column (5) presents the estimates from a dynamic DID event study (93). Column (6) controls for the number of COVID-related new cases and deaths and whether the state has any stay-at-home orders in place. Column (7) controls for the Twitter post index for "Asian(s)." Column (8) excludes Washington, New York, and California. Column (9) excludes tweets from users who are likely Twitter bots. All regressions control for county fixed effects and year-month fixed effects. Standard errors are clustered by county.

** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 1.4: Predictors of Tweeting Ch-word among First-Time Ch-word Users after the First Local COVID-19 Diagnosis

VARIABLES	(1) P(ch-word) (t+1)	(2) P(ch-word) (t+1)
Anti-Asian user(t)	0.281***	0.259***
	(0.072)	(0.072)
Anti-minority(t)	1.156	1.112
	(1.360)	(1.368)
COVID consp.(t)	1.314	1.177
	(1.861)	(1.845)
Trump(t)	0.325***	0.297**
	(0.122)	(0.122)
McCarthy(t)	-0.184	-0.097
	(0.456)	(0.456)
McConnell(t)	-1.969***	-2.045***
	(0.351)	(0.449)
Pelosi(t)	-0.373	-0.419
	(0.284)	(0.285)
Schumer(t)	-0.031	-0.106
	(0.379)	(0.379)
CBS(t)	-0.636	-0.692
	(0.824)	(0.825)
CNN(t)	0.164	0.173
	(0.277)	(0.277)
Fox(t)	-0.430	-0.386
	(0.348)	(0.346)
Account years	-0.002	-0.002
	(0.005)	(0.005)
Log(followers)	-0.032**	-0.031**
	(0.013)	(0.013)
Log(followings)	-0.010	-0.009
	(0.020)	(0.020)
New diagnoses		-0.000
		(0.000)
New deaths		0.000
		(0.000)
Observations	174,164	174,164
R-squared	0.002	0.004
Outcome mean	1.251	1.251

Notes: This table presents the relationship between first-time ch-word users' likelihood of tweeting the ch-word in a day (in the unit of percentage point) and their Twitter activity in the day before as well as their baseline characteristics. See note to Figure 1.4 for definitions of the independent variables. The data is at the user×day level, and all regressions control for county, year-of-week, and day-of-week fixed effects. Standard errors are clustered by user.
** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 1.5: Relationship between Trump Tweets and Racial Animus Nationwide

VARIABLES	(1) Twitter ch-word	(2) Twitter ch-word	(3) Log(incidents)	(4) Log(incidents)
China-and-COVID(t)	0.0482**	0.0493**	0.0799*	0.0888**
	(0.0234)	(0.0246)	(0.0453)	(0.0398)
China only(t)	-0.0126	-0.0130	-0.0592	-0.0332
	(0.0131)	(0.0133)	(0.0815)	(0.0844)
COVID only(t)	0.0008	0.0006	-0.0014	0.0004
	(0.0038)	(0.0040)	(0.0146)	(0.0143)
New diagnoses		-0.0000		0.0000*
		(0.0000)		(0.0000)
New deaths		0.0001		0.0001
		(0.0001)		(0.0003)
Observations	123	123	45	45
R-squared	0.519	0.522	0.812	0.829
Outcome mean	.344	.344	3.1932	3.1932

Notes: The table presents the relationship between the number of President Trump's tweets about COVID-19 and/or China and racial animus nationwide. The outcome variable in columns (1) and (2) is the daily number of ch-word tweets per 100,000 "the" tweets nationwide between January 1, 2020 and May 2, 2020. The outcome variable in columns (3) and (4) is the natural log of the daily number of anti-Asian hate incidents nationwide from AP3CON Stop AAPI Hate Reporting system between March 19 and May 2, 2020. We categorize the president's tweets that include "china", "chinese", "huawei", "xi", "COVID", "COVID-19", "corona", "coronavirus", "virus", "epidemic", or "pandemic" into three categories: "China-and-COVID" is the daily number of the president's tweets mentioning both China and COVID-19; "China only" those mentioning only China; and "COVID only" those mentioning only COVID-19. "New diagnoses" and "New deaths" are the daily number of COVID-related new cases and deaths in the United States. All regressions control for year-week fixed effects and day-of-week fixed effects. Standard errors are clustered by date.
** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Chapter 2

Doctor Who — Can physicians from low socio-economic status families reduce the socio-economic gradient in health?

There is a strong positive relationship between socio-economic status (SES) and health in developed economies (19, 30). The *health-SES gradient* is observed in many dimensions of health: low-SES individuals have worse self-reported health, more chronic conditions, and shorter life expectancy (69). The gap in life expectancy in the US between a college and a high school graduated men is 7 years (24, 63). We observe a similar SES gradient in health in countries with universal health care access and the most equal income distributions. Mitigating this inequality in health is at the top of the policy agenda globally.

A large literature studies how either patient characteristics (21, 77) or physician behavior (84, 35, 87) explains the health-SES gradient. In this paper, we investigate the importance of the interaction between physicians and patients. Understanding the interaction between the two has important policy implications for optimizing physician-patients matches and efficiency in government health care spending. Existing interventions to dampen the gradient often involve safety nets that provide patients with more stable income flow and less discontinuity in care, rather than physician-patient match quality . Previous studies have found that similarities between physicians and patients, such as gender and racial concordance or family ties, can improve patient health (17, 43, 1, 42, 49). However, the relationship between physicians' and patients' SES and the health-SES gradient is unexplored, despite the interest in the nature of the health-SES gradient.

In this paper, we ask: can primary care physicians from low-SES families reduce the SES-gradient in health? We focus on mortality as a main outcome of health and study the potential mechanisms by investigating different causes of death and patient health behaviors. Since physicians are highly educated, we use the physicians' childhood SES to define their SES. Patients are low-SES if their own highest level of education is primary school.[1] Unlike gender and race, physicians' SES is unobserved and difficult to infer by the patient. We focus on the primary care physicians (henceforth physicians) because the quality of physician-patient relationship is especially important in this setting. Primary care physicians' responsibilities cover almost all aspects of everyday health; they provide continuous interaction with patients, make diagnoses, prescribe drugs, act as gatekeepers to medical specialists, and work with patients to manage chronic conditions.

We use Danish population-wide administrative data of patients between ages 30-70 to study SES concordance effects. The Danish setting is ideal for the research question, as it allows us to track families across generations, and merge this information with physicians' practices, patients' healthcare utilization, and health outcomes. Universal healthcare coverage in Denmark allows us to zoom in on the effect of physician-patient match and rule out effects attributed to differences in healthcare costs and insurance selection.

The primary challenge in answering our research question causally is that physician-patient matches may be endogenously created. We exploit variation induced by clinic closures, a

[1] Primary school in this setting is equivalent to completing ninth grade where children are approximately 16 years old.

cause for physician-patient separation that is plausibly exogenous to patients' health trajectories (89, 38). Despite separation being plausibly exogenous, there remains concerns that selection exists in the physician assignment post clinic closures.[2] We address this concern by comparing high- and low-SES patients within groups that have the same physician before and after the clinic closes in a triple differences design. We compare the health behavior and outcome between high- and low-SES patients (first difference) before and after closure (second difference) who get new physicians from either a high- or low-SES family (third difference).

We find that SES concordance between physicians and patients decreases the SES gradient in mortality, measured by the difference in mortality between high- and low-SES patients, by 24.8%. The reduction comes from lower mortality rates for low-SES patients who are matched with low-SES physicians in the post period, whose behaviors in turn are particularly sensitive to their physician match. High-SES patients' mortality does not depend on their physician's SES. This means that the reduction in SES-gradient in mortality is *not* caused by harming the high-SES patients, but improving the health of low-SES patients. Importantly, we do *not* find other attributes of the physician, including academic performance, graduating institution, experience, or gender, to contribute to the effect we find.

To explore the origin of the reduction in SES-mortality gradient, we first break down mortality by cause. We focus on deaths by chronic conditions, as primary care physicians hold the central role for the diagnosis and management of these conditions (79). We find that the effect on overall mortality is driven by a large reduction in cardiovascular mortality, especially driven by men. Next, we explore how SES concordance affects patients' health behaviors. We find that low-SES patients matched with low-SES physicians receive more care at the intensive margin (more visits to physicians, more services per visit, and higher reimbursement to medical specialists), but not at the extensive margin (likelihood of making *any* office visits). Furthermore, we find that SES concordance increases treatment of chronic conditions for low-SES patient through

[2] We do not find any evidence that patients select their new physicians based on physicians' SES, potentially because this characteristic is unobserved by the patient.

more disease detection for new patients and better adherence for pre-existing patients. For example, we find that SES concordance increases adherence to statins, a medicine that prevents major heart attacks, among patients who are previously diagnosed. Both previously and newly diagnosed diabetic patients display higher use of annual diabetic check ups.

We hypothesize that concordance in physician-patient SES may affect patient health through the following channels: (1) the patient establishes better trust and partnership with the physician (92, 1); (2) the physician is more cognizant of the low-SES patients' health conditions and risks due to occupation and lifestyle constraints; (3) the physician is able to make relevant information more salient to the patient and is better able to understand the patient's way of communicating (105), thereby improving patients' health literacy (17). While direct tests for these hypotheses are not feasible in our data, we provide suggestive evidence related to the second and third. Related to the second channel, we find that physicians whose family members suffer from a chronic condition reduce the SES mortality gradient. The third hypothesis can be tested via patient adherence to medical guidelines (105), and the adherence results described above provide suggestive evidence of medical communication effectiveness.[3]

Our paper makes three novel contributions. First, we demonstrate that physician-patient SES concordance can mitigate the SES inequality in health. Second, we bridge the literature on health inequality to the literature on physician practice style. Third, our study demonstrates that childhood SES is a relevant and important factor for how physicians interact with patients. We discuss our contribution to three strands of literature below.

First, our paper builds upon a literature studying physicians practice styles (see e.g. (15) for a review). Differences in physicians' behavior translate into differences in quality of care (88, 38). What affects the physician's practise style? Studies show that physician's skill (22, 35), their medical training or quality (25, 84), and their personal believes about the benefit of a treatment (23) matters for their practise style. (89) studies the effect of discontinuity in care on

[3]We test the first hypothesis on trust and partnership by looking at whether patient churn differs between treatment and control groups, and we do not find this to be the case.

patient health in Denmark and finds that disruption in care increases reimbursement per visit and the detection of chronic conditions. We use a similar design and setting as (89), but estimate the effect of being matched with a similar physicians in terms of SES on top of the effect of discontinuity in care. We contribute to the above literature by showing that the physicians' family background is an important factor when studying differences in practice styles.

Second, this paper is closely related to a literature on matching quality as a mechanism that improves outcomes. In educational settings, (31) finds demographically similar teachers improve student outcomes. (57) finds that having a female boss increase the chance of advancing a rank for female workers. In the medical setting, (1) studies racial physician-patient concordance using a randomized controlled experiment. They estimate that racial concordance between physician and patient can reduce the the black-white gap in cardiovascular mortality substantially, and the improvement is largely driven by better communication. In non-experimental settings, (42, 43) and (49) find that physician-patient concordance in terms of race and gender reduces within-hospital mortality. Having familial access to medical expertise, a form of close social concordance, can improve health (17) and change health behaviors, although the evidence is mixed (5). We contribute to this strain of the literature, by focusing on a type of concordance that is under-explored, not directly observable, but universally policy relevant as it addresses the SES gradient in health directly.

2.1 Institutional Settings and Data

This section describes the institutional setting and data that allows for our identification strategy.

Denmark has tax-funded universal public health insurance that provides free and equal access for all citizens. Primary care clinics are privately owned, and are reimbursed on a mixed capitation and fee-for-service system. Primary care physicians are gatekeepers of the healthcare system; they perform initial diagnosis, treat illnesses, prescribe medication, manage chronic conditions, and refer patients to medical specialists. The tasks they face vary widely and often require intensive communication and continuous relationship with the patient (48, Chapter 1). SES concordance may matter especially in the primary care setting since a common cultural background and familiarity in low-SES lifestyle constraints may make low-SES physicians more cognizant of their health risks and conditions; it may also help facilitate medical communication (99).

Our identifying variation is induced by clinic closures, a vast majority of clinic closures (74%) are due to retirement.[4] New assignment of physicians and patients take place in three ways upon closures: (1) if the old physician choose to sell the clinic to another physician, the patient list is sold over along with the clinic; (2) if the clinic is not sold, patients can choose a new primary care physician online, conditioning on the availability of open clinics that accept new patients; (3) if patients do not make an active choice, they are assigned a clinic by the municipality.

In the second scenario, patients are informed about the number of physicians in the clinics, as well as the physicians' name, gender, and age upon making a choice. From this informationit is difficult for patients to infer the type of childhood household the physicians is from. Patients cannot observe the physicians' graduating institutions on this web-page either. In the period of interest, there were three medical schools in Denmark following similar curriculum and provide similar quality of training.[5] Patients can only choose among clinics that accept new patients. In the analysis period, many municipalities had a critical shortage of primary care physicians and many clinics did not accept new patients, restricting the choice of the patient. [6]

[4]Retirement is defined as the average age in the clinic being over 60 years at the time of clinic closure following (89).

[5]University of Copenhagen (UCPH), Aarhus (AU), and Odense (SDU). Aalborg University introduced a program in Medicine in 2010. The University of Copenhagen is the most popular institution to study Medicine, as measured both in terms of number of applicants and GPA cut-off.

[6]Clinics can stop the intake of new patients if they have more than 1600 patients per physician, and have to stop taking any new patients when the number reaches 2500. Clinics must take all patients that choose them when the list is open.

2.1.1 Data

To study the effect of physician-patient SES concordance on health and healthcare utilization, the ideal data requires linking each physician to demographic information of their parents, and merging this with information with their patients' health, healthcare utilization, and demographics. The Danish population-wide administrative data are one of the few data sources that allows for such an analysis on the population level. We describe how the analysis sample and variables of interest are constructed constructed below.

Constructing the analysis sample

To construct the patient analysis sample, we start with all adults between ages 30-70 in the entire Danish population between 1995 and 2017. We use the Danish National Health Service Register and follow (54) to link every adult to the corresponding primary care clinic on a annual basis.[7] We find clinics that close between 1999 and 2016 and define the closure year as the last year with registered services for the clinic. We include patients the first time they experience a clinic closure, and define their new clinic, as the clinic that patients are connected to in the year after closure of their original clinic. We observe 776 clinic closures affecting more than 480,000 adult patients in the analysis period, see Table 2.2. We add demographic information about each patient in the clinic, including their highest level of education. Our main analysis sample is balanced in the pre-period, such that we observe patients four years before clinic closure. The patients may pass away in the post period, and their mortality is a core outcome of interest.

After linking patients to clinics, we use Service Provider Registry to add the ID of the physicians in the clinic. Using physician IDs, we obtain physicians' demographics and their parents' levels of education in the registers. We are only able to match patients to physicians at the clinic level. Around 61% of clinics are non-solo; on average, there are 1.8 physicians per clinic in Denmark. For this reason, we the aggregate physician SES to the clinic level. In the

[7]We can match patients and GPs with more than 98 % accuracy using this algorithm (54).

main analysis, a clinic is defined as low-SES if it has one or more low-SES physicians working there. We use two alternative definitions in section 2.3.4 as robustness checks. We also check the robustness of our results, using a sub-sample of physicians whose parents we are able to observe.

Measurement of socio-economic status

We use highest level of completed education as a main determinate of SES. We define a patient to be low-SES if he/she has primary school as the highest level of completed education, which corresponds to 9 years of schooling. To identify physicians' SES, we use their parents' highest level of education. A physician is defined as low-SES if at least one parent has primary school as their highest level of completed education. Parental education is missing for most people born before 1960 in the Danish data (see Appendix Figure B.1 Panel D). This means that for most physicians born before 1960, we are unable to identify their SES. They make up 79% of the primary care physicians working in closing clinics and 34% of physicians working in non-closing clinics in our sample. In our main analysis, we assume that physicians for whom we do not observe their SES are high-SES.[8] We discuss how this affects our identification strategy in section 2.2 and test the robustness of this assumption in section 2.3.4.

Measurement of health behaviors

After defining the population of interest, we construct the relevant outcome variables. Patient mortality is a primary outcome of interest. We identify patient mortality and cause of death using the Cause of Death Registry. We use the Health Insurance Registry to identify the number of visits the patient have at the clinic, the number of services the physician conducts for each patient visit, and the total expenditure the physician is reimbursed by the region for the services

[8]Most physicians born before 1960 attend medical school between 1959 and 1976, when most students in medical schools are from high-SES families (102).

provided to the patient.[9] Number of visits and services provided per visit per year are calculated conditioning on having at least one visit that year. We also use the Health Insurance Registry to identify if the patient receives any specialized care, as well as physician reimbursement amount.

Measures related to chronic conditions

To study mechanisms of mortality effects, we focus on the four most unequally distributed chronic conditions. They account for the majority of the global and national burden of diseases, are leading causes of deaths, and primary physicians are central to the management of these conditions (79): cardiovascular conditions (CVC), cancer, diabetes and chronic obstructive pulmonary disease (COPD) (95).[10] Many of the common chronic conditions are under-diagnosed. E.i. (39) finds that CVC have a under-diagnosis rate of 30-60%, COPD 70-80%, diabetes 20-50%. Although primary care plays a central role in managing chronic conditions (79, 103), diagnosis are only recorded in hospital claims in the Danish data.[11] Absent of accurate records of diagnosis, we use outcomes related to the different chronic condition such as first-line treatments or services related to the detection of these conditions.[12] Note that using treatment to infer diagnosis is imperfect. While we are unable to give precise estimates on whether physicians are under-diagnosing or over-treating, improvements in health outcomes may suggest under-diagnosis in the pre-period.

The four conditions have the following in common: (1) they have a close links with health behaviors such as smoking, exposure to pollutants, and diet, (2) early detection of these diseases leads to higher survival rates, (3) The diagnosis process requires communication between primary care and patient, and (4) reducing disease progression in the early stages often

[9] Examples of a service in the Danish data are blood tests, consultation, or phone consultation. Visits can be in-person office visit or phone consultation.

[10] Cause of death is coded according to ICD-10. See Appenix Table B.1 for the used ICD-10 codes.

[11] The patients who are diagnosed in hospitals might have been diagnosed in non-hospital settings prior to hospital admissions; they are also at more severe stages of these conditions.

[12] We use Anatomical Therapeutic Chemical (ACT) classifications to code medical treatments; see Appendix Table B.2 for an overview of the codes used.

do not involve invasive treatments, but lifestyle changes (such as smoking cessation, limiting alcohol intake, balanced diet, and exercise) or medication.

Note that since the data does not capture patients' changes in health behaviors outside of the clinic, such as smoke cessation and dieting, our analysis misses potential effects in the early stages of these chronic diseases. Effects of early stage interventions, especially on mortality or hospitalization, may take longer to observe. The potential channel of early-stage effects implies that SES concordance may have longer-lasting impact on the health-SES gradient.

Cardiovascular Conditions (CVC) Cardiovascular conditions are the most common cause of death in the developed countries (76). Around 20% of deaths are caused by CVC in our sample. Low-SES patients are around 40% more likely to be treated for a CVC, and 57% more likely to die from a CVC. Guidelines for primary care physicians include assessing patients' risk of cardiovascular conditions using multivariate risk prediction algorithms (26), putting primary care at the center of identifying high-risk patients and preventing acute hospitalizations arising from chronic CVC. To infer CVC diagnosis in our data, we use prescriptions with statins and ACE inhibitors. These medications are considered first-line to treat hyperlipdemia and hypertention (26), and statins has been shown to reduce CVC mortality and major coronary events by 70 percent (81). Patients should not stop taking statins once they start; adherence is therefore key to survival.[13]

Chronic obstructive pulmonary disease (COPD) COPD is a group of chronic lung conditions that cause obstructed airflow from the lungs commonly caused by long term exposure to irritating particulate matters such as cigarette smoke, dust, or fumes. It is often misdiagnosed in the early stages, and the process of diagnosis involves a conversation between the physician and patient about exposure to irritants, family history, and symptoms (27). Although COPD is

[13]Statins have also been shown to increase blood sugar levels and increase the risk of diabetes II for exposed groups (80).

progressive, it could be well managed through smoking cessation alone in the early stages, and medication when the condition progresses. We infer COPD diagnosis using (1) prescriptions of common COPD medications[14], and (2) avoidable hospitalizations due to COPD.[15]

Diabetes Around 8% of the Danish adult population are diagnosed with diabetes. Individuals of low SES are around twice as likely to be diagnosed with diabetes compared to high-SES individuals (64). Diabetes is also closely associated with lifestyle – a healthy diet and regular exercise can delay or prevent it from occurring. In addition, diabetes are a common cause for heart diseases and strokes (28). Guideline published by the American Diabetes Association refers to a care model with "proactive practice teams and informed activated patient" as first-line. The chronic care model involves an annual checkups of diabetes complications. Hence, we look at the following diabetes related treatments (1) annual diabetes checkups with primary care physicians and (2) prescription of metformin.[16]

Cancer Cancer is the chronic disease that causes the most death in Denmark (59). While breast cancer is the most common type of cancer, lung cancer is the most common cause of cancer death (94). Lung cancer is often diagnosed after the disease has spread, as symptoms do not appear at early stages; 1-year survival rate is around 50-60 percent. Early detection is therefore key in increasing likelihood of survival. This is in contrast to breast cancer, where the one-year survival rate is over 97 percent for Danish women (98, 97). Different from the three diseases described above, the diagnosis and treatment primarily take place in specialists' offices or in hospital settings. A primary physician's role therefore takes place at the initial stage by making referrals to specialists. To study the physicians behavior in relation to cancer, we look at

[14] Long-acting muscarinic antagonists (LAMA) and Long-acting β2-agonists (LABA). See Appendix Table B.2 for the used ATC codes.

[15] Avoidable hospitalizations can be avoided with appropriate care in the primary care sector. Avoidable hospitalizations are commonly used to assess physician performance and physician-patient relationship (see e.g. (70)).

[16] Metformin is first-line pharmacotherapy to treat people with type 2 diabetes since the 1950s. Annual diabetes checkups are only recorded in the years 2006-2011 and regressions using this outcome therefore contains fewer observations than the other outcomes.

patients use of these specialist services. We focus on the referrals related to detection of breast and lung cancer, (1) radiology for breast cancer and (2) thorax scans (x-rays and CT-scans) to detect lung-cancer. We focus on the first time that patients receive these examinations.

2.1.2 Descriptive Statistics

Table 2.2 shows summary statistics on the clinic and physician level. We have a total of 3,137 clinics and 9,096 physicians in our sample. We see large differences across closing and non-closing clinics. Clinics that close in our sample are more likely to be solo clinics and are older on average. Because the physicians working in the closing clinics on average are older, we are unable to their SES in most cases. From Appendix Figure B.1, we see that physicians are advantageously selected. Compared to the average population, physicians are are less likely to have a parent with primary school education. Of the physicians we can observe their SES, 25% of these is defined as low-SES, and if we aggregate SES on the clinic level, we find that 28% of clinic employ a physician defined a low-SES. There are differences between high- and low-SES physicians as displayed in the last two columns of Table 2.2. Low-SES physicians are more likely to be female, they are slightly older, and less likely to have a degree from the University of Copenhagen (UCPH). The fact that low-SES physicians are older are likely related to the increase in the average level of education over the past decades. From Appendix Figure B.1, we see a clear decline in the proportion that a parent with primary school education for both the overall population and physicians. [17]

Table 2.3 shows summary statistics on the Danish population between ages 30-70, our analysis sample, and our analysis sample by SES of the patient. The patients who experience a clinic closure in the period of interest are older and more ethnic Danish. This selection is caused by the fact that clinic closures are more concentrated in rural parts of the country. Patients in

[17] As a robustness check, we use physicians parents' educational rank in the whole adult population to measure their SES. Figure B.1 shows that physicians' parents' educational rank is fairly stable across the period.

our analysis sample are similar to the total population in terms of SES and gender. Patients that are low-SES are more likely to be female, older, ethnic Danish and less likely to be married.[18] From the table, we also see that high- and low-SES patients are equally likely to have a low-SES physician.

Socio-economic inequality in health

While Denmark has equal access to health care and education, we still observe a large inequality in health. Figure 2.1 shows mortality rates by patient education and physician SES in the full population adjusted for age, gender, and year fixed effects. The figure shows that patients with primary school as their highest level of schooling have the highest probability of dying. The gradient is nonlinear in education: the decline in mortality of going from primary school to vocational education is larger than from vocational education to college education, and from college to university.[19] On average, 0.75% of patients with primary school die in a given year, while the same is true for 0.51, 0.39 and 0.33% for those with a high school degree, college or a university degree (average in higher than primary school groups: 0.48). In sum, the patients with primary school as their highest level of completed education are (0.75-0.48/0.48*100=) 56% more likely to die in given year. From the figure we also see that low-SES patients who are assigned with a low-SES physician have a lower mortality rate compared to low-SES patients who are assigned to a high-SES physician. The gap in mortality between low-SES and high-SES patients is 12% lower when low-SES patients are matched with a low-SES physician after adjustments in the overall population. This figure suggests that studying the population whose highest level of education is primary school is the most worthwhile and that the SES of the physician can impact low-SES patients' health.

Figure 2.2 summarizes the health-SES gradients across outcomes in the full population.

[18] Immigrants' levels of education are coded differently from non-immigrants, resulting in some missing values. Our results are robust to excluding non-ethnic Danish patients.

[19] Primary school, vocational education and college education corresponds to finishing ninth grade, upper secondary education, and short cycle tertiary degree in the United States, respectively.

A positive value means that patients with low SES have higher utilization or experience the incidence at a higher rate. The gradient for death related to COPD is more than 150%; death related to CVC has a 58% gradient. The difference in mortality reflect that low-SES patients have worse health, possibly with more chronic conditions. We see positive gradients in most of the outcomes related to health behaviors: low-SES patients are more likely to visit their physician in a given year, they have more visits, have more services per visit. This difference might reflect that individuals with low SES are more likely to have chronic conditions and co-morbidities and thereby need consultation with their GP more often. However, we also see that despite worse overall health, low-SES patients are less likely to be in contact with a medical specialist. We are unable to address whether certain care is over- or under-treatment. Even if some care provided is excess treatment, low-SES patients still have worse health, as shown by the mortality gradient.

2.2 Identification Strategy

An ideal experiment to study our research question would be to separate a representative group of patients from their existing physicians and randomly assign them to physicians of different SES. To mimic such an experiment, we use clinic closures as they are plausibly exogenous to patients' health trajectories and exploit the variation from the re-assignment of patients to physicians after these closures. We use this setup in a triple-differences design. The first difference compares outcomes of interest for low-SES patients before and after they join a low-SES clinic. Since this difference includes a discontinuity-of-care effect from the separation of patients from their initial physicians, we use low-SES patients who join high-SES clinics in the post period as a control group; this creates our second difference. Since there are potential systematic differences between high- and low-SES physicians, we introduce a second control group, consisting of high-SES patients who either are matched with a high- or low-SES physician post clinic closure. This gives us the third difference.

We highlight that our design mimics a randomized experiment as closely as possible. First, because we are interested in the adult population, an ideal experiment would need to separate patients from their existing physician, creating a similar discontinuity of care. Second, due to the practical importance and government mandate of having primary care close to patients' residence, combined with limited clinic availability of open clinics, it is difficult for an experiment to assign patients to physical clinics randomly.

Although the separation is plausibly exogenous, there remains concern that selection exists in the formation of new physician-patient pairs. To address this concern, we first test for selection. While we find that patients of the same on gender, age, and ethnicity are more likely to be matched, we find no evidence that this is true for SES, see Table 2.4.[20] The reason could be that physicians' SES is not observed by the patients, therefore the patient is unable to select a new physician based on this characteristic. We further address the concern of endogenous selection by employing a trajectory fixed effect in our triple difference identification strategy. Trajectory fixed effects refers to taking fixed effects on the pre-post closure physician interaction. The triple interaction coefficient therefore compares high- and low-SES patients who had the same pre-closure physician and post-closure physician. This strategy not only accounts for the fact that there might be selection of the post-closure physician, but also for the fact that low-SES physicians are different from high-SES physicians on several dimensions, see Table 2.2.[21]

[20]We do not find patient gender, age, ethnicity, or previous diagnosed chronic conditions to matter for the physician they are matched with, after controlling for physician characteristics (gender, age, ethnicity, clinic type), see Appendix Tables B.3 and B.4. In addition, we see no signs of either the treatment or control groups selecting into physicians that graduated from a particular institution, potentially because this information is not readily available upon choosing a physician. Ideally, we would also investigate the effect on health from gender, age, and ethnicity concordance between physician and patients. However, selection along these dimensions makes a causal analysis infeasible using our design.

[21]A second concern regarding identification is potential non-random assignment of physician SES to other physician characteristics, if a particular group of patients benefit more from a certain physician characteristic. We test if observable physician characteristics contribute to changes in patient mortality and health behaviors in Section 2.3.

2.2.1 Estimation Equation

The triple-difference estimate of having patient-physician SES concordance is estimated in the following equation,

$$
\begin{aligned}
y_{ijt} = \tau \times post_{it} \times SES_j^p \times SES_i + \alpha \times post_{it} \times SES_j^p + \rho \times post_{it} \times SES_i \\
+ \delta \times SES_j^p \times SES_i + \iota SES_i + \sigma \times Post_{it} + \gamma(PCP_i) + x_{it}^p \beta + \varepsilon_{ijt}
\end{aligned}
$$

where y_{ijt} is a measure of health or health care utilization for patient i, who get physician j at time t. SES_i is an indicator that takes the value one if the patient is defined as low SES. The variable $Post_{it}$ takes the value one in post-closure years and zero in the years before the clinic closure. We include four years prior to and three years after the clinic closure.[22] x_{it}^p includes patient-specific characteristics, such as age, gender, ethnicity. SES_j^p takes the values one if the new physician after a clinic closure is from a family with a low SES and zero otherwise. We hold this constant even if the patient change physician in the post period. PCP_i is trajectory fixed effects, taking fixed effect on the pre-post physician level. Most of our outcome variables are indicators. In these cases, we use a linear probability model to estimate the parameters. We cluster standard errors by patient ID.

The triple interaction term, $post_{it} \times SES_j^p \times SES_i$, indicates the difference in health or health care utilization between high- and low-SES patients who get a physician from a low-SES family following a clinic closure, compared to the same difference for patients who get a physician from a high-SES family following a closure. τ is the estimate we use to calculate the gradient.

We define treatment using only post-closure physician SES. This treatment definition means that we assume that all closing clinics are of the same SES, and that SES concordance with the previous physician does not have dynamic lasting effects. We make an implicit assumption that all closing clinics are high-SES. Making this assumption gives us a *reduced form* estimate

[22]We use three years after clinic closures, as our event study design shows that the effect disappears in later periods.

of the effect of SES concordance. To produce *treatment on the treated*, our result should be weighed by the fraction of patients that have a high-SES physician in the pre-period and are reallocated to a low-SES physician in the post period. The treatment on the treated estimate should be numerically close to the reduced form estimate that we present here, since we expect that most physicians in the closing clinics are indeed high-SES, as discussed in Section 2.1.1.

2.2.2 Identifying Assumption and Validity

The key identifying assumption in our empirical design is the parallel trend assumption. The design requires that patients' underlying trends in health care utilization and health does not systematically differ by the SES of the physician they get after the clinic closure. To test for parallel trends, we present graphical evidence by examining how outcomes of interest change in years around clinic closures. We do so by employing a dynamic double difference strategy for high- and low-SES patients separately. The estimating equation is

$$y_{ijt} = \sum_{r \neq -1}^{r=5} \theta \times I_r + \sum_{r \neq -1}^{r=5} \theta \times I_r \times SES_j^p + x_{it}^p \beta + x_{jt}^d \phi + \kappa(GP_i^{-1}) + \varepsilon_{ijt} \tag{2.1}$$

where I_r is an indicator that takes the values 1 in period r. GP^{-1} is previous physician fixed effects, and x^d is the new physician controls including age, gender, ethnicity, and graduating institution.[23]

2.3 Effects of physician-patient SES concordance

This section presents results on how physician-patient SES concordance affects patient health and health behaviors. We first look at how SES concordance affects all-cause mortality.

[23] In this equation, we are not able to account for trajectory fixed effects as the equation is estimated separately for high- and low-SES patients.

To investigate the origin of the concordance effect, we first break down mortality by causes, focusing on deaths related to chronic conditions. We then study potential pathways that physician-patient interaction could affect mortality, by looking at general patient health behaviors and behaviors specific to chronic conditions. Lastly, we present suggestive evidence on potential mechanisms and study threats to the validity of our results.

2.3.1 SES Concordance Effects on Mortality

We begin by presenting the results on all-cause mortality in an event study design following equation 2.1. Figure 2.3 shows coefficients from two separate dynamic difference-in-differences regressions. The x-axis denotes years since clinic closure and the y-axis shows the difference in effect depending on the patient's new physician's SES. The solid line shows the treatment effect of low-SES patients who have a low-SES physician after experiencing clinic closures, relative to low-SES patients who have high-SES physicians after closures. And the dashed line show the same effect for high-SES patients.

All patients are alive at the event of a clinic closure to identify the new physician's SES, therefore the mortality estimates in the pre-periods are zero by design. In order to test for the parallel trends assumption in mortality, we use deaths that take place between years -4 and 0 in the closing clinics.[24] We define treatment and control to deceased patients at the clinic level using the physician re-assignment of the patients that are alive when clinic closures take place. We assume the passing patients would have matched with a low-SES physician if more than 50% of the patients who are alive are matched with a low-SES physician. We show in Figure **??** shows that treatment and control groups are on the same mortality trajectory prior to clinic closures.

Since year 0 is the year of clinic closure, year 1 is the first year that all patients are re-assigned to a new physician. We see that mortality immediately decreases by 0.9 percent-

[24] We use pre-trends to test for the parallel trend assumption on non-mortality outcomes by interacting the event time with treatment dummies, as shown in Figures 2.5 and 2.4.

age points for low-SES patients in the first year they are matched with a low-SES primary care physician, relative to low-SES patients that switch to a high-SES physician. Meanwhile, mortality rates for high-SES patients does not depend on their physician's SES, as shown in the dashed line. We focus on the *relative* change in mortality between the solid and dashed lines in the triple differences design.

From Figure 2.3, we see that the effect disappears four years after clinic closures. The disappearing of the effect could be caused by high-SES physicians and low-patients becoming familiar with each other over time, which could improve medical communication; it could also be that SES concordance delays, rather than removes, deaths; or that patients voluntarily leave a physician if they are unsatisfied with the initial assignment. [25]

Table 2.5 shows the triple differences estimates using mortality as the outcome using varying controls. Our estimate of interest, the coefficient for the triple-interaction term, is robust to controlling for patients characteristics, old physicians fixed effects, individual fixed effects, and trajectory fixed effects. The triple differences estimate indicates that the treatment group (low-SES patients matched with low-SES physicians in the post period) experience a 0.134 percentage point decrease in the probability of dying, relative to comparison groups. Our preferred specification is shown in column 5 and uses the triple differences design described in equation 3.3.

For ease of interpretation and to translate this effect into changes in the SES-mortality gradient, we compare this triple differences estimate to the mortality gradient for patients of high-SES physicians post-closure. Table 2.5 column 5 shows that the mortality gradient between high- and low-SES patients of high-SES physicians is 0.54 percentage points. This indicates that, in the post-period, the SES-mortality gradient for patients at low-SES physicians decreases from

[25]The dynamic effect may come from a combination of these explanations. Reassuringly, most of the disappearing of the effect is caused by low-SES patients who have low-SES physicians returning to the mortality level of the group of low-SES patients who have high-SES physicians, see Appendix Figure B.2. Deaths caused by CVC and cancer among the treatment group is lower than the average between years 0 and 3, and higher in year 4, indicating a delay in mortality. The next section discusses mortality by causes in detail.

0.54 to 0.41 by 24.8% in our preferred specification.[26]

Next, we examine the effect on mortality by sub-populations to study which groups are the most susceptible to the SES concordance effect in Table B.7. The observed effect on mortality is the most pronounced for men and unmarried individuals. As shown in the first two columns of Table B.7, SES-mortality gradient drops by 28.7% for men while the effect is 16.8% for women. The effect size is similar for the older and younger birth cohorts, while the effect is entirely driven by the ethnic Danish sample.

The role of chronic conditions in mortality effects

What drives the significant decline in mortality when low-SES patients have low-SES primary care physicians? To investigate the underlying mechanisms, we breakdown mortality by cause and focus on deaths that are caused by the four most common and unequally distributed somatic chronic conditions: cardiovascular conditions (CVC), cancer, diabetes, and chronic obstructive pulmonary disease (COPD).

Table 2.6 column 1 shows that in the 3 years following clinic closure, the treatment group experiences 0.043 percentage points lower probability of dying from CVC than the comparison groups. Comparing the triple difference result to the baseline, we find that SES concordance lowers this CVC-mortality gradient by 42.6%. This effect is almost twice the size of the point estimate for overall mortality, suggesting that the reduction in deaths due to CVC account for an substantial part of the overall mortality-gradient reduction. Given the acute nature of CVC deaths, results on CVC mortality also aligns with the fact that we observe overall deaths to drop immediately after clinic closures. From column 2, we also see a decline in cancer mortality in the first three years following clinic closure, which reduces the SES-gradient by 24%. In Table 2.6 columns 3 and 4 report the effect of SES concordance on mortality related to diabetes and COPD. We do not find any overall significant effect on these outcomes.

[26]This is calculated by (0.134/0.541)*100.

Table B.8 reports the triple differences estimates on the effect on different causes of death by gender. We see that the SES-gradient in CVC mortality is largest for men, reducing the SES-gradient on overall CVC mortality for men by 51.5%, while we do not find a significant effect on CVC mortality for women. However, we find that SES concordance reduce the SES-gradient in cancer mortality by 28.6% for women. [27] Appendix Figure **??** shows the event study figures for CVC mortality for men and cancer mortality for women. We see that men's CVC mortality and women's cancer mortality declines in the first three years after clinic closure, and returns and exceeds the control group level of mortality in year four. This pattern suggest that the SES concordance between physician and patient delay, rather than, avoid mortality, potentially because of earlier detection of cancer, we elaborate on this in the next section. Table B.9 shows that the effect on CVC mortality is mainly driven by the older sample. Low-SES patients born before 1958 have a significantly reduced mortality related to CVC (50.4% reduction in the gradient) and cancer (42.6%). While we observe a significantly reduced mortality-gradient in the younger sample in COPD-related deaths.

2.3.2 SES Concordance Effects on Health behaviors

We have shown that SES concordance decrease the SES gradient in mortality and that this is driven by reduced or delayed deaths caused by CVC and cancer. Next, we study potential pathways in which the concordance can reduce mortality. We look at patient health behaviors related to (1) general healthcare utilization and (2) behaviors related to chronic conditions.

We first present results on general healthcare utilization at the extensive and intensive margins. On the extensive margin, we study if SES concordance induce patients to make *any* visit to their primary care physician. Table 2.7 column 1 shows that patients in the treatment and control groups are equally likely to make at least one visit to their new physicians within the first

[27] Deaths by individual types of cancer do not show significant affects; the largest point estimates for lung-cancer, see Appendix Table B.10

three years of clinic closure. The fact that we do not observe any effect on the extensive margin could be because a relatively high baseline office visit rate, as 84% of the Danish population make at least one office visit per year. On the intensive margin, we study the number of visits per year, mean number of services per visit, and how these effects translate to health spending on fee-for-services physician reimbursement. Figure 2.4 shows the event study graphs for number of visits and total primary care physician fee-for-services reimbursement. We see an increase in number of visits, and total reimbursements for low-SES patients when there is SES concordance, while we see little or no effect across clinic closure for high-SES patients. Increased contact with the physician may originate from the need for more care due to more detection of diseases or better adherence of treatment guidelines (see section 2.3.2); it may also be the contributing factor to the increased detection of diseases. It could be that patients are more health-aware or feel more comfortable with the physician, and schedule more visits given the same health condition. [28] Importantly, the estimates for four years prior to closure suggest that patients in treatment and control groups are on similar health trajectories.

The triple differences results are shown in Table 2.7 columns 2-3. Low-SES patients make more visits to their physician at baseline, and SES concordance increases number of visits by 0.124 per year. This corresponds to an 8.5% *increase* of the baseline gradient.[29]

We also see that the treatment group receive more services per visit in the 3 years post closure in Table 2.7 Column 3. SES concordance for low-SES patients thus *increases* this gradient by 22.6%. The increase in the number of visits per year and number of services per visit are both far greater for the older patients (born before 1958), as shown in Table B.13. Lastly, the increase in number of visits per year and number of services per visit translate to increased fee-for-services physician reimbursement by a total of 2.7 dollars per year, as shown in Appendix

[28] An alternative explanation is that the quality of each visit is lower, leading to more visits. However, considering the decline in mortality, this seem like an unlikely explanation.

[29] In Table 2.7, we see an increase in the likelihood of having an office visit, the number of visits, and services per visit in the post period for both treatment and control groups, see the coefficient for the variable *Post* . This is consistent with (89), that finds increased use of care after clinic closures.

Table B.11. The increase in spending is driven by women, which could be caused by a large increase in number of services per visits in the female sample (see Appendix Tables B.14 and B.12). We also see an increase in spending to medical specialists, and an increase in references to specialized care for men, which reduces the SES gradient in medical specialist visits by 34.3%.

Health behaviors related to chronic conditions

In order to gain further insights on how SES concordance affects chronic condition deaths, we look at health behavior related to these conditions. While we are unable to see physicians' diagnosis of the chronic conditions, we can observe how they treat patients. The outcomes of interest for each condition is defined in Section 2.1.1. In sum, we find that SES concordance leads to higher detection of chronic conditions and better adherence to treatment guidelines of chronic conditions, which may an important factor in explaining the reducing in SES-gradient in mortality.

Cardiovascular Conditions (CVC) Since we find a large effect on CVC mortality, we look at how use of first-line prescription medications, statins and ACE inhibitors, respond to SES concordance. In Figure 2.5 Panel B, we see that statins prescriptions immediately increases after low-SES patients get matched with low-SES physicians post clinic closure, while we see no such effect for high-SES patients. Pre-closure estimates display parallel pre-trends. Triple-differences results in Table 2.8 column 1 shows that SES concordance increases statins prescriptions by 0.286 percentage points while we do not find any effect on ACE use. Since low-SES patients are more likely to be prescribed statins at baseline, concordance increases the SES-gradient by 6.2%. This effect is driven by men (see Appendix Table B.16). Combined with the decrease in CVC mortality, the results suggest that low-SES patients are under-diagnosed or under-treated for CVC at baseline and that the higher use of statins prevent or delay deaths in the years after clinic closure. In line with the effect on CVC morality disappearing in year 4, we see in Appendix that men's use of statins also disappears in year 4.

Cancer We look at the first time that patients are tested for breast and lung cancer, since primary care physicians play a crucial role in the decision to test for these types of cancer but treatment of cancer takes place at hospitals. While we do not observe statistically significant effects for these outcomes for the population overall, we find that the older cohort is more likely to receive lung cancer examinations when they have a low-SES physician after clinic closure, as shown in Table B.17 column 3. The result on first lung cancer examinations provides suggestive evidence that the effect may go through earlier detection of lung cancer. Since lung cancer has a low survival rate, early detection is especially important.

Chronic obstructive pulmonary disease (COPD) The variables of interest related to COPD includes both medication and avoidable hospitalization due to COPD. We do not find COPD medication prescription to respond to the treatment. However, we observe a stark reduction in avoidable COPD hospitalizations, as shown in Figure 2.5 Panel A. Our preferred triple differences estimate in Table 2.8 column 4 shows that SES concordance reduces the SES-gradient in COPD avoidable hospitalizations by 14.2% on a basis of 0.866 percentage points Although COPD is more common for women, we find that the that the reduction in the hospitalizations related to COPD is driven by men, with a 25.3% reduction in the SES-gradient (see Table B.16 column 4). When we breakdown the sample by birth cohorts, we show that the older patients drive the decrease in COPD hospitalization, while we see no effect for the younger sample (See Table B.17).

Diabetes Following treatment guidelines of diabetes, we study how metformin prescriptions and annual diabetes check ups respond to SES concordance. Figure 2.5 shows that the low-SES patients with low-SES physicians experience a drastic and persistent increase the number of diabetes checkup visits (2.5 percentage points in year 1), while the high-SES patients with low-SES physicians also experience an increase in the first year (1.7 percentage points). Triple differences results in Table 2.8 column 6 shows that SES concordance *increases* the SES gradient by 46.3% from the baseline gradient of 2.43 percentage points. Since diabetes is a cause for CVC, better management of diabetes could explain some of the reduction in CVC mortality. While we see both genders increase the number of diabetes checkup visits, the older sample sees a greater increase compared to the younger one.

Adherence and detection effects Primary care physicians play a key role in the detection of chronic conditions, and patients' adherence to medical guidelines (56, 71, 46). We break down effects on health behaviors into a detection effect and an adherence effect. Adherence refers to patients with a pre-exisiting condition continuing the respective treatment. Detection refers to patient that are not previously treated for a condition, thereby likely newly diagnosed, starting treatment. We group patients by whether they received treatment before they experienced a clinic closure ("new patients" and "pre-existing patients"), and study their use of a treatment or mortality when they are matched with a low-SES physician.

Table B.18 shows the effect of SES concordance on adherence and detection. The first two columns show that overall mortality decreases for both new and pre-exisitng patients with any one of the four chronic conditions we address above. Columns 3 and 4 show that SES concordance affects CVC mortality mainly through a detection effect.[30] Columns 5 and 6 show that the use of statins only increases for patients who are prescribed statins before clinic closures, suggesting that SES concordance increase low-SES patients adherence to statins. The fact that we do not find increase in statins prescription, but a large reduction on CVC mortality among new patients, suggests that our data only captures part of the mechanism that prevent CVC deaths. Patients' changes in health behaviors outside of the clinic, such as smoke cessation, exercising, and dieting, may also contribute to the reduced SES gradient in CVC mortality.

[30]We do see a large estimate for pre-existing patients, although it is not statistical significant.

In contrast to treatment for CVC, both adherence and detection effects exist for diabetes patients. This could be because CVC is more likely to be under-diagnosed compared to diabetes, by the nature of the disease (39).

2.3.3 Mechanisms

In this section, we first study three proposed mechanism of concordance effect: trust, physician cognizance of patient health risks, and quality of medical communication. While direct tests for these hypotheses are not feasible in our data, we present suggestive evidence in this section. Other mechanisms than the ones mentioned here might be at play aswell. Next, we conduct two extension analysis related to the internal and external validity of our results. We address an identification concern regarding the potential non-random assignment of physician SES and other characteristics. We then investigate the concordance effect can be generalized to other patient-patients groups.

Trust and rapport First, we investigate whether being matched with a low-SES physicians makes low-SES patients establish better trust and partnership with their physician. Trust and partnership is essential in the primary care physician setting (44). Since better partnership might result in longer lasting relationships between the two, we look at patient churn to study this channel. We find no evidence that SES concordance increase the length of the physician-patient relationship in the post period.

Physician cognizance of low-SES patient health risks Low-SES patients experience more chronic conditions, therefore physicians from low-SES families might gain personal familiarity outside of professional settings with these conditions. To study the channel of physician cognizance of low-SES patient health risks, we investigate if physicians who have parents' with chronic conditions reduce the SES gradient in mortality.[31] We define the physician to have personal experience with these conditions if a parent had died from one of the conditions or have received treatment for one of more of the conditions.[32] Appendix Table B.22 shows that physicians cognizance of patient health risks is a relevant channel. From the table, we see that physicians who have a parent who has been treated or died from a chronic condition, reduce the SES-gradient in SES gradient in all-cause mortality for their patients three years after clinic closure. This pattern is also evident when looking personal experience with with CVC or cancer and mortality related to these conditions.[33]

Improvement of communication A third potential channel is that SES concordance improves communication quality between physicians and patients, allowing physicians to make relevant information more salient and increasing health literacy. The medical literature uses adherence rates and avoidable hospitalizations to proxy for patient-physician communication quality (see e.g. (44, 107, 70)). We find increased adherence to medical guidelines on diabetic check ups and statins prescription, as well as decreased aviodable hospitalizations with COPD. These results suggest improved medical communication quality is a relevant channel.

Internal Validity - the role of other physician characteristics

A threat to the internal validity in our research design is that physicians' SES is correlated with other physician characteristics. As shown in Table 2.2, low-SES physicians are on average older, more likely to be female, and less likely to have a degree from the University of Copenhagen. Could any of these factors be the driving our estimates? For instance, do low-SES

[31]We look at the four chronic conditions of interest: CVC, cancer, diabetes, COPD.

[32]We do not look at the timing of the parents diagnosis or death, and the physician is defined as being exposed to these conditions if a parent has died of a certain condition at some point in the analysis period. The reason behind this assumption is that, we would expect parent to be affect by a certain condition before the time of death. For example, patients that pass away due to CVC likely would have had an increased risk before the most adverse incident.

[33]The effect on CVC mortality is only marginally significant on the 13% level.

patients benefit more from having a more experienced physician relative to a less experienced one? We conduct a similar analysis to our main specification, but replace the treatment dummy with a dummy for the another physician characteristic. Appendix Table B.19 shows that matching the most experienced clinics with low-SES patients does not reduce the SES-gradient in mortality. Neither does matching patients with clinics that have more male, more UCPH-trained, or more ethnic Danish physicians.

We also examine whether SES-gradient in health and healthcare utilization decreases if low-SES patients are matched with physicians of the best quality. In other words, can we substitute low-SES physicians' knowledge about low-SES patients with high physician quality? Since physician quality is hard to measure, we proxy for physicians' quality using their academic performance (GPA) upon entering medical school.[34]

We define physicians as "high quality" if their grades are among the top 30% in the whole physician population. We conduct a similar analysis to our main specification, but replace the treatment dummy with a dummy for high quality physician.[35] Table B.21 show that physicians of higher academic performance does not affect patient outcomes differently compared to physicians with lower academic performance. This suggests that higher quality physicians can not substitute low-SES physicians.

The above suggest that observed physician characteristics including gender, experience, ethnicity, graduating institution, and physician academic performance do not explain our findings.

External validity - generalization of the concordance effects

The share of Danish population that have primary school as their highest level of education has decreased overtime, as shown in Appendix Figure B.1 panel C. By using primary school

[34] High school GPA is observable for the youngest physicians in the sample, graduating high school after 1985 and after. We observe physician GPA for around 25 percent of the sample.

[35] We aggregate physician school grades to the clinic level.

education as the definition for SES, we label someone who is primary-school-educated who is born in 1940 the same as someone born in 1970, while his/her educational "rank" is much lower for the one born in the later years. Appendix Figure B.1 panel B shows that the physicians' parents' educational rank have been relatively stable across the study period. In this section, we investigate if our results are robust to an alternative definition of low-SES and if our results can be generalized to other educational groups.

First, we test the robustness of our results using educational rank within a birth cohort to define SES. We define physicians to be low-SES if they have a parent who is among the bottom 30 percent educated in their birth cohort. We keep primary school as the definition of a patient being low-SES. Appendix Table B.23 shows that the estimates are robust with this definition. By substituting the *level* of parental education by the *rank* of parental education, we show that exposure to physicians from low-SES households continues to be important even as the share of physicians with primary-school-educated parents decreases.

Next, we test whether our results on SES concordance can be generalized to patient populations with higher levels of education. For instance: for patients who have vocational school as their highest level of education, would their health outcomes improve when they are matched with a physician who has a parent with vocational school as their highest level of education? To assess whether our results apply more broadly, we perform the same analysis following equation 3.3, but change our definition of low-SES to vocational school and college education. As shown in Table **??** column 2 and 3, we do not find that educational concordance improves the health of groups of patients. This aligns with our findings in event studies, such as Figure 2.3, in which we do not see high-SES patients' mortality to depend on their physicians' SES. Baseline mortality-SES gradient by patient education levels in Figure 2.1 also shows that primary-school-educated patients show the largest gap in mortality. The most disadvantaged population may be most susceptible to the physician they are match with due to worse health at baseline, while better-educated patients do not benefit from SES concordance as much because they are healthier with better health literacy.

2.3.4 Robustness Checks

In this section, we discuss robustness checks in relations to data limitations we face and the assumptions we made in addressing those limitations.

Addressing potential selection using primary care shortage Our preferred specification in section 2.2 addresses physician re-assignment selection concerns by restricting the treatment and comparison groups to have the same physicians before and after clinic closures. We do this by taking "trajectory" fixed effects, namely, fixed effect on the pre and post closure physician interaction. An alternative to this "with-in" group comparison is to make use of the primary care shortage in Denmark. Over the last 10 years, the number of physicians in Denmark has decreased by 7 percent, while the number of citizens, old people, and individuals with chronic diseases have increased (75). This resulted in a critical shortage of physicians where most clinics do not accept new patients. In 2017, 67 percent of all clinics had closed intake of new patients. The number of clinics that has closed intake of new patients varies substantially between municipalities: certain areas have no clinics that accept new patients (74). When clinic closures take place in a municipality and year with extreme primary care shortage, the choice of a new physician is extremely limited. Clinics would only accept a new patient when a patient moves to another municipality or passes away.

We run our main analysis using a subsample of patients who experience a clinic closure in municipalities and years with extreme primary physicians shortage. We define primary care shortage to occur in municipalities and years where the average patient per clinic exceeds 1600.[36] Closures in 458 clinics containing more than one million patients in our analysis sample satisfy this criteria. We use this sub-sample of patients to conduct the analysis on our main outcomes

[36] Physicians can close their intake of patients when the number of patients exceeds 1600.

from section 2.3. Appendix Table B.25 shows that our results are robust to using this sub-sample of patients that have a limited choice of a new physician.

Alternative aggregations of Physician SES Claims data from Denmark allows us to connect each patient to the primary care clinic, rather than a specific physician within the clinic. The average clinic has 1.8 physicians. In this section, we present versions of our analysis by aggregating physician SES to clinic SES in two alternative ways. When a clinic has more than one physician, we construct three variables, min, max, and mean of the physicians' SES.

In the main analysis, we defined a patient being matched with a low-SES physician if at least one of the physicians in the corresponding clinic was defined as low-SES (using a "max" function). In this case, there is a positive probability that the patient sees a physician with a low educational family background. As robustness checks, we repeat our analysis for our main outcomes defining physician SES using the "min" and "mean" functions. The min function takes the value 1 if we define all physicians in the clinic as being low-SES. In this case, we are certain that the patient sees a low-SES physician. We also use the "mean" function; this gives us the share of physicians from a low-SES family and measures the probability that the patient sees a physician with a low educational background. As shown in Appendix Table B.26, the point estimates are robust to these alternative definitions.

Missing physician SES As described in Section 2.1.1, we are unable to identify the SES of a physician if he or she is born before 1960, and this applies to 37% of non-closing physicians. Based on the age of physicians with missing SES information, we assume that they are high-SES in our main analysis. As a robustness check, we complement the main analysis by discarding this assumption, and instead restrict our sample to physicians whose SES we can observe. Appendix Table B.27 Panel A shows our main results using this subsample and specification described in equation 3.3. In Table B.27 Panel B and C, we repeat this analysis, using the min and mean functions to aggregate physician SES to the clinic level, as described in section 2.3.4. The table shows that our results are robust to excluding observations with missing SES information.

Excluding Non-ethnic Danish patients A data limitation is that immigrants' education information is not always recorded. In the main analysis, we assume that immigrants with missing education are high-SES. For robustness, we exclude any non-Danish patient and repeat the main analysis in Appendix Table B.28 and show that most of our main outcomes are robust.

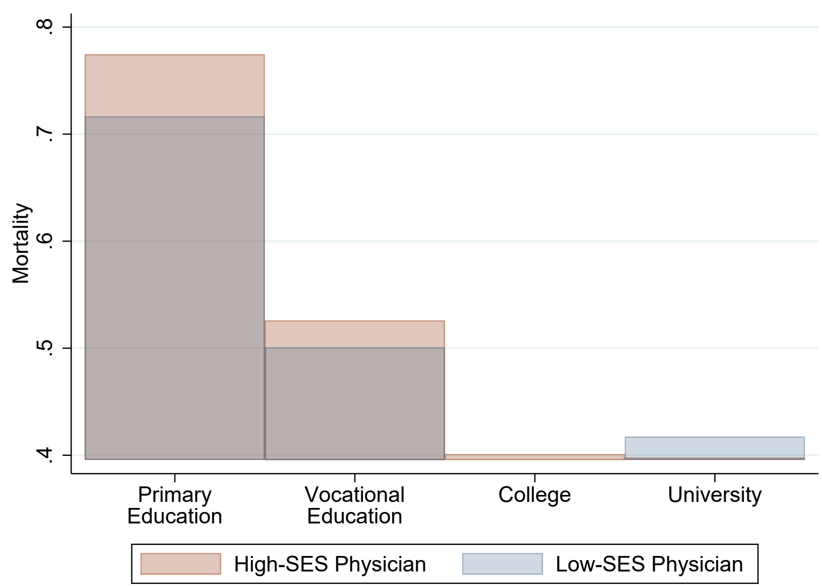
2.4 Conclusion

This paper studies the effect of physician-patient SES concordance on the socio-economic gradient in health. We exploit variation in SES concordance between physicians and patients that are induced by clinic closures and use physicians parents' highest level of education to measure their SES. We find that SES concordance lowers low-SES patients' mortality, while the mortality of high-SES patients are unaffected by the SES of their physician, leading to a reduction in the SES-gradient in health. Mortality effects are driven by a reduction in deaths caused by CVC. To study how concordance reduces patient mortality, we look at patients' health behaviors. We find that when low-SES patients are matched with low-SES physicians, they increase healthcare utilization on the intensive margin by having more offices visits and receiving more services per visit. In addition, SES concordance increases treatment of chronic conditions for low-SES patients; the effect comes from a higher disease detection rate of new patients, and a higher adherence rate for pre-existing patients.

This paper opens up several avenues for future research. Within health economics, studies that dig deeper into the mechanisms for how medical communication affects healthcare patterns would be especially valuable. Our paper also has implications beyond the medical setting; any transactions that involve coordination, especially the matching of pairs, such as teacher-student or manager-worker, may have matching quality as an input for their production functions.

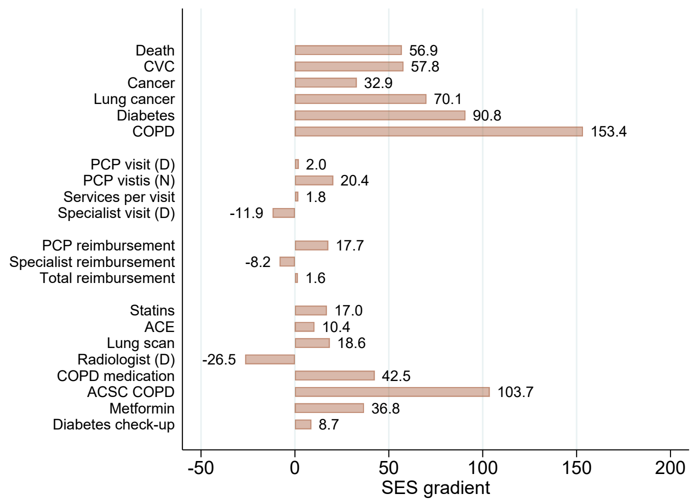
2.5 Acknowledgement

Chapter 2, in full, is currently being prepared for submission for publication of the material. Ida Lykke Kristiansen; Yanying Sheng. The dissertation author was a primary investigator and author of this paper.



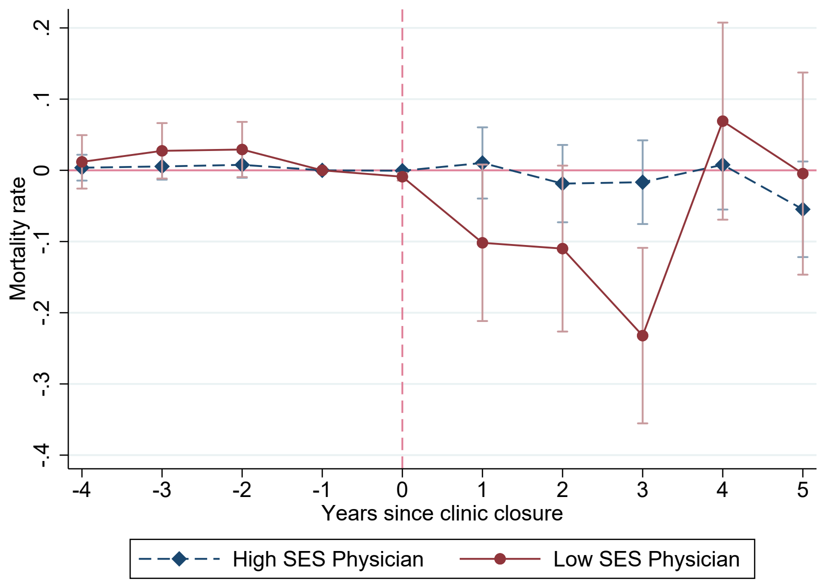
Note: The figure plots mortality rates of patients with high- and low-SES physicians by education levels in the full Danish adult population, adjusted for age, gender, and year fixed effects.

Figure 2.1: Physician SES and Mortality by Patient Education



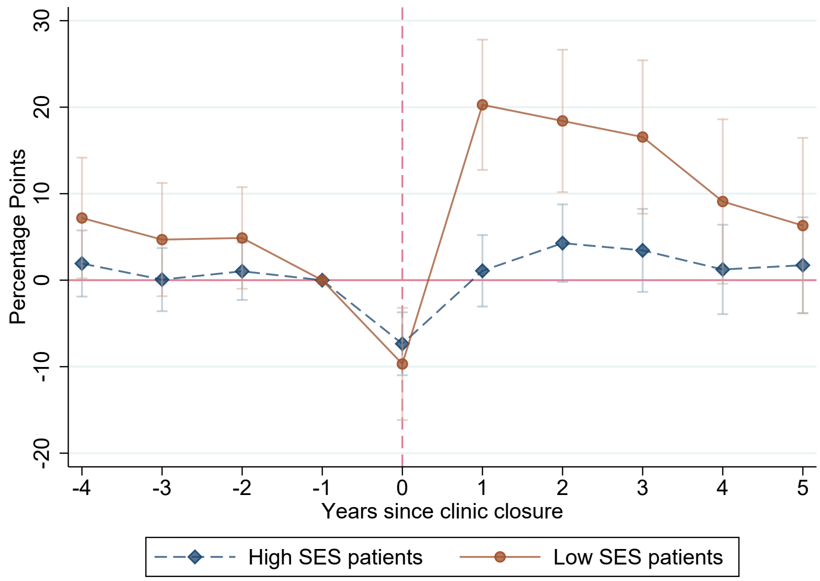
Note: The figure presents the Health-SES gradient by outcomes of interest at baseline. See Section 2.2 for definition of patient SES. The gradient is defined as the "excess" part of an outcome for low-SES patients relative to high-SES patients, weighed by the high-SES outcome. For example, the mortality gradient is calculated as $(lowSESmortality - highSESmortality)/(highSESmortality) \times 100$. The outcomes are adjusted for age, gender and year fixed effects. The following abbreviations are used: CVC stands for cardiovascular conditions, COPD stands for chronic obstructive pulmonary disease, PCP stands for primary care physician, D stands for dummy, N stands for numerical variable, ACSC stands for ambulatory care sensitive condition.

Figure 2.2: Health-SES Gradient by Outcomes of Interest

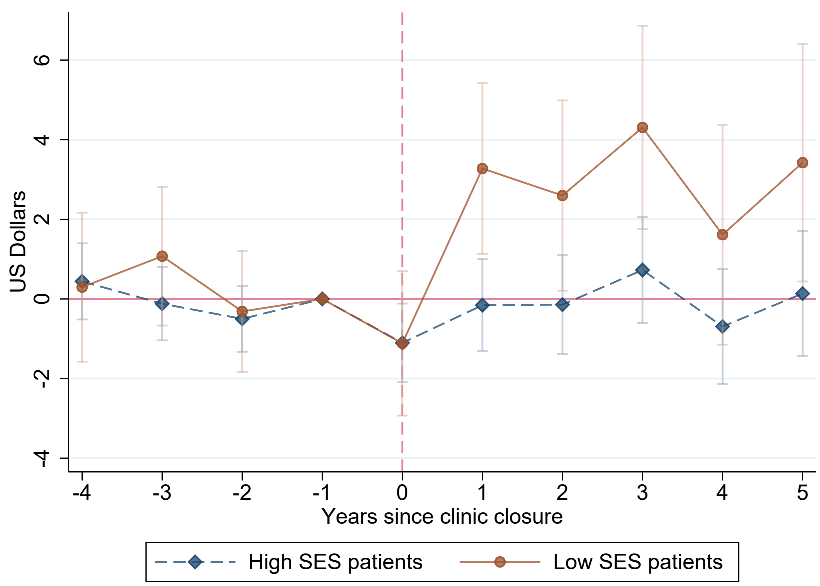


Note: The figure presents the effect of physician-patient SES concordance on mortality. For event periods on and after year zero, the solid (dashed) line plots the estimates and 95 percent confidence intervals of the coefficients on the event dummies in equation 2.1 using mortality as outcome for low-SEShigh-SES) patients. Treatment is defined as the patient being matched with a low-SES physician. For event periods smaller than zero, the solid (dashed) line plots the likelihood of dying on the clinic level. For patients dying in the pre-period, treatment is defined as 50% of patients in the same clinic are matched with a low-SES physician in the post period. Both regressions control for old physician fixed effects, year fixed effects, new physician characteristics (mean age, share of male physicians, share of ethnic danish physicians, solo clinic dummy, number of physicians in the clinic, and physicians' graduating institution), and patient characteristics (age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and level of education). The estimation results can be found in Appendix Tables B.5 and B.6. Standard errors are clustered by patient ID.

Figure 2.3: The Effect of Physician-patient SES Concordance on Mortality



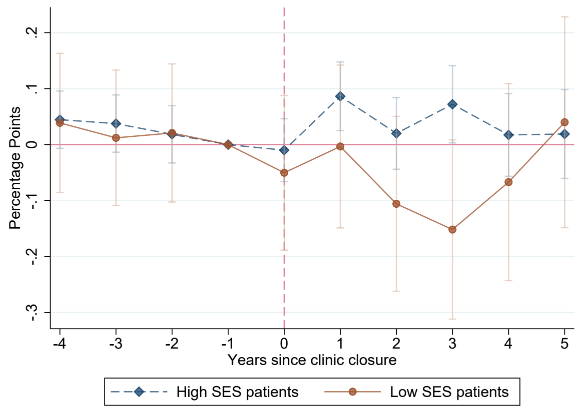
A: Number of office visits



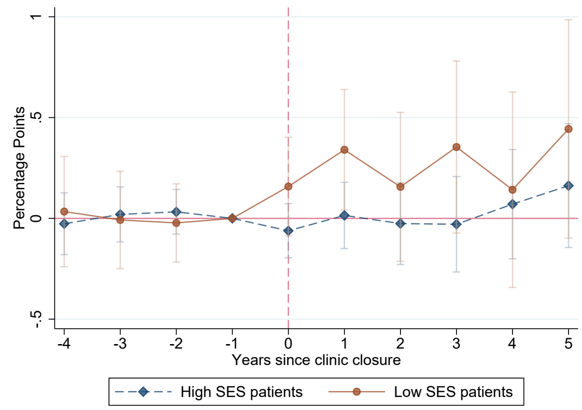
B: Physician services reimbursement

Note: The figure presents the effect of physician-patient SES concordance on the number of office visits and physician services reimbursement. See section 2.1.1 for the definitions of these outcomes. The solid (dashed) line plots the estimates and 95 percent confidence intervals of the coefficients on the event dummies in equation 2.1 for low-SEShigh-SES) patients. Treatment is defined as the patient being matched with a low-SES physician. Both regressions control for old physician fixed effects, year fixed effects, new physician characteristics (mean age, share of male physicians, share of ethnic danish physicians, solo clinic dummy, number of physicians in the clinic, and physicians' graduating institution), and patient characteristics (age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and level of education). The estimation results can be found in Appendix Tables B.5 and B.6. Standard errors are clustered by patient ID.

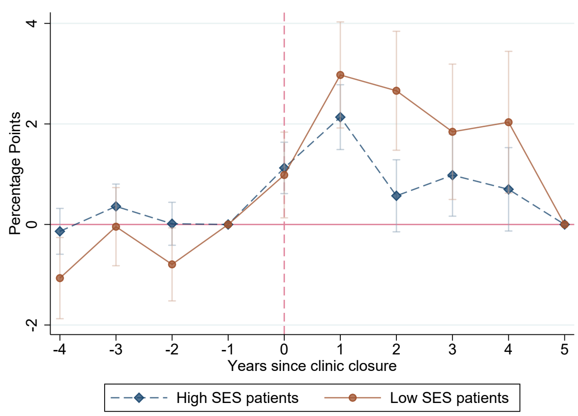
Figure 2.4: The Effect of Physician-patient SES Concordance on Health Behaviors



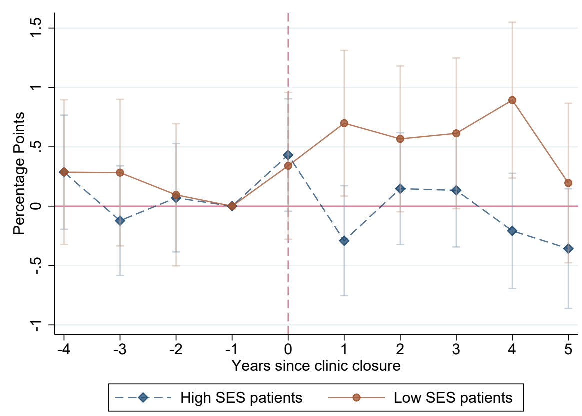
A: COPD avoidable hospitalization



B: Statins prescription



C: Diabetes check-up visits



D: Lung scan for older women

Notes: The figure presents the effect of physician-patient SES concordance on having a COPD avoidable hospitalization, being prescribed statins, and having a diabetes check-up visit. See section 2.1.1 for the definitions of these outcomes. The solid (dashed) line plots the estimates and 95 percent confidence intervals of the coefficients on the event dummies in equation 2.1 for low-SEShigh-SES) patients. Treatment is defined as the patient being matched with a low-SES physician. Both regressions control for old physician fixed effects, year fixed effects, new physician characteristics (mean age, share of male physicians, share of ethnic danish physicians, solo clinic dummy, number of physicians in the clinic, and physicians' graduating institution), and patient characteristics (age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and level of education). The estimation results can be found in Appendix Tables B.5 and B.6. Standard errors are clustered by patient ID.

Figure 2.5: The Effect of Physician-patient SES concordance on Health Behaviors Related to Chronic Conditions

Table 2.1: Summary Statistics - Clinics

Solo	0.611	0.501	0.948	0.487	0.419
Number of doctors in clinic	1.828	2.083	1.053	2.005	2.475
SES max	0.220	0.278	0.041	0.000	1.000
SES min	0.092	0.109	0.037	0.000	0.473
SES mean	0.142	0.176	0.039	0.000	0.690
Non-missing SES	0.501	0.637	0.085	1.000	1.000
Number of patients	2381.9	2701.2	1410.5	2539.4	3195.5

height

Notes: The table presents the summary statistics of clinics.

Table 2.2: Summary Statistics - Physicians and Clinics

	(1) All	(2) Non-closing sample	(3) Closing sample	(4) High-SES	(5) Low-SES
Physicians					
Male	0.531	0.495	0.681	0.370	0.325
Year of birth	1963.5	1966.3	1951.9	1975.9	1972.9
Born before 1960	0.427	0.343	0.782	0.000	0.000
Ethnic Danish	0.897	0.887	0.940	0.982	0.987
Low-SES	0.246	0.237	0.328	0.000	1.000
Non-missing SES	0.566	0.632	0.288	1.000	1.000
University of Copenhagen (UCPH)	0.523	0.505	0.599	0.527	0.424
University of Southern Denmark (SDU)	0.162	0.181	0.085	0.219	0.280
Aarhus University (AU)	0.280	0.279	0.283	0.251	0.287
Other University	0.035	0.035	0.034	0.003	0.009
Clinics					
Solo	0.611	0.501	0.948	0.487	0.419
Number of doctors in clinic	1.828	2.083	1.053	2.005	2.475
SES max	0.220	0.278	0.041	0.000	1.000
SES min	0.092	0.109	0.037	0.000	0.473
SES mean	0.142	0.176	0.039	0.000	0.690
Non-missing SES	0.501	0.637	0.085	1.000	1.000
Number of patients	2381.9	2701.2	1410.5	2539.4	3195.5
Number of physicians	9,096	7,352	1,744	3,212	794
Number of clinics	3,137	2,361	776	682	518

Notes: The table presents physician and clinic characteristics. Physicians are low-SES if one of their parents has primary school as his/her highest level of education. Clinics are low-SES if at least one of its physician is low-SES. Clinics in columns 4 and 5 are from the non-closing sample.

Table 2.3: Summary Statistics - Patients

	(1) Population	(2) Analysis sample	(3) High-SES	(4) Low-SES
Male	0.504	0.510	0.532	0.454
Year of birth	1959.4	1957.4	1958.3	1954.9
Ethnical Danish	0.877	0.908	0.883	0.971
Low SES	0.288	0.285	0.000	1.000
Married	0.707	0.755	0.784	0.683
PCP low SES	0.360	0.204	0.204	0.203
PCP visit	0.812	0.835	0.828	0.852
Number of visits	5.064	5.148	4.766	6.108
Number of services per visit	1.415	1.435	1.423	1.465
Medical specialist	0.130	0.135	0.132	0.143
Total reimbursement	294.0	318.6	314.3	329.3
Death	0.083	0.053	0.043	0.080
Death from CVC	0.017	0.009	0.007	0.014
Death from COPD	0.004	0.003	0.002	0.005
Death from diabetes	0.002	0.002	0.001	0.003
Death from cancer	0.002	0.002	0.001	0.003
Chronic Condition	0.242	0.298	0.286	0.330
Number of chronic conditions	0.322	0.423	0.399	0.482
CVC	0.144	0.194	0.182	0.227
COPD	0.021	0.027	0.020	0.042
Diabetes	0.157	0.201	0.197	0.213
Number of observations	4,651,432	488,505	349,380	139,125

Notes: The table presents patient characteristics in different patient samples. See Appendix Table B.2 and B.1 for the used ICD and ATC codes. Patients are Low-SES if they have primary school as the highest level of completed education. Having a primary care physician (PCP) who is low-SES is defined as having a physician who has a parent with primary school as the highest level of completed education.

Table 2.4: Test for Selection in Patient-Physician Reassignment

	(1)	(2)	(3)	(4)
Physician characteristics	Low SES	Male	Non-ethnic Danish	Age>60
Patient characteristics				
Low SES	0.00468			
	(0.00585)			
Male		0.03484***		
		(0.00585)		
Non-ethnic Danish			0.03049***	
			(0.01079)	
Age >60				0.00251*
				(0.00151)
Observations	474614	474614	474614	474614
Patient characteristics	Y	Y	Y	Y
New physician characteristics	Y	Y	Y	Y
Old physician fixed effects	Y	Y	Y	Y

Notes: The table tests for selection in patients' assignment to new physicians post clinic closures. The table shows coefficients from regressing physician characteristics on patients having the same characteristic one year after clinic closure. The coefficients are the likelihood of physicians sharing the same characteristics with the patient. The regressions includes both new physician (on the clinic level) and patient controls excluding the variable under investigation. New physician controls includes: average age, share of male physicians, share of ethnic Danish physicians, dummy for being a solo clinic, number of physicians in the clinic, graduating institutions, and SES. Patient controls includes the following dummy variables: male, gender, non-ethnic Danish, married, and low SES. Standard errors are clustered at the old-physician level.

Table 2.5: The Effect of Physician(PCP)-patient SES Concordance on Mortality

VARIABLES	(1) Death	(2) Death	(3) Death	(4) Death	(5) Death
PCP low SES x Patient low SES x Post	-0.00130***	-0.00131***	-0.00130***	-0.00144***	-0.00134***
	(0.00038)	(0.00038)	(0.00038)	(0.00043)	(0.00039)
PCP low SES x Patient low SES	0.00000**	0.00002*	0.00001		0.00012**
	(0.00000)	(0.00001)	(0.00002)		(0.00006)
Patient low SES x Post	0.00536***	0.00505***	0.00503***	0.00500***	0.00508***
	(0.00022)	(0.00022)	(0.00022)	(0.00025)	(0.00022)
PCP low SES x Post	-0.00008	-0.00007	-0.00009	-0.00010	-0.00008
	(0.00016)	(0.00016)	(0.00016)	(0.00018)	(0.00016)
PCP low SES	-0.00000	-0.00001	-0.00005		
	(0.00000)	(0.00001)	(0.00006)		
Patient low SES	-0.00001*	0.00027	0.00022		-0.00012
	(0.00000)	(0.00023)	(0.00032)		(0.00071)
Post	0.00537***	0.00474***	0.00487***	0.00494***	0.00488***
	(0.00009)	(0.00009)	(0.00011)	(0.00011)	(0.00011)
Outcome mean	.00234	.00234	.00234	.00234	.00234
Gradient for high-SES physicians	.00541	.00541	.00541	.00541	.00541
Effect %	-24.0	-24.2	-24.0	-26.6	-24.8
Observations	3,749,654	3,749,654	3,749,654	3,749,654	3,749,654
Patient Characteristics	N	Y	Y	Y	Y
Old PCP FE	N	N	Y	N	N
Patient ID FE	N	N	N	Y	N
Old x new PCP FE	N	N	N	N	Y

Notes: The table presents the effect of physician-patient SES concordance on mortality. All columns report the estimates from the triple differences equation 3.3 with different controls. Patient characteristics includes age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education. Gradient for high-SES physician is the difference in the outcome variable between high and low-SES patients who have high-SES physicians in the post period, calculated as $(lowSESoutcome - highSESoutcome)$. The effect in percentage is calculated as $(Tripledifferenceestimate/gradientforhigh - SESphysician) \times 100$. Column 5 is our preferred specification. Standard errors are clustered by patient ID.
** p <0.01, ** p <0.05, * p <0.1.

Table 2.6: The Effect of Physician(PCP)-patient SES Concordance on Mortality Caused by Chronic Conditions

Cause of death	(1) CVC	(2) Cancer	(3) Diabetes	(4) COPD
PCP low SES × Patient low SES × Post	-0.00043***	-0.00044*	0.00006	-0.00004
	(0.00016)	(0.00025)	(0.00007)	(0.00009)
Outcome mean	.00042	.00098	.00007	.00011
Gradient for high-SES physicians	.00101	.00182	.00017	.00048
Effect %	-42,6	-24.2	35.3	-8.4
Observations	3,749,654	3,749,654	3,749,654	3,749,654
Patient Characteristics	Y	Y	Y	Y
Old x new PCP FE	Y	Y	Y	Y

Notes: The table presents the effect of physician-patient SES concordance on mortality caused by chronic conditions. All columns report the estimates from the triple differences equation 3.3. Patient characteristics include age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education. Gradient for high-SES physician is the difference in the outcome variable between high and low-SES patients who have high-SES physicians in the post period, calculated as $(lowSESoutcome - highSESoutcome)$. The effect in percentage is calculated as $(Tripledifferenceestimate/gradientforhigh - SESphysician) \times 100$. Standard errors are clustered by patient ID.
** p <0.01, ** p <0.05, * p <0.1.

Table 2.7: The Effect of Physician(PCP)-patient SES Concordance on Healthcare Utilization

VARIABLES	(1) PCP visit (Dummy)	(2) PCP visit (N)	(3) Services per visit (N)	(4) Specialist visit (Dummy)
PCP low SES x Patient low SES x Post	-0.00127	0.12377***	0.01343**	0.00337
	(0.00174)	(0.03298)	(0.00557)	(0.00233)
PCP low SES x Patient low SES	0.00045	-0.02179	-0.00296	0.00076
	(0.00169)	(0.03717)	(0.00423)	(0.00213)
Patient low SES x Post	-0.00186*	-0.06696***	0.00345	-0.00100
	(0.00097)	(0.01829)	(0.00306)	(0.00131)
PCP low SES x Post	0.00054	0.04292***	-0.00234	0.00320**
	(0.00098)	(0.01522)	(0.00295)	(0.00126)
Patient low SES	-0.09659***	0.33882	-0.12085	-0.01442
	(0.03215)	(0.53709)	(0.09800)	(0.05738)
Post	0.00505***	0.40673***	0.09732***	-0.00330***
	(0.00074)	(0.01012)	(0.00222)	(0.00095)
Outcome mean	.83866	6.24079	1.44509	.33085
Gradient for high-SES physicians	.02435	1.4598	.05943	-.01524
Effect %	-5.2	8.5	22.6	-22.1
Observations	3,749,654	3,140,867	3,749,654	3,749,654
Patient Characteristics	Y	Y	Y	Y
Old x new PCP FE	Y	Y	Y	Y

Notes: The table presents the effect of physician-patient SES concordance on healthcare utilization. All columns report the estimates from the triple differences equation 3.3. Patient characteristics include age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education. Gradient for high-SES physician is the difference in the outcome variable between high and low-SES patients who have high-SES physicians in the post period, calculated as $(lowSESoutcome - highSESoutcome)$. The effect in percentage is calculated as $(Tripledifferenceestimate/gradientforhigh - SESphysician) \times 100$. Standard errors are clustered by patient ID. ** p <0.01, ** p <0.05, * p <0.1.

Table 2.8: The Effect of Physician-patient SES Concordance on Health Behaviors Related to Chronic Conditions

CONDITIONS	CVC		Cancer		Diabetes		COPD	
	Statins	ACE	Lung scan	Radiologist	Metformin	Office visit	Medication	Hospitalization
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PCP low SES × Patient low SES × Post	0.00286*	0.00104	0.00096	0.00040	0.00028	0.01126***	-0.00123**	-0.00076
	(0.00169)	(0.00172)	(0.00108)	(0.00058)	(0.00097)	(0.00251)	(0.00049)	(0.00113)
Outcome mean	.10415	.12554	.03279	.01987	.04311	.09522	.00563	.05568
Gradient for high-SES physicians	.04643	.04235	.00887	-.0032	.02309	.0243	.00866	.03277
Effect %	6.2	2.5	4.5	8.8	1.2	46.3	-2.3	-14.2
Observations 3,749,654	3,749,654	3,262,358	3,749,654	3,749,654	2,123,957	3,749,654	3,749,654	
Patient Characteristics	Y	Y	Y	Y	Y	Y		
Old x new PCP FE	Y	Y	Y	Y	Y	Y		

Notes: The table presents the effect of physician-patient SES concordance on health behavior related to the four most common and unequal chronic conditions. All columns report the estimates from the triple differences equation 3.3. Patient characteristics include age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education. Gradient for high-SES physician is the difference in the outcome variable between high and low-SES patients who have high-SES physicians in the post period, calculated as $(lowSESoutcome - highSESoutcome)$. The effect in percentage is calculated as $(Triple difference estimate/gradient for high - SES physician) \times 100$. Standard errors are clustered by patient ID.
** p <0.01, ** p <0.05, * p <0.1.

Chapter 3

Employer Reputation and the Online Labor Market: Evidence from Glassdoor.com and Dice.com

3.1 Introduction

Extensive research in economics has shown that employees and job seekers not only care about wages, but also non-monetary compensation such as fringe benefits, job amenities, and culture (e.g. (61)). The literature on compensating differentials usually assume that workers are certain about this trade-off when calculating workers' willingness to pay for such benefits. However, uncertainties arise when workers search for jobs, not only because wage differences are unclear at this stage, but also that there exists asymmetric information around these attributes. Traditional means to address these asymmetries in product markets are evaluations of the previous consumers' perception via word of mouth, brand name, or media ranking systems (7). Together, they form a reputation on the product (32). Similarly, evaluations on an employer form a reputation of the employer in the labor market.

An abundant literature has addressed the role of reputation in regulating markets that have incomplete contracts or moral hazard problems (78, 13, 96). Theoretically, reputation of firms are usually modeled as an accumulated asset that firms invest or disinvest in, reputational incentives then determine firms' investment decisions. Empirical work in this space, such as (8) and (11), test for how reputation regulates product and gig-economy markets. There is less causal evidence on the role of reputation on the labor market outside of the gig economy. Although employment may be viewed as an experience good, reputation of employers in the non-gig labor market is different. Compared to the product market, the stakes are higher since the end result is a long-term employment, rather than a purchase; compared to the gig-economy, the reputation of an employer is not only about the fulfillment of contracts, but also employee satisfaction beyond compensation.

Employer reputation is hard to measure. Existing papers use rankings of the best employers to work for as measure of employer reputation (100). Second, data on job postings have only become available in the past decade, and data on job applications is even rarer. As technology drastically lowered the cost of delivering information, websites such as Glassdoor.com and Indeed.com became a part of the market of evaluation. By aggregating individual reviews, these websites provide a way of measuring reputation for each employer. Similarly, job board websites such as LinkedIn.com and Dice.com allow for tracking of applications to job postings.

Leveraging the availability of such data, we ask: what is the effect of reputation on the demand and supply sides of the labor market? The effects of reputation is ex ante unclear. The supply side may respond in the following ways: first, since evaluations are crowd-sourced, and providing evaluation is costly, evaluations may be under-supplied and subject to selection (90). Aggregating evaluations at the extremes may result in inaccurate measure of "employment quality". If job seekers perceive displayed reputation to be biased signals of quality, they may not view it as a valuable source of reputation at all, and turn to alternative sources, such as brand name recognition. Online reputation would not affect firms' number of applications received in this scenario. Second, if job seekers find displayed reputation credible, we may expect to see that firms with better reputation draw *more* and/or *better* applicants. Third, if job seekers perceive displayed reputation as a signal for the competitiveness of the screening and interview process, a firm with better reputation may attract an applicant pool that is not bigger, but better. Both the second and third scenarios imply that reputation facilitates better employer-employee match quality.

On the demand side, displayed reputation change inter-temporally as new evaluations are submitted. If employers are cognizant of their contemporaneous displayed ratings, and that they perceive the displayed reputation to have an impact on prospective employees, they may respond by conducting more recruiting while having higher displayed rating. On the reverse, if search is costly, they may advertise for fewer positions when the rating is lower than the anticipated rating in the future. It could be that when a firm experiences an improvement in displayed reputation, they recruit for positions that are previously unfilled; or that firms preemptively conduct more recruiting. In addition, firm reputation is correlated with firm performance (12), which could lead to creation of more positions, hence more job postings.

The main identification challenge in uncovering the impact of reputation on the labor market is that reputation when accurately measured should reflect the true quality of an employment at a firm. Fringe benefits that affects employee satisfaction, such as the employer's generosity of health care insurance, may even be explicitly stated on a job posting. Therefore the effects from displayed reputation and the underpinnings of reputation can hardly be separated.

In this paper, we narrow the gap in the literature by studying how crowd-sourced reputation affects both the supply and demand sides of the labor market. We do so by leveraging the nature of evaluation websites' operations. Since processing individual evaluations, at times contradictory ones, is costly (7), these websites aggregate individual evaluations into a composite score, and round the score to one decimal place for display. Rounding thresholds provides us with exogenous variations, showing two different scores for employers with almost identical underlying composite scores and quality of employment. We use a Regression Discontinuity Design (RDD) around the rounding threshold and compare employers that are barely rounded up and barely rounded down. On the supply side, we study whether job seekers value reputation, and on the demand side, we investigate whether employers' recruiting behavior respond to their own displayed reputation. We obtain displayed reputation data from the leading employer review website, Glassdoor.com, and job postings and applications data from an online job board, Dice.com.

We find that the number of views received per job posting is roughly 3% higher for firms that are barely rounded up, compared to those barely rounded down, while the number of application per posting increases by 7% when a firm is barely rounded up. The difference in application and view responses aligns with the fact that the cost of applying is higher than that of viewing a job posting. At the same time, employers that are barely rounded up recruit more actively by (1) re-activating more old postings, which suggests that they now recruit for positions that are previously not filled, and (2) posting more new positions, which is suggestive that they use the superior rating to conduct more recruiting.

What is the relationship between crowd-sourced reputation online and offline reputation, such as brand name or rankings produced by the media (e.g. Fortune Best Places to Work)? Literature has documented that online and offline reputation are likely substitutes in product markets (58, 52). We test whether this holds in the labor market using two variables to proxy for off-line reputation: whether the employer is a public firm and the size of the firm. We find that both supply and demand side effects are strongest for employers that are private and have fewer employees.

To study the mechanism for supply-side effects, we study the role of information in two ways. Theoretical literature suggests that reputation is an asset that accumulates over time. Following this literature, we find that firms with sticky ratings show stronger effects around the rounding thresholds in all variables of interest. Second, we study the volume of information. We compare firms that have a greater or smaller number of cumulative evaluations at the time of job postings, and find that firms with fewer cumulative evaluations that are less established display stronger effects. This shows that the effect is stronger when information is scarce.

To examine whether the demand side responds to the *levels* of displayed reputation or *changes* treatment status, we compare firms that are barely rounded up in time t and barely rounded down in time $t-1$ with those that are barely rounded down in both time periods. Firms indeed respond to this change in treatment with estimates twice as large the main analysis focusing on levels of reputation. However, the supply side is not sensitive to this week-to-week change in treatment status, possibly because few job seekers track ratings of potential employers weekly.

The remainder of the paper is organized as follows. Section 2 describes the institutional setting and our data. Section 3 discusses our empirical strategy. We discuss our main results and robustness checks in Section 4, discussions of mechanisms in section 5, and conclude in Section 5.

3.2 Institutional Background and Data

In this section, we discuss how we construct our working sample, and institutional details that may affect interpretation and identification.

3.2.1 Reputation data from Glassdoor.com

Glassdoor.com hosts anonymous employer reviews that are voluntarily submitted to the website. Each reviewer can only submit one review per employer per year regarding a job position at a firm; they can separately evaluate interview processes with more than one firm. After an evaluation is submitted, it goes through a moderation process that is either automated or by human eye. Reviewers can expect that their evaluation shows up on the company home page with a one-week lag. The content of each review may contain information including overall rating, employment status, and text including the title of an evaluation, as well as the "pros and cons". They can provide their job title, tenure at the firm, and office location by choice. Although evaluations are anonymous, reviewers may be worried about employers' ability to locate them using information such as job titles and tenure, hence over forty percent reviewers omit this information. The overall rating provided within each evaluation are integers from one to five, with one being the worst. As shown in Figure 3.2, the number of five-star ratings has increased over our sample period, resulting in an inflation of the displayed rating at the employer level. In our sample period, the medium review is 3.5.

Individual reviews are displayed on an employer's main page. The page provides firm characteristics such as it industry, most recent firm size, and firm age. Since reviews are voluntary, and some employees are bounded by Non-Disclosure Agreements, reviews are likely under-supplied (90). In order to encourage supply of evaluations, Glassdoor hides most of its content to those who have not yet submitted any review, and only displays the aggregated firm-level rating to someone who visits the website without an account. The displayed rating can be also viewed via a Google search without entering Glassdoor.

3.2.2 Job posting and applications data from Dice.com

The outcomes of interest are constructed using DHI Vacancy and Application Flows database constructed by (29) (the DHI data, henceforth). The DHI data contains job postings and job applications data from Dice.com, an online job board where employers pay to advertise job openings, and employers submit resumes and applications without additional charge. Dice.com focuses on industries that hire highly-skilled workers, such as technology, finance, business operations, etc. Therefore, the job seekers we focus on in this paper are skilled workers searching for long-term employment.

From DHI data, we obtain employer information including employer location, number of employees, ownership structure, and whether the company is a third-party labor market institution recruiting on behalf of other employers. At the job posting level, we can locate where the job takes place, job title, job description, contract terms, and wages, if employers choose to display them. We observe the date-time stamp when postings first becomes available to receive applications, when they are deactivated for display, or when they are re-activated. The cost associated with each posting increases by how long it is active. Daily costs in advertising for a position incentivizes employers to de-activate a posting once a position is filled, or if they do not expect to find a good candidate. At the day$\times$firm level, the DHI data tracks the number of seconds each posting is active, the number of views it receives, and its daily applications. It also allows for second-by-second tracking of each job seekers' activity on the platform. For each job seeker, we observe the IP address from which they submit each application, their work authorization status, and a self-reported job title.

We focus on job postings for US-based employment, which means we exclude jobs that are remote, but includes jobs that allow for telecommuting so long as a job location is specified. Our final job posting and job application sample includes 181,884 employers and over 6 million job postings located across the U.S.. As shown in Table 3.2, the average employer has 5 active postings per month, with average vacancy duration of 40 days.

3.2.3 Constructing analysis sample

Our working sample contains 181,884 employers on Glassdoor that meet two conditions: 1) can be perfectly matched with at least one employer in the DHI database, 2) have more than 10 reviews on Glassdoor. We focus on employers on Glassdoor because employer pages are unique on Glassdoor, while the same employer in DHI database can have more than one account. We set the second restriction because Glassdoor only calculate and display the aggregated rating for employers with more than 10 reviews. These restrictions leave us with more than 3 million company reviews from Glassdoor and 3+ million vacancies from DHI database in our working sample during 2012-2017.

The unit of observation is at the firm$\times$week level. When aggregating job posting to the firm level, we only keep data from the first seven days of a job posting. This is because job postings attract 75% of its applications in the first seven days (29). This can be because applicants prefer "younger" postings if they believe that the position is not yet filled, or that they have an advantage in being the early candidate. Second, since the job board ranks job postings on the web page, the posting ranked higher on the first page may gain more attention. Ranking algorithms is a central problem job board websites face. The age of a posting is one of many determinants of the position it takes on a web page. Third, the duration of vacancies may be affected by (1) how fast a position is filled, and (2) whether companies strategically de-activiate and re-activiate positions to gain attention. The former would reflect the company's ability to attract workers. Smaller number of applications received over the whole job posting period does not indicate that the firm is *less* attractive to job seekers, it might even mean that the firm is *more* attractive if the duration is short. For this reason, we only consider applications within the first seven days of a posting. Doing so allows us to control for firm heterogeneity in how long they decide to keep a posting active.

In order to interpret outcomes in percentage terms when a firm is barely rounded up, we transform the count of applications and job postings into inverse hyperbolic sine, rather than taking the log. This is because a large number of firms and job postings post zero positions, or receive zero applications in a week. Tables 3.2 and 3.1 report summary statistics on the labor demand and supply sides of our working sample.

3.3 Identification Strategy

Our main strategy is a sharp Regression Discontinuity design (RDD) based on the distance between an employer's weekly-updated raw rating on Glassdoor and the rounding thresholds. The local linear regression specification is given by:

$$\mathrm{Y}_{it} = \beta_1 RoundedUp_{it} + \beta_2 f(X_{it} - C_t) + \beta_3 RoundedUp_{it} \times f(X_{it} - C_t) +$$
$$\mathrm{sector}_s + month_m + threshold_c + \varepsilon_{it}$$

In this model, y_{it} denote the outcome variables of interest for employer i in week t, including the inverse hyperbolic sine of the count of new postings, re-activated postings, deactivated postings, average number of views per posting, and average number of applications per posting, etc. *Rounded Up*$_{it}$ is an indicator which equals 1 if the raw aggregate rating is above a rounding threshold, and the coefficient of interest β_1 tells us the effect of a 0.1-star increase in an employer's displayed rating around the cutoff. $(X_{it} - C_t)$ is the difference between the raw rating of employer i in week t and corresponding threshold. We include *Rounded Up*$_{it} \times f(X_{it} - C_t)$ to allow for differential slopes on either side of the rounding thresholds. We also include month by year, sector, and threshold fixed effects to account for time, sector, or threshold common shocks. We use local linear regressions with a triangular kernel for all regressions. The coefficient of interest is β_1, it measures how much the outcomes of interest differs for employers that are barely rounded up relative to those barely rounded down.

The identifying assumption is that the rounding process is as good as random around the thresholds, hence orthogonal to employer's firm performance or employment quality. We provide evidence showing that this assumption likely holds below.

3.4 Validity of RD Design

In this section, we verify the key identifying assumption of RDD and potential threats to identification in our setting.

In order for treatment to be assigned randomly, employers should not be able to manipulate whether they are rounded up or down when they are close to a rounding threshold. It is important to check for manipulation in our setting since firms may have an incentive to produce more positive evaluations without investing in reputation among employees. They may do so by posting fake positive evaluations if reviewers' true employment is hard to be verified. We argue that perfect manipulation is hard to achieve in practice. First, Glassdoor.com prevents abusing of the site by developing proprietary detection algorithms. Second, although rating inflation is in place in our sample period, as shown in Figure 3.2, it is hard for employers to manipulate around the threshold as they do not know their raw rating and how close they are to the threshold. To prevent employer manipulation, the aggregation algorithm that sums up individual reviews is also proprietary, weighting reviews depending on evaluation characteristics, such as time since the previous evaluation or voted helpfulness of an evaluation, so that an individual employer rating would not change drastically by a single evaluation. Third, we formally verify that perfect manipulation does not exist by plotting the density of the centered raw aggregate ratings in Figure 3.1 and show that the density is smooth across rounding thresholds. We also run the McCrary test (62) on the density of raw ratings and fail to reject the null hypothesis that the density is continuous at the rounding thresholds (p value = 0.24).

One drawback due to data limitation is that our running variable is not continuous. Displayed rating is at one decimal place, while the underlying raw rating we have is two decimal places. The concern is that when the running variable is discrete, one would have too few observations near the treatment threshold. A data-driven approach to choosing the optimal bandwidth may lead to a window that is too large and the estimate may be biased by observations at the ends of the bandwidth. This data feature should not be a concern in our particular case, since there are a lot of observations at each running variable value, allowing us to identify the treatment effect close to the rounding thresholds. We follow (55) and do not cluster standard errors at the running variable level, but use Eicker-Huber-White standard errors. While we present main results using 0.04 bandwidth, we include results using different bandwidths in the Appendix.

Since the treatment variable and outcomes come from different sources, the effect we are estimating is the reduced form effect of barely rounding up. Estimating the treatment on the treated would require weighing our estimates by the proportion of Dice.com applicants that make a visit to Glassdoor.com, or searches for an employer's Glassdoor rating on Google. We highlight that Glassdoor is the largest website that hosts employee reviews, while the ideal job applications data should come from Glassdoor.com.

Next, to test for nonrandom treatment assignments across rounding thresholds, we compare employer characteristics of those barely rounded up or down. Table 3.3 shows RDD regression results using several firm characteristics to check that they are not discontinuous at rounding thresholds. First, we check whether the number of reviews submitted in each week is discontinuous, which would be indicative of strategic gaming of the firm when they are around rounding thresholds; we do not find this to be true. Second, if employers are systematically different across thresholds in terms of their average recruiting behavior over the whole sample period, the effects we observe may be attributed to unobserved firm characteristics, rather than labeling effect of reputation. Therefore, we construct variables that reflect employer behavior, including average vacancy duration across the whole sample duration and vacancy duration for the positions posted in week t, as well as whether the firm makes wages available on job postings. Lastly, we also check for jumps in firm characteristics, such as the number of employees (filled positions) and whether an employer is a public firm at the time of postings, neither show any discontinuity. These suggest that employers above and below the rounding thresholds are similar, and these employer attributes are orthogonal to treatment assignment.

3.5 Results

In this section, we examine whether and how supply and demand sides of the labor market respond to displayed reputation around rounding thresholds. The first set of results separately examine the effect on job seeker behavior and firm behavior at the equilibrium, since job seeker response can affect that of the firm, and vice versa. The second set of results address the relationship between online displayed reputation and other sources of reputation. Lastly, we examine potential channels of job seeker response by zooming in on the role of information, and whether employers respond to changes or levels of displayed reputation.

Figure 3.4 plots the inverse hyperbolic sine of the number of application per job posting in the first seven days of a posting, suggesting a significant increase in the number of applications received per posting just above the rounding threshold. In other words, firms that are barely rounded up experience an in crease in their ability to attract *more* workers. Corresponding results using local linear regression with triangular weights are reported in Table 3.5. Our preferred specification in column (3) shows that when employers are rounded up, the number of applications received per posting in the first seven days of a posting is 7% higher than the number of applications received per posting for those that are barely rounded down. Estimates remain numerically similar when we exclude year-month- and sector, or threshold-fixed effects in columns (1) and (2).

For robustness, we estimate column (3) in Table 3 using alternative inference methods in Appendix Table 11. Appendix Table 11 column 1 through 10 further show that our results are robust to varying the bandwidths between 0.02 and 0.03, and estimating with higher order polynomials or alternative kernel functions, and clustering by employer and the running variable, respectively.

We also plot how number of views per job posting receives in the first seven days in Figures 3.4. While there is a jump in the number of views at the threshold, it is half the size of the estimate in the number applications a posting draws. This is potentially because the cost of clicking on an application is lower than that of applying for one. A job application leads up to a screening process involving interviews, while viewing a position is private and merely reveals interest. Applicants are also a lot more likely to search for a company's information before applying, rather than before clicking into a job posting. Corresponding regression results are shown in Table 3.5. Results are robust to varying fixed effects, bandwidths, inference methods, kernel function, and higher order polynomials.

We examine supply side behavior around the threshold by looking at when and how they post job positions. We define job postings by the following types: new postings, re-activated postings, and deactivated postings. The first two variables correspond to active recruitment of an employer. De-activation and re-activation of a posting within the same week may imply the recruiter's attempt to boost the rank of a posting on a web page to draw attention, which signals a lack of ability to recruit the fitting candidates in the first place. We define the gap between the date of new posting and the date of permanent removal to be vacancy duration, except for instances when employers re-use the same job posting for multiple recruiting seasons. We do not use vacancy duration as an outcome of interest because it reflects both supply and demand side behavior. A shorter duration may correspond to more applications received in the first few days of a posting or employers' attempt to keep postings "young".

Figure 3.5 plots the inverse hyperbolic sine of the number of new job postings and re-activated postings in a week, suggesting a significant increase in active recruiting behavior just above the rounding threshold. We also see an increase in the number of de-activated postings. Corresponding regression results are reported in Table 3.4. Our preferred specification in column (3) shows that when employers are rounded up, the number of new posting in the first seven days of a posting is 4.3% higher than the number of new postings from firms that are barely rounded down. The number of re-activated postings also increase by 4.8%. Estimates remain numerically similar when we exclude year-month- and sector, or threshold-fixed effects in columns

(1) and (2). The reactivation of existing postings implies that firms make use of when they are above the threshold to search for positions that are not previously filled. The number of more new postings may be a combination of searching for previously failed searches, and preemptive search when the firm is barely rounded up. The preferred estimates for number of re-activated and de-activated postings in the same week are numerically similar, suggesting that some of these postings reflect recruiters' attempt to draw job seekers' attention. To study whether employers' response comes from levels of displayed reputation or a change in their own displayed reputation, we conduct a separate analysis in Section 3.6.

Heterogeneous effects at different thresholds In the above analysis, we stack multiple thresholds in a regression. However, is the effect the same across different thresholds? For example, do job seekers respond to a displayed rating change from 1.9 to 2 (corresponding threshold is 1.95) the same way as they do from 2.9 to 3 (corresponding threshold is 2.95)? At which point do job seekers stop responding to reputation differences, and at which point do employers stop changing their recruiting behavior? We separate our sample in to five by thresholds, and plot the the estimates and confidence intervals in Appendix Figure 10. We find that application responses is the strongest when the underlying reputation rating changes from 2.94 to 2.95, and displayed reputation changes from 2.9 to 3. Since the median displayed rating is 3.5, this implies that job seekers care the most about employee reputation when it is not the worse half of the entire pool. The estimate is closer to zero and not statistically significant when reputation is at the highest and lowest values.

The role of sticky reputation in rounding effects To study the mechanism for supply-side effects, we study the role of sticky information. The management literature on firm reputation suggest that reputation is an intangible asset that is sticky by nature when measured by media exposure and ranking systems. The more sticky reputation, the more certain perceived beliefs are in the market. Following this literature, we hypothesize that when displayed reputation is sticky, the effect around the rounding threshold is stronger. As RDD focuses on observations nearest to rounding thresholds, firms near the thresholds are by construction sensitive to treatment status, a single new evaluation may pull the displayed rating up or down. Sticky rating may appear when (1) new evaluations are near the rounding threshold, (2) new evaluation is weighted down due to a large number of cumulative evaluations or other built-in features of the aggregation algorithm, or (3) the company rarely receives a new evaluation.

We test for the role of reputation stickiness by dividing the sample by stickiness. A firm in week t is in the more sticky sample when the firm's displayed reputation is unchanged between $t-6$ and $t-1$, since five weeks is the medium time that firms in our sample have invariable ratings. In the less sticky sample, the firms nearest to the thresholds experience less change in underlying ratings than those farthest away from the threshold. As shown in Figure 3.6 and Table 3.8, the effect is only present in the more sticky sample, and is twice the size of the effect in the overall analysis. This suggests that both the employers and the job seekers value information that are the most certain about.

Relationship with off-line reputation What is the relationship between crowd-sourced reputation and traditional types of reputation, such as brand name, ranking systems, and media rhetoric? In product markets, consumers treat chained restaurants differently than they do independent smaller restaurants (58); franchised restaurants are more likely than chain restaurants to free-ride on the brand name when it comes to restaurant hygiene because of differences in profit maximization (52). We test whether the same holds in the labor market in our sample by looking at heterogeneous effects between firms that have different levels of off-line reputation. We use two variables to proxy for off-line reputation: (1) whether the employer is a public firm, (2) whether the firm as more or less than 400 employees.[1] We find that both supply and demand side effects are disappear and are close to zero for employers that are public or have fewer

[1]The median firm in our working sample has 400 employees. Firms with more than 500 hundred employees are often classified as large.

employees, while the smaller firms have twice the effect than the overall sample.

In addition to off-line reputation, we also use the number of cumulative reviews at the time of job postings to proxy for how established a firm is on Glassdoor.com. Consistent with (90), we find that effects are strongest in firms that have fewer than 34 cumulative reviews at the time of a job posting. This could potentially be attributed to job-seekers' response to reputation depends on the scarcity of information.

Demand-side response to changes in displayed reputation The main analysis finds that firms that are barely rounded up post more new positions and reactivate more existing positions. To examine whether the demand side respond to the *levels* of displayed reputation or *changes* in their treatment status, we conduct the following exercise. For each firm$\times$week, we limit our sample to those that are untreated below rounding thresholds at time $t-1$, and use outcomes in time t to run the same RDD regression. This exercise compares firms that barely become treated in time t and those that stay untreated. We find that firms indeed respond to changes in treatment status much stronger. Those that barely round up post 18% more new positions and re-activate 4.7% more existing positions, suggesting that the effect we see in the main analysis may come from firms responding to changes in own-reputation across time. The supply side is not sensitive to this week-to-week change in treatment status, potentially because job seekers do not track changes in employer reputation at such a granular time unit (weekly).

3.6 Robustness

In this section, we present robustness checks addressing concerns regarding potential outliers, and conduct an alternative analysis on the job posting level.

3.6.1 Job level analysis

In the main analysis, observations are at the employer×week level. The benefit of doing so is that it allows us to analyze employer response to treatment and helps us understand the heterogeneity in human resource practices between firms. The draw back is that the supply-side outcome variable averages the total number of applications received for a firm over the total number of job postings; a change in this outcome variable may come from the demand side *and/or* the supply side. In addition, the average treats a 7-day-old posting from the previous week the same as "young" posting that is 1-day-old, while the number of applications received is correlated with the age of postings. Since we are not able to single out supply and demand side effects, we conduct an exercise at the job posting level in order to zoom in on supply-side responses. Similar to the main analysis sample, we keep jobs applications and job views in the first seven days to control for the role of age effects.

Following our main RDD specification, the estimating equation for job-level analysis is

$$\mathrm{Y}_{itj} = \beta_1 RoundUp_{ijt} + \beta_2 f\left(X_{ijt)} - C_t\right) + \beta_3 RoundUp_{ijt} \times f\left(X_{ijt} - C_t\right) + \\ \mathrm{sector}_{ij} + month_t + threshold_c + \varepsilon_{ijt}$$

where j denotes the job posting of firm i at week t. We include the similar set of controls that remove labor market shocks at month t, sector of the firm i, and threshold c. Regression results are presented in Table 3.10. On the job level, we no longer observe jumps near rounding thresholds. We discuss the seemly contradictory results below.

First, the fact that demand-side respond to rounding cutoffs implies that the job postings are not distributed smoothly around roundingn thresholds. While firms cannot manipulate their displayed ratings, manipulation can be done by changing recruiting behavior inter-temporally. The strategic response of the firm makes a job-level RDD subject to bias; it could be driven by characteristics of the firms that choose to recruit strategically, rather than true applicant behavioral response.

Second, if the number of applications is unaffected by the treatment in our firm-level analysis, a jump in the average application received would correspond to a decrease in the number of postings for the firms that are barely rounded up. However, the contrary is found in the main analysis; treated firms increase, rather than decrease, total active postings by increasing more new and re-activated positions.

3.6.2 Concerns related to outlier

Graphical representation of the RDD in Figure 3.4 shows that the outcome variables of interest does not have a clear upward trend upward as the running variable increases. To address concerns that the jump at the cut-off is driven by noise to the left of the rounding thresholds, we conduct the following exercises. Since the sample average barely to the left of the rounding threshold is lower, and triangular weighting weighs observations closest to the threshold the most, one may worry that the triangular kernel exacerbates outlier concerns. Therefore, we replace triangular kernel with uniform kernel and present our results in the appendix, and show that our results are robust to this change in specification.

An alternative approach to changing the kernel function is to conduct a Difference-in-differences exercise, treating those above a threshold as treatment group, and those below as control. The identifying assumption would be that the firms above the threshold should trend in the same way as the postings below cutoffs. This may not hold in our sample, especially for those that are further away from rounding thresholds.

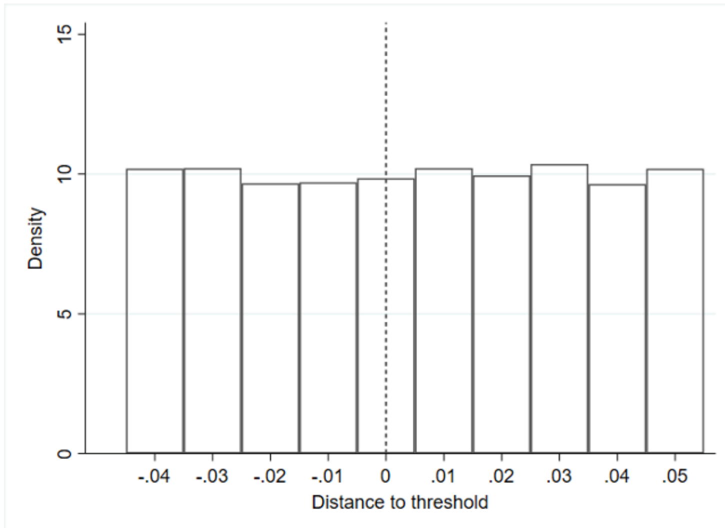
3.7 Conclusion

This paper studies the effect of displayed employer reputation on job-seeker behavior and employer recruiting behavior. We use novel data combining reviews from Glassdoor.com with job posting and applications from Dice.com. Raw crowd-sourced employer ratings are rounded when displayed to job seekers. By exploiting the rounding threshold, we identify the causal impact of Glassdoor ratings using a regression discontinuity design. We find that displayed employer reputation affects employer's ability to attract workers, especially when the displayed rating is sticky. Employers respond to the rounding threshold by posting more new positions and re-activating more job postings. The effects are the strongest for firms that are private, smaller, and less established, suggesting that online reputation is a substitute for offline reputation. Our work contributes the personnel economics literature and extends what we know about how firms approach human resources problems from both the workers' and the firms' perspectives (72).

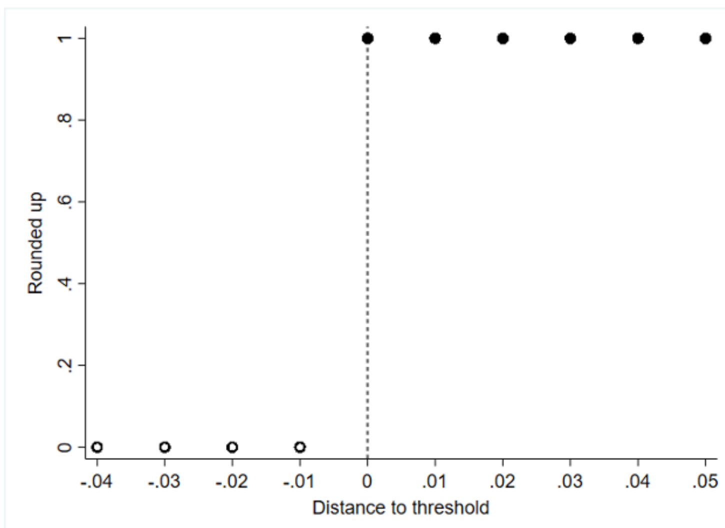
There are many unanswered questions related to the impact of firm reputation on labor markets. First, better employer reputation may not only draw *more* applications to a job posting, but also *better* applicants. A natural extension would be to study the quality of applicants in addition to the quantity. Second, this paper does not pin down how supply side effects, or firms' belief of supply side effects, influence the demand side. It would be meaningful to disentangle one effect from the other. Third, when firms attract *more* and/or *better* workers, does the match quality between employer and employees improve? Does the screening process become more or less costly?

3.8 Acknowledgement

Chapter 3, in full, is currently being prepared for submission for publication of the material. Ke Ma; Yanying Sheng. The dissertation author was a primary investigator and author of this paper.



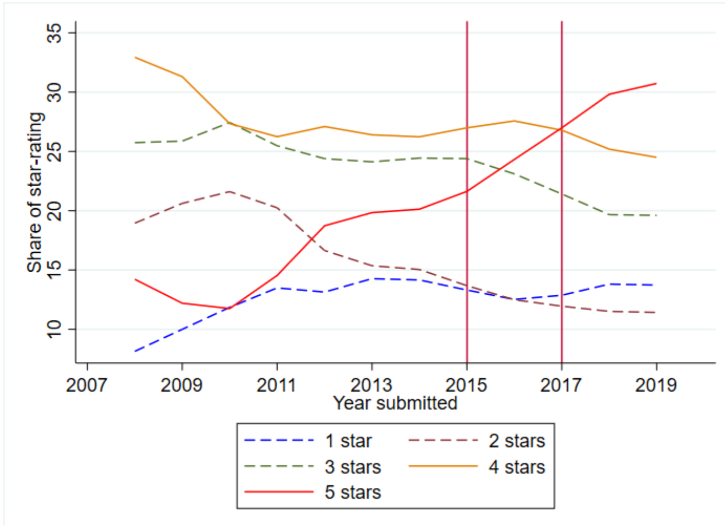
A: Density of Centered Firms



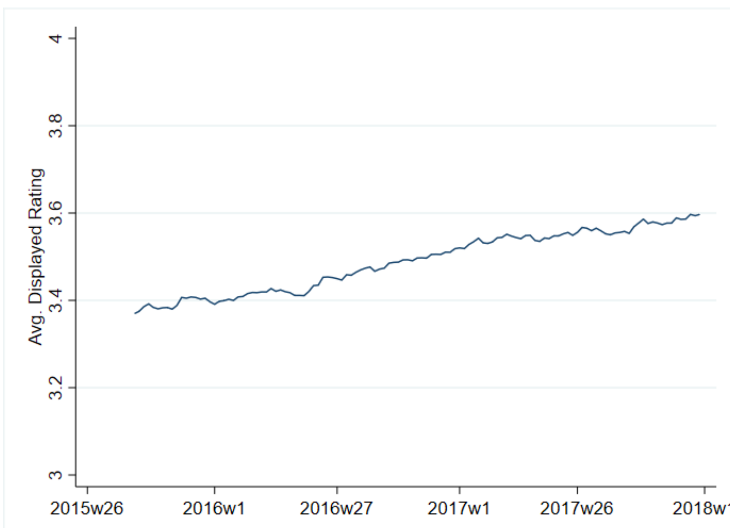
B: First Stage

Note: The figure plots two tests on the validity of the main RD sample. Panel A plots the density of firms centered at rounding thresholds. Panel B plots the first stage of the main analysis described in equation 1. The dashed vertical line denote the rounding threshold and is normalized to zero.

Figure 3.1: Validity of RD



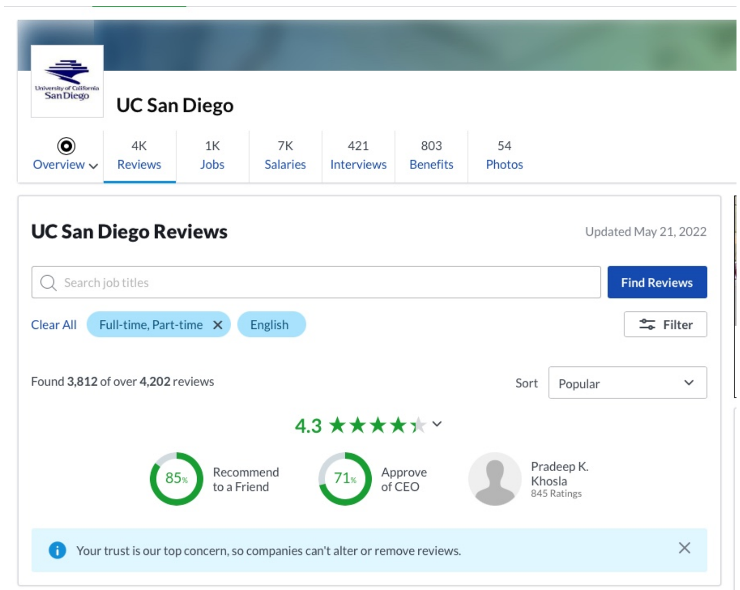
A: Raw Evaluation



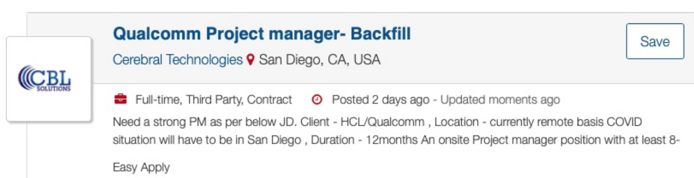
B: Average Displayed Reputation

Note: The figure shows the proportion of evaluations submitted over time (Panel A) and the resulting displayed reputation over time (Panel B).

Figure 3.2: Displayed Reputation Summary Statistics



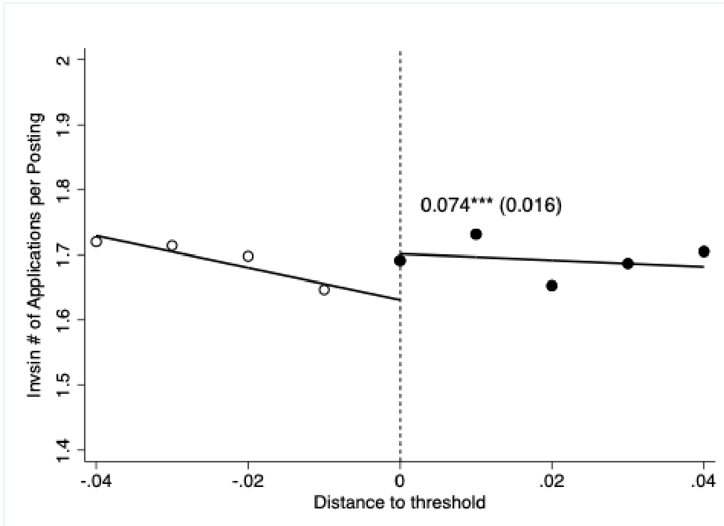
A: Employer Page on Glassdoor.com



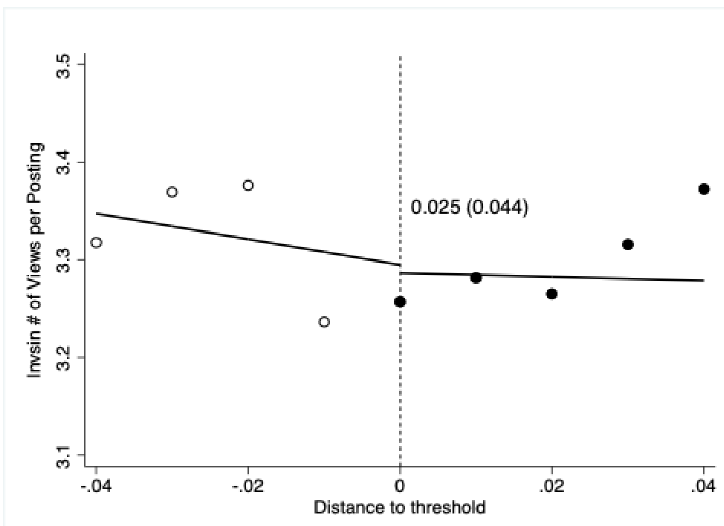
B: Job Posting on Dice.com

Note: The figure shows an example employer page with displayed reputation on Glassdoor.com (Panel A) and an example job posting on Dice.com (Panel B).

Figure 3.3: Example Employer Page and Job Posting



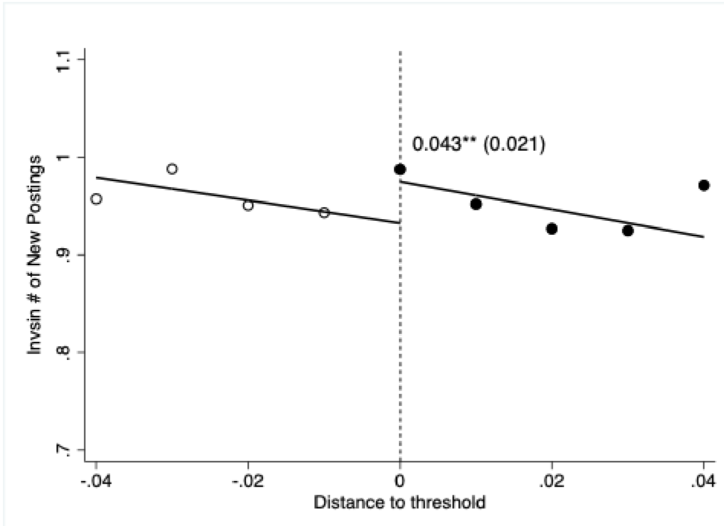
A: IHS(Applications per Posting)



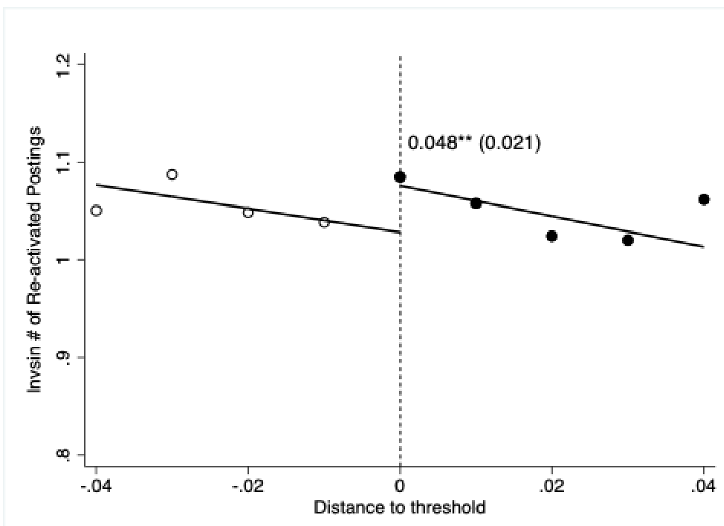
B: IHS(Views per Posting)

Note: Each observation is the average weekly inverse hyperbolic sine of the number of applications per posting (panel A) and number of views per posting (panel B). Dashed vertical line denotes rounding threshold and is normalized to 0. The solid lines are estimated using a local linear regression with triangular weights and firm $\times$ week data following equation 1. Corresponding regression results can be found in Table 3.5. Standard errors are calculated using Eicker–Huber–White standard errors.

Figure 3.4: The Effect of Displayed Reputation on Job Seeker Behavior



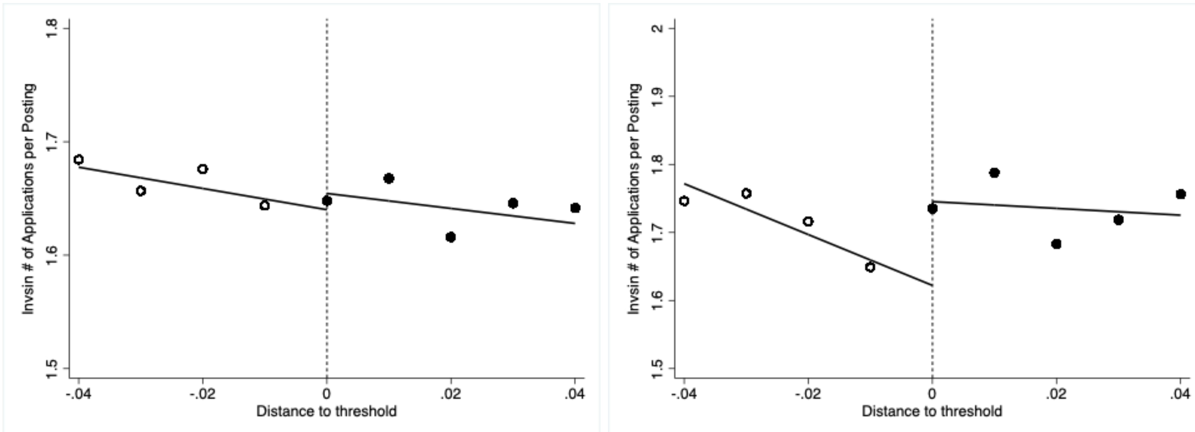
A: IHS(New Posting)



B: IHS(Reactivated Posting)

Note: Each observation is the average weekly inverse hyperbolic sine of the number of new postings (panel A) and the number of re-activated posting (panel B). Dashed vertical line denotes rounding threshold and is normalized to 0. The solid lines are estimated using a local linear regression with triangular weights and firm × week data following equation 1. Corresponding regression results can be found in Table 3.5. Standard errors are calculated using Eicker–Huber–White standard errors.

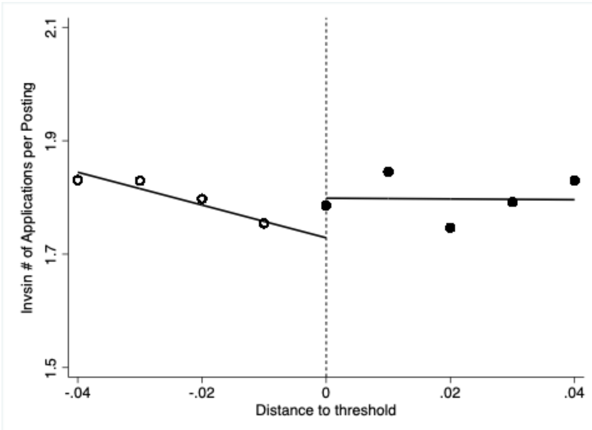
Figure 3.5: The Effect of Displayed Reputation on Employer Behavior



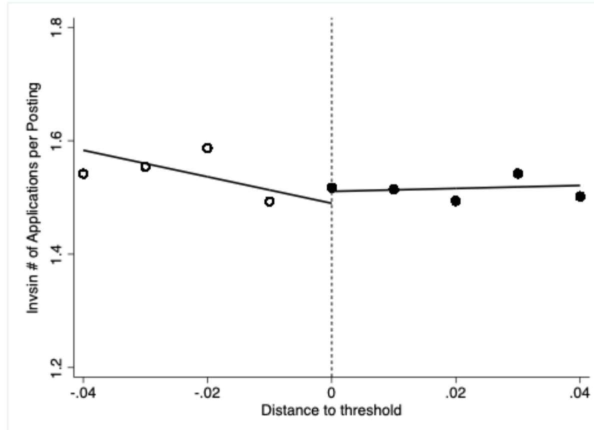
A: Less Sticky **B:** More Sticky

Note: The figure plots the heterogeneous effect of displayed reputation by displayed reputation stickiness. Each observation is the average weekly inverse hyperbolic sine of the number of applications per posting for firms with less sticky displayed ratings (Panel A) and more sticky ratings (Panel B). Stickiness is defined in Section 3.5. Dashed vertical line denotes rounding threshold and is normalized to 0. The solid lines are estimated using a local linear regression with triangular weights and firm $\times$ week data following equation 1. Corresponding regression results can be found in Table 3.8. Standard errors are calculated using Eicker–Huber–White standard errors.

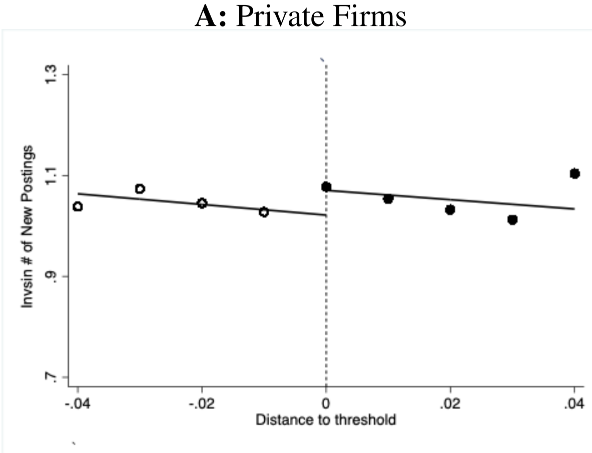
Figure 3.6: Heterogeneous Effect of Displayed Reputation by Stickiness



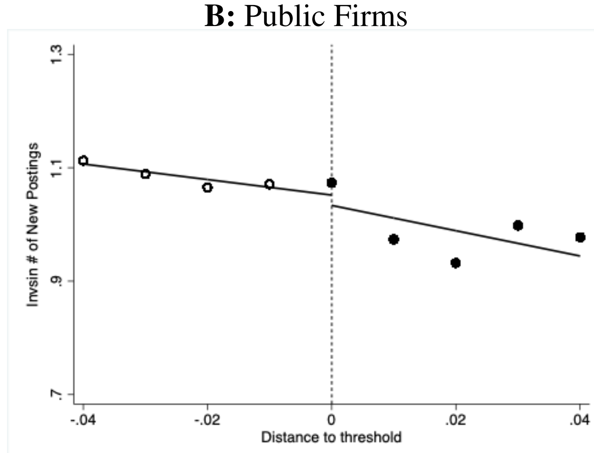
A: Private Firms



B: Public Firms



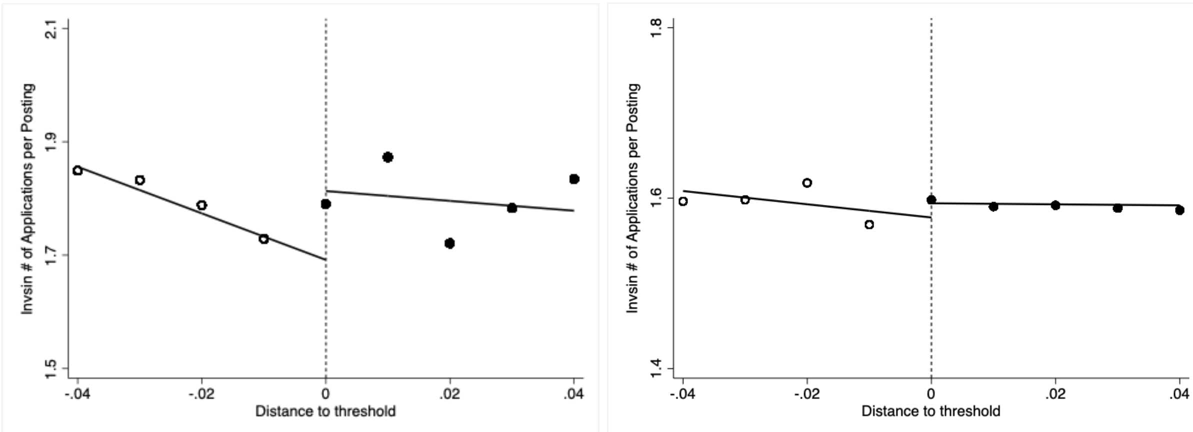
C: Private Firms



C: Public Firms

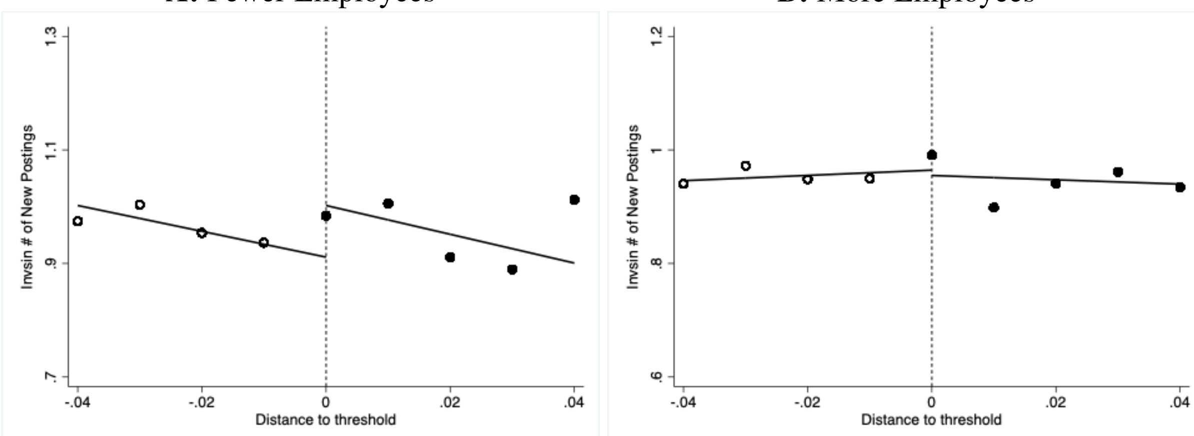
Notes: The figure plots the heterogeneous effect of displayed reputation by firm type. Each observation is the average weekly inverse hyperbolic sine of the number of applications per posting (Panel A and B) and the number of new posting (Panel C and D) for public firms (Panel A and C) and private firms (Panel B and D). Dashed vertical line denotes rounding threshold and is normalized to 0. The solid lines are estimated using a local linear regression with triangular weights and firm $\times$ week data following equation 1. Corresponding regression results can be found in Table 3.6. Standard errors are calculated using Eicker–Huber–White standard errors.

Figure 3.7: Heterogeneous Effect of Displayed Reputation by Firm Type



A: Fewer Employees

B: More Employees

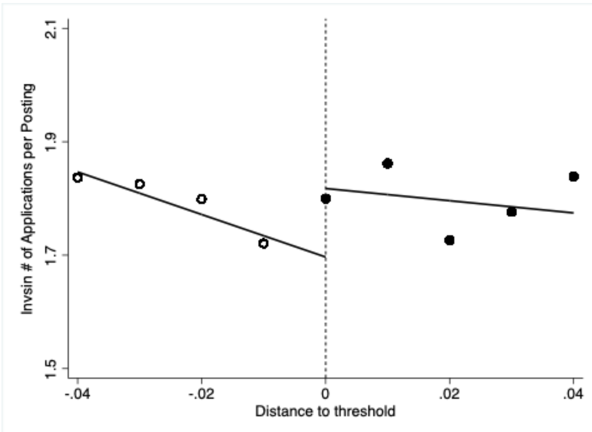


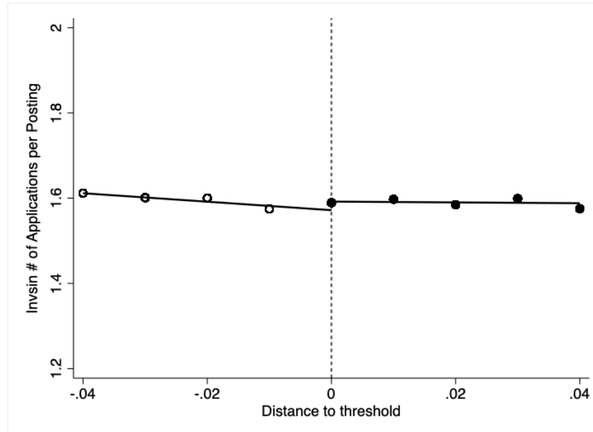
C: Fewer Employees

C: More Employees

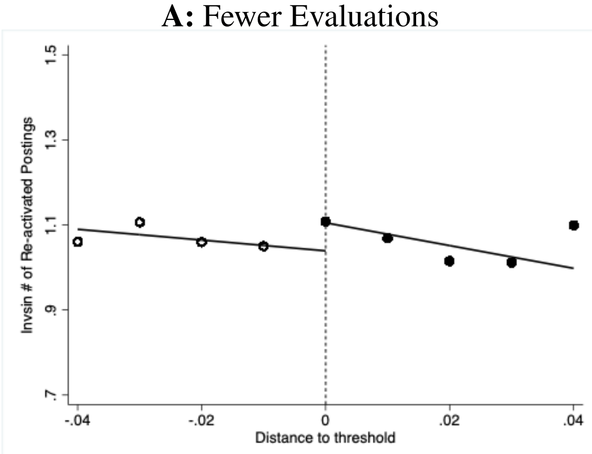
Notes: The figure plots the heterogeneous effect of displayed reputation by firm size. Each observation is the average weekly inverse hyperbolic sine of the number of applications per posting (Panel A and B) and the number of new posting (Panel C and D) for firms with fewer than 400 employees (Panel A and C) and firms with more than 400 employees (Panel B and D). Dashed vertical line denotes rounding threshold and is normalized to 0. The solid lines are estimated using a local linear regression with triangular weights and firm × week data following equation 1. Corresponding regression results can be found in Table **??**. Standard errors are calculated using Eicker–Huber–White standard errors.

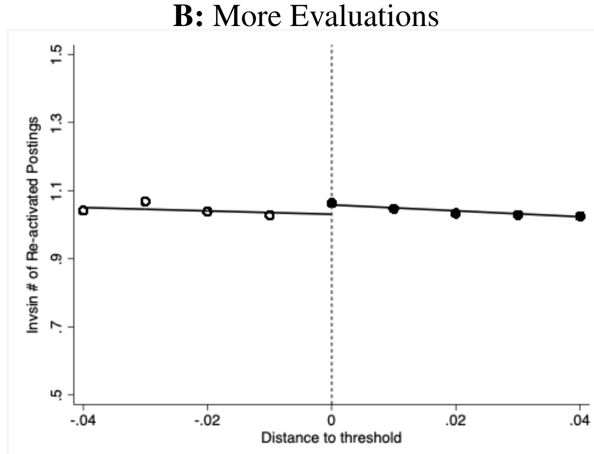
Figure 3.8: Heterogeneous Effect of Displayed Reputation by Firm Size





A: Fewer Evaluations **B:** More Evaluations





C: Fewer Evaluations **C:** More Evaluations

Notes: The figure plots the heterogeneous effect of displayed reputation by firms' number of evaluations received. Each observation is the average weekly inverse hyperbolic sine of the number of applications per posting (Panel A and B) and the number of new posting (Panel C and D) for firms with fewer cumulative evaluations (Panel A and C) and more cumulative evaluations (Panel B and D). Dashed vertical line denotes rounding threshold and is normalized to 0. The solid lines are estimated using a local linear regression with triangular weights and firm $\times$ week data following equation 1. Standard errors are calculated using Eicker–Huber–White standard errors.

Figure 3.9: Heterogeneous Effect of Displayed Reputation by Cumulative Evaluations

Table 3.1: Summary Statistics - Supply Side

	# of Reviews Received per Week	Vacacy Duration (All postings)	# of Views per Posting	# of Applications per Posting
By Sector				
Information Technology	1 (5.44)	40 (25)	55 (69)	5.5 (7.7)
Business Services	0.9 (4.2)	34 (18)	52 (59)	4.7 (6.3)
Finance	2.2 (8.4)	45 (18)	60 (69)	3.3 (3.9)
Manufacturing	1.1 (2.2)	46 (19)	61 (75)	3 (4.1)
By Employer Type				
Public company	4.6 (11.7)	42 (18)	61 (69)	3.3 (4.2)
Private company	0.5 (2.2)	38 (23)	54 (66)	5.1 (7.2)
By # of Employees				
>= 400	2.4 (7.5)	40 (19)	59 (70)	3.9 (5.5)
<400	0.2 (0.6)	39 (25)	52 (62)	5.2 (7.3)
Total	1.3 (5.5)	40 (22)	56 (67)	4.5 (6.5)

Notes: The table presents summary statistics for the demand side of our main regression samples. All variables are measured at the firm$\times$week level.

Table 3.2: Summary Statistics - Demand Side

	Obs (%)	Displayed Rating	# of New Postings	# of Re-activated Postings	# of Dropped or De-activated Postings
By Sector					
Information Technology	66,994 (37%)	3.53 (0.64)	5.83 (25)	7 (35)	7 (33)
Business Services	42,426 (23%)	3.63 (0.67)	10 (50)	13 (74)	13 (72)
Finance	9,419 (5%)	3.39 (0.53)	1.6 (5)	2 (7)	2 (7)
Manufacturing	8,213 (4.5%)	3.29 (0.51)	0.7 (2)	0.9 (2)	0.9 (2)
By Employer Type					
Public company	30,855 (17%)	3.3 (0.48)	10 (62)	14 (90)	13 (87)
Private company	119,156 (66.5%)	3.54 (0.66)	5 (19)	6 (25)	6 (24)
By # of Employees					
>= 400	92,689 (51%)	3.36 (0.53)	7 (40)	9 (56)	9 (54)
<400	89,195 (49%)	3.62 (0.69)	3 (16)	4 (22)	4 (22)
Total	181,884	3.49 (0.63)	5 (31)	7 (43)	6 (42)

Notes: The table presents summary statistics for the demand side of our main regression samples. All variables are measured at the firm×week level.

Table 3.3: Validity of RD - Discontinuity of Firm Characteristics

VARIABLES	(1) # of Employees	(2) Weekly # of Reviews	(3) Public	(4) Vacancy Duration (All postings)	(5) Vacancy Duration (Start Week)	(6) Post Wage Dummy (Start Week)
Rounded up	-152.780	0.096	0.006	0.083	0.417	-0.008
	(632.392)	(0.086)	(0.006)	(0.416)	(0.588)	(0.007)
Observations	163,346	163,346	163,346	163,346	77,284	77,284
Bandwidth	.04	.04	.04	.04	.04	.04
Mean (Level)	8487	1.333	0.170	39.44	33.88	0.159

Notes: The data are at the firm×week level consisting of firms in both Glassdoor.com and Dice.com and have raw employer reputation within 0.04 from the rounding thresholds. Rounded up(t) equals 1 if the firm is above a threshold at time t. All regressions are estimated using a linear RD model and triangular weights. Standard errors are calculated using Eicker–Huber–White standard errors.
** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 3.4: The Effect of Displayed Reputation on Firm Behavior

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Invsin # of New Postings			Invsin # of Re-activated Postings			Invsin # of Dropped or De-activated Postings		
Rounded up	0.043*	0.037*	0.043**	0.048**	0.043*	0.048**	0.038*	0.036	0.041*
	(0.022)	(0.022)	(0.020)	(0.023)	(0.023)	(0.021)	(0.022)	(0.022)	(0.021)
Observations	163,346	163,346	162,450	163,346	163,346	162,450	163,346	163,346	162,450
Adjusted R-squared	0.000	0.012	0.116	0.000	0.010	0.115	0.000	0.007	0.111
Bandwidth	.04	.04	.04	.04	.04	.04	.04	.04	.04
Mean (IHS)	0.956	0.956	0.957	1.053	1.053	1.054	1.056	1.056	1.057
Mean (Level)	5.096	5.096	5.117	6.351	6.351	6.377	6.283	6.283	6.310
Month FE		Yes	Yes		Yes	Yes		Yes	Yes
Sector FE			Yes			Yes			Yes
Threshold FE			Yes			Yes			Yes

Notes: The data are at the firm×week level consisting of firms in both Glassdoor.com and Dice.com and have raw employer reputation within 0.04 from the rounding thresholds. Rounded up(t) equals 1 if the firm is above a threshold at time t. All regressions are estimated using a linear RD model and triangular weights. Standard errors are calculated using Eicker–Huber–White standard errors.
** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3.5: The Effect of Displayed Reputation on Job Seeker Behavior

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Invsin # of Views per Posting			Invsin # of Applications per Posting		
Rounded up(t)	0.025	0.030***	0.029***	0.075***	0.074***	0.073***
	(0.044)	(0.010)	(0.010)	(0.017)	(0.017)	(0.016)
Observations	163,346	163,346	162,450	163,346	163,346	162,450
Adjusted R-squared	0.000	0.927	0.927	0.000	0.010	0.053
Bandwidth	.04	.04	.04	.04	.04	.04
Mean (IHS)	3.310	3.310	3.315	1.694	1.694	1.693
Mean (Level)	55.85	55.85	55.94	4.514	4.514	4.507
Month FE		Yes	Yes		Yes	Yes
Sector FE			Yes			Yes
Threshold FE			Yes			Yes

Notes: The data are at the firm×week level consisting of firms in both Glassdoor.com and Dice.com and have raw employer reputation within 0.04 from the rounding thresholds. Rounded up(t) equals 1 if the firm is above a threshold at time t. All regressions are estimated using a linear RD model and triangular weights. Standard errors are calculated using Eicker–Huber–White standard errors.
** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 3.6: The Effect of Displayed Reputation by Firm Type

VARIABLES	(1) Invsin # of New Postings	(2) Invsin # of Re-activated Postings	(3) Invsin # of Dropped or De-activated Postings	(4) Invsin # of Views per Posting	(5) Invsin # of Applications per Posting
Panel A: Public Company					
Rounded up	-0.008	-0.013	-0.046	0.004	0.005
	(0.048)	(0.051)	(0.050)	(0.020)	(0.030)
Observations	27,749	27,749	27,749	27,749	27,749
Adjusted R-squared	0.140	0.135	0.137	0.927	0.049
Mean (IHS)	1.035	1.160	1.172	3.510	1.527
Mean (Level)	9.532	13.30	13.18	60.67	3.295
Panel B: Private Company					
Rounded up	0.043	0.050*	0.056**	0.015	0.076***
	(0.027)	(0.028)	(0.028)	(0.011)	(0.022)
Observations	106,100	106,100	106,100	106,100	106,100
Adjusted R-squared	0.103	0.102	0.097	0.934	0.043
Mean (IHS)	1.054	1.152	1.153	3.238	1.800
Mean (Level)	4.747	5.570	5.492	54.57	5.130
Bandwidth	.04	.04	.04	.04	.04
Month FE	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes
Threshold FE	Yes	Yes	Yes	Yes	Yes

Notes: The data are at the firm×week level consisting of firms in both Glassdoor.com and Dice.com and have raw employer reputation within 0.04 from the rounding thresholds. Rounded up(t) equals 1 if the firm is above a threshold at time t. All regressions are estimated using a linear RD model and triangular weights. Standard errors are calculated using Eicker–Huber–White standard errors.
** p <0.01, ** p <0.05, * p <0.1.

Table 3.7: The Effect of Displayed Reputation by Firm Size

VARIABLES	(1) Invsin # of New Postings	(2) Invsin # of Re-activated Postings	(3) Invsin # of Dropped or De-activated Postings	(4) Invsin # of Views per Posting	(5) Invsin # of Applications per Posting
Panel A: # of Employees >= 400					
Rounded up	-0.010	-0.004	-0.023	-0.001	0.014
	(0.027)	(0.028)	(0.028)	(0.012)	(0.020)
Observations	83,451	83,451	83,451	83,451	83,451
Adjusted R-squared	0.168	0.167	0.164	0.924	0.061
Mean (IHS)	0.950	1.053	1.057	3.488	1.593
Mean (Level)	6.642	8.602	8.493	59.41	3.883
Panel B: # of Employees <400					
Rounded up	0.097***	0.101***	0.105***	0.059***	0.127***
	(0.030)	(0.031)	(0.031)	(0.016)	(0.026)
Observations	78,999	78,999	78,999	78,999	78,999
Adjusted R-squared	0.103	0.103	0.100	0.931	0.038
Mean (IHS)	0.965	1.055	1.058	3.132	1.799
Mean (Level)	3.507	4.027	4.003	52.26	5.166
Bandwidth	.04	.04	.04	.04	.04
Month FE	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes
Threshold FE	Yes	Yes	Yes	Yes	Yes

Notes: The data are at the firm$\times$week level consisting of firms in both Glassdoor.com and Dice.com and have raw employer reputation within 0.04 from the rounding thresholds. Rounded up(t) equals 1 if the firm is above a threshold at time t. All regressions are estimated using a linear RD model and triangular weights. Standard errors are calculated using Eicker–Huber–White standard errors.
** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 3.8: The Effect of Displayed Reputation by Stickiness

VARIABLES	(1) Invsin # of New Postings	(2) Invsin # of Re-activated Postings	(3) Invsin # of Dropped or De-activated Postings	(4) Invsin # of Views per Posting	(5) Invsin # of Applications per Posting
Panel A: Less Sticky (# of consecutive weeks with same rating <5)					
Rounded up	0.031	0.036	0.017	-0.013	0.013
	(0.026)	(0.027)	(0.027)	(0.012)	(0.021)
Observations	74,041	74,041	74,041	74,041	74,041
Adjusted R-squared	0.116	0.116	0.116	0.926	0.038
Mean (IHS)	0.838	0.922	0.874	3.485	1.652
Mean (Level)	4.378	5.385	5.180	59.85	4.258
Panel B: More Sticky (# of consecutive weeks with same rating >= 5)					
Rounded up	0.050	0.053*	0.053*	0.069***	0.131***
	(0.031)	(0.032)	(0.031)	(0.015)	(0.024)
Observations	88,409	88,409	88,409	88,409	88,409
Adjusted R-squared	0.111	0.108	0.095	0.929	0.069
Mean (IHS)	1.058	1.165	1.211	3.172	1.728
Mean (Level)	5.736	7.208	7.256	52.66	4.716
Bandwidth	.04	.04	.04	.04	.04
Month FE	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes
Threshold FE	Yes	Yes	Yes	Yes	Yes

Notes: The data are at the firm×week level consisting of firms in both Glassdoor.com and Dice.com and have raw employer reputation within 0.04 from the rounding thresholds. Rounded up(t) equals 1 if the firm is above a threshold at time t. All regressions are estimated using a linear RD model and triangular weights. Standard errors are calculated using Eicker–Huber–White standard errors.
** p <0.01, ** p <0.05, * p <0.1.

Table 3.9: The Effect of Changes in Displayed Reputation

VARIABLES	(1) Invsin # of New Postings	(2) Invsin # of Re-activated Postings	(3) Invsin # of Dropped or De-activated Postings	(4) Invsin # of Views per Posting	(5) Invsin # of Applications per Posting
Rounded up	0.183***	0.208***	0.151***	0.043***	0.032
	(0.028)	(0.029)	(0.029)	(0.014)	(0.022)
Observations	69,406	69,406	69,406	69,406	69,406
Adjusted R-squared	0.114	0.112	0.107	0.925	0.054
Mean (IHS)	0.958	1.054	1.059	3.329	1.697
Mean (Level)	5.224	6.547	6.403	56.51	4.531
Bandwidth	.04	.04	.04	.04	.04
Month FE	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes
Threshold FE	Yes	Yes	Yes	Yes	Yes

Notes: The data are at the firm×week level consisting of firms in both Glassdoor.com and Dice.com, have raw employer reputation within 0.04 from the rounding thresholds at time t, and have raw employer reputation within 0.04 below the rounding thresholds at time $t-1$. Rounded up(t) equals 1 if the firm is above a threshold at time t. All regressions are estimated using a linear RD model and triangular weights. Standard errors are calculated using Eicker–Huber–White standard errors.
** p <0.01, ** p <0.05, * p <0.1.

Table 3.10: Job-level Effect of Displayed Reputation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Invsin # of Views				Invsin # of Applications			
Rounded up	-0.068***	-0.040	-0.029	-0.028	-0.004	0.001	-0.019	-0.019
	(0.023)	(0.032)	(0.022)	(0.022)	(0.036)	(0.037)	(0.022)	(0.022)
Observations	3,531,988	3,531,988	3,528,732	3,528,732	5,136,718	5,136,718	5,130,261	5,130,261
Adjusted R-squared	0.001	0.022	0.048	0.078	0.000	0.004	0.034	0.045
Bandwidth	.04	.04	.04	.04	.04	.04	.04	.04
Mean	3.237	3.237	3.237	3.237	0.594	0.594	0.594	0.594
Month FE		Yes	Yes	Yes		Yes	Yes	Yes
Sector FE			Yes	Yes			Yes	Yes
Threshold FE			Yes	Yes			Yes	Yes
Days online FE				Yes				Yes

Notes: The data are at the job$\times$week level consisting of firms in both Glassdoor.com and Dice.com and have raw employer reputation within 0.04 from the rounding thresholds. Rounded up(t) equals 1 if the job posting is from a firm above a threshold at time t. All regressions are estimated using a linear RD model and triangular weights. Standard errors are calculated using Eicker–Huber–White standard errors.
** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3.11: RDD with Alternative Bandwidths

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	Invsin # of New Postings		Invsin # of Re-activated Postings		Invsin # of Dropped or De-activated Postings		Invsin # of Views per Posting		Invsin # of Applications per Posting	
Rounded up	0.062***	0.060*	0.068***	0.066**	0.058**	0.050	0.041***	0.055***	0.092***	0.124***
	(0.024)	(0.032)	(0.025)	(0.033)	(0.025)	(0.033)	(0.012)	(0.016)	(0.019)	(0.026)
Observations	126,556	89,340	126,556	89,340	126,556	89,340	126,556	89,340	126,556	89,340
Adjusted R-squared	0.119	0.122	0.117	0.120	0.113	0.116	0.928	0.930	0.054	0.055
Mean (IHS)	0.955	0.953	1.053	1.052	1.056	1.054	3.306	3.289	1.688	1.683
Mean (Level)	5.069	5.007	6.349	6.249	6.285	6.222	55.68	55.25	4.450	4.412
Bandwidth	.03	.02	.03	.02	.03	.02	.03	.02	.03	.02
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Threshold FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Appendix A

Appendix For Chapter One

A.1 Rescaled Google Search Index

Google Trends reports the search index in either a time series or a cross-sectional format. To construct a panel data set for each media market and time, we need to extract the search index in each media market separately. However, the search index reported by Google Trends is the search rate normalized by the maximum search rate in an extraction and is not comparable across extractions. To build a panel of search indexes that are normalized by the same base, we rescale the search index using the following method.

In a time series extraction of the search index in media market m over period T, the search index in median market m at time t is approximately

$$SearchIndex_{mt,T} = 100 \times \frac{\frac{Searchesincluding``chink(s)''_{mt}}{Totalsearches_{mt}}}{\max\limits_{t \in T}\{\frac{Searchesincluding``chink(s)''_{mt}}{Totalsearches_{mt}}\}} \tag{A.1}$$

Meanwhile, in a cross-sectional extraction of the search index at time t for all media markets

$m \in M$, the search index in media market m at time t is approximately

$$SearchIndex_{mt,M} = 100 \times \frac{\frac{Searchesincluding``chink(s)''_{mt}}{Totalsearches_{mt}}}{\max\limits_{m \in M}\{\frac{Searchesincluding``chink(s)''_{mt}}{Totalsearches_{mt}}\}} \tag{A.2}$$

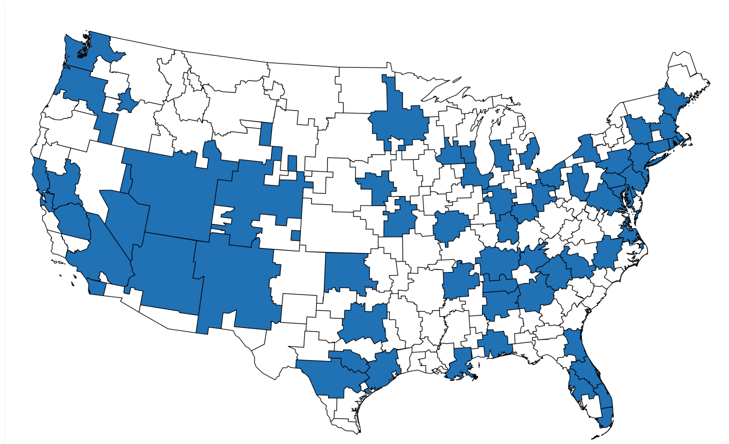
If we are willing to assume that the numerators in equations A.1 and A.2 are the same, then we can calculate the ratio of the two denominators as

$$Ratio_{m.MT} = \frac{\max\limits_{t \in T}\{\frac{Searchesincluding``chink(s)''_{mt}}{Totalsearches_{mt}}\}}{\max\limits_{m \in M}\{\frac{Searchesincluding``chink(s)''_{mt}}{Totalsearches_{mt}}\}} = \frac{SearchIndex_{mt,M}}{SearchIndex_{mt,T}} \tag{A.3}$$

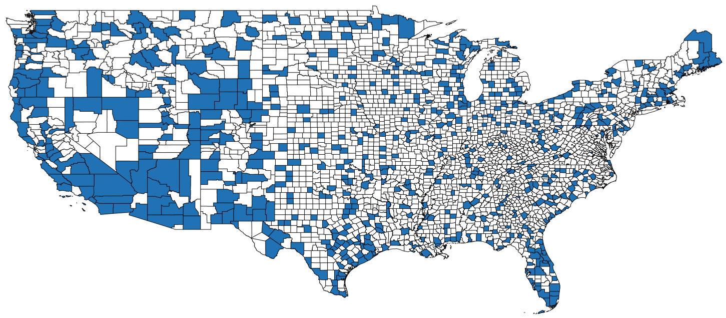
when both search indexes are non-zero. We can scale the time series search index over period T in each media market $m \in M$ by multiplying it with the corresponding $Ratio_{m,MT}$. The resulting time series are normalized by the same $\max\limits_{m \in M}\{\frac{Searchesincluding``chink(s)''_{mt}}{Totalsearches_{mt}}\}$. However, Google Trends returns a zero value when the absolute level of search in a given media market and time is below an unreported threshold, under which the rescaling does not work. After extracting cross-sectional search indexes on all possible weeks in the sample period, we can at best back out the rescaled search index for 35 media markets using Huntsville-Decatur (Florence) media market's search rate on March 15, 2020, as the base. Alternatively, we can back out 29 media markets using Wilkes Barre-Scranton media market's search rate on March 29, 2020, and 29 media markets using Buffalo media market's search rate on April 5, 2020, as the base. When combined, these three measures cover 50 media markets.

Note that Google calculates the search index using a random sample of searches, which can be different across extractions. As a result, the numerators in equations A.1 and A.2 are similar but may not be exactly the same. To the extent that these two numerators are not the same, we may be introducing measurement errors to the dependent variable and attenuating the main effects.

A.2 Additional Figures & Tables



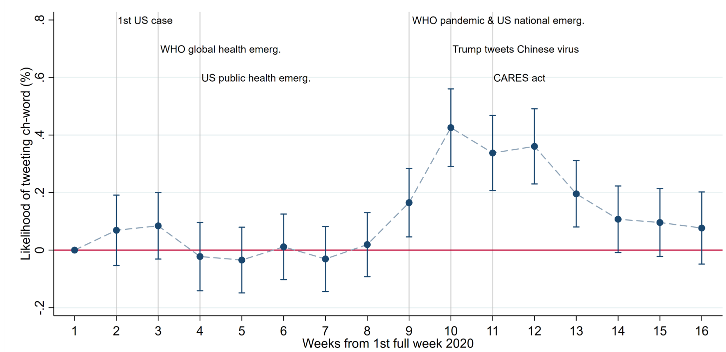
A: Google media market



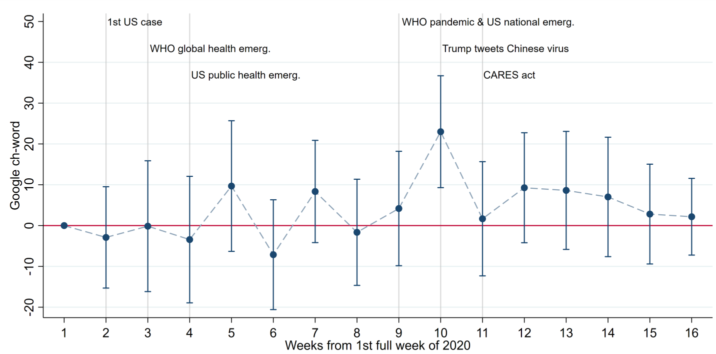
B: Twitter county

Note: The figure plots the locations of media markets with Google data (panel A) and counties with Twitter data (panel B).

Figure A.1: Location of Media Markets and Counties with Data



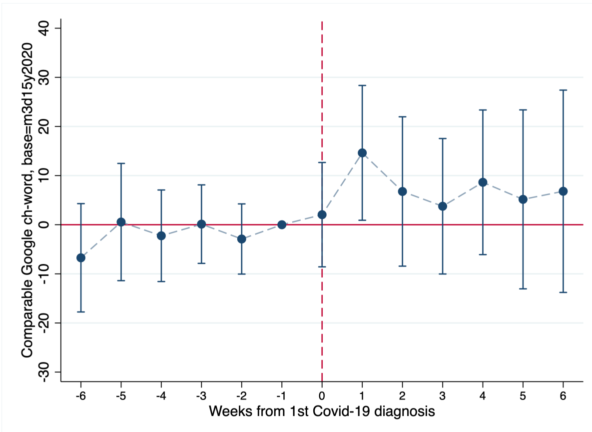
A: Likelihood of tweeting the ch-word



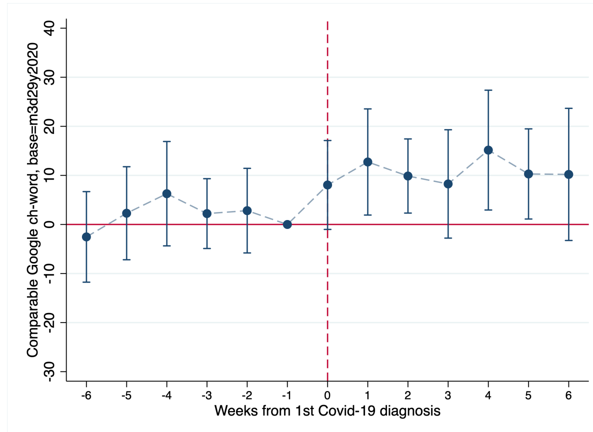
B: Google search index

Note: The figure presents the relationship between the timeline of important COVID-19 developments and the evolution of racial animus in the United States. Panels A and B plot the estimates and 95 percent confidence intervals of the coefficients on the event dummies in equation 1.2 using user's likelihood of tweeting the ch-word and the racially charged Google search index as the outcome, respectively. Regressions control for week-of-year fixed effects and user fixed effects (panel A) or media market fixed effects (panel B). Standard errors are clustered by user (panel A) or by media market (panel B).

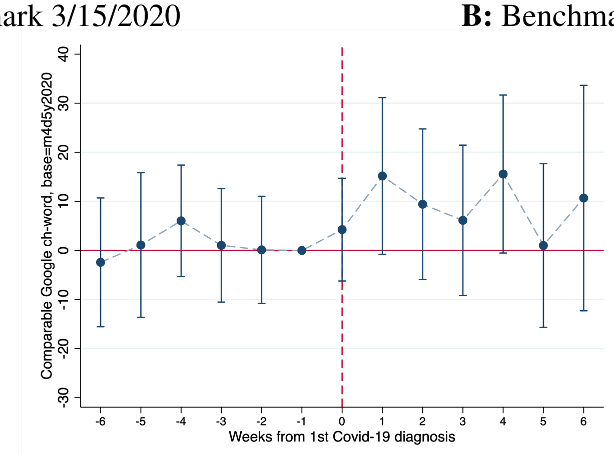
Figure A.2: Timeline of COVID-19 Developments and Evolution of Racial Animus



A: Benchmark 3/15/2020



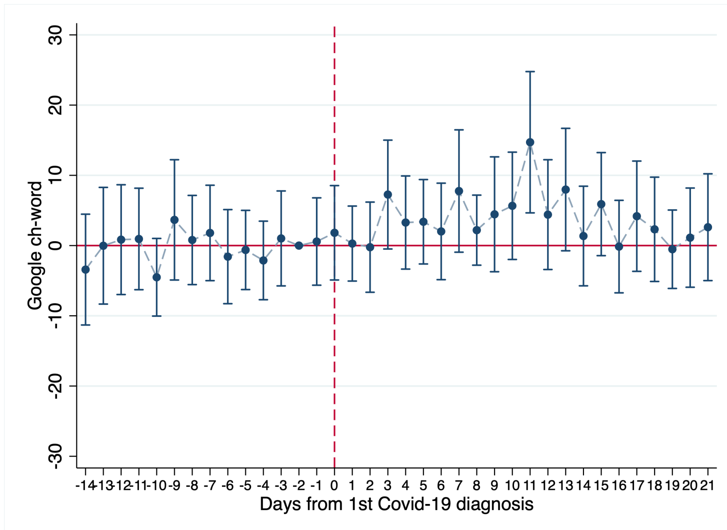
B: Benchmark 3/29/2020



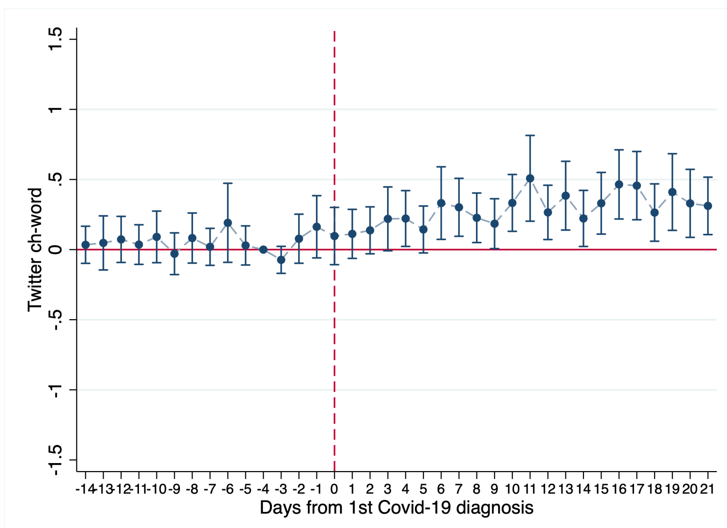
C: Benchmark 4/5/2020

Notes: The figure presents the effect of the first local COVID-19 diagnosis on various *rescaled* racially charged Google search indexes. Panels A, B, and C plot the estimates and 95 percent confidence intervals of the coefficients on the event dummies in equation 3.3 using an area's racially charged Google search rate scaled by Huntsville-Decatur (Florence) media market's search rate on March 15, 2020, by Wilkes Barre-Scranton media market's search rate on March 29, 2020, and by Buffalo media market's search rate on April 5, 2020 as the outcome, respectively. See A.1 for the definitions of these indexes. Specifications mirror those in column (1) of Table 1.2.

Figure A.3: The Effect of the First Local COVID-19 Diagnosis on Racial Animus
Rescaled Google Search Index



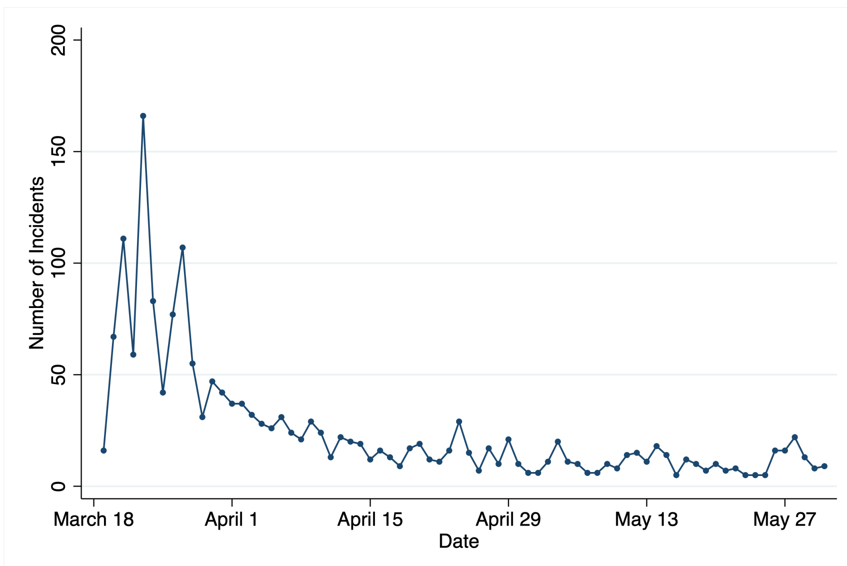
A: Google search index



B: Twitter post index

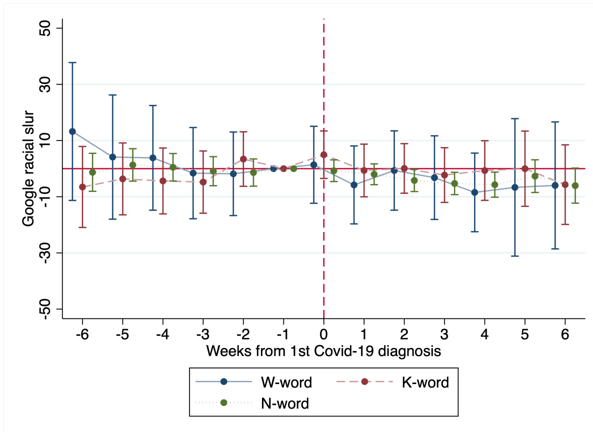
Note: The figure presents the effect of the first local COVID-19 diagnosis on the *daily* racially charged Google search index and Twitter post index. Panels A and B plot the estimates and 95 percent confidence intervals of the coefficients on the event dummies in equation 3.3 using the Google search index and the Twitter post index as the outcome, respectively. Regressions control for year-month fixed effects, day-of-week fixed effects, and media market fixed effects (panel A) or county fixed effects (panel B). Standard errors are clustered by media market (panel A) or by county (panel B).

Figure A.4: The Effect of the First Local COVID-19 Diagnosis on *Daily* Racial Animus

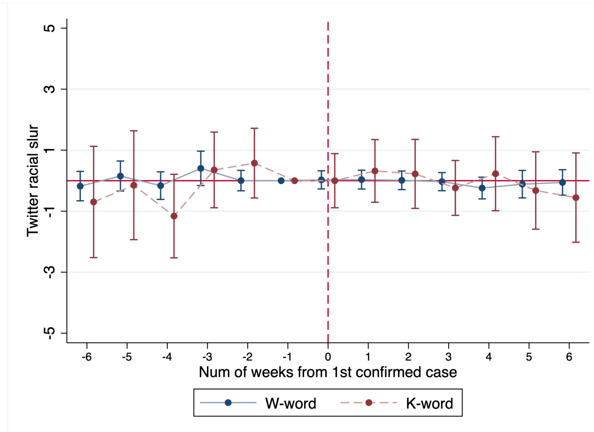


Notes: This figure presents the daily number of hate incidents from AP3CON Stop AAPI Hate Reporting system between March 19, 2020 (start of the data) and September, 2020.

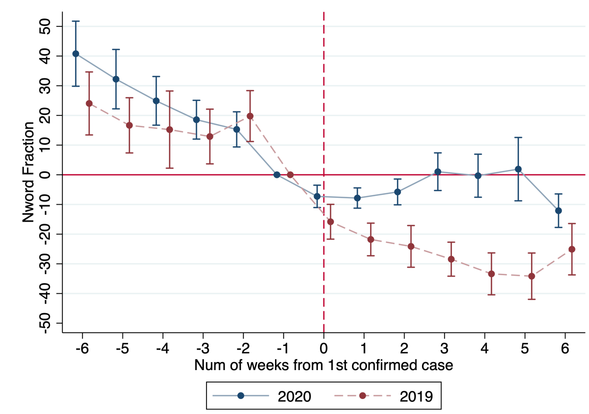
Figure A.5: Self-Reported Hate Incidents in the United States



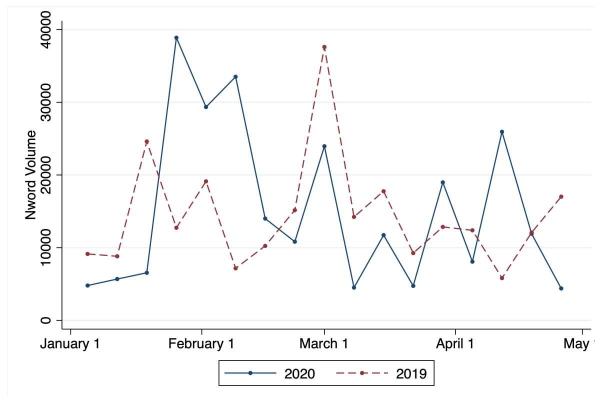
A: Google search index



B: Twitter post index



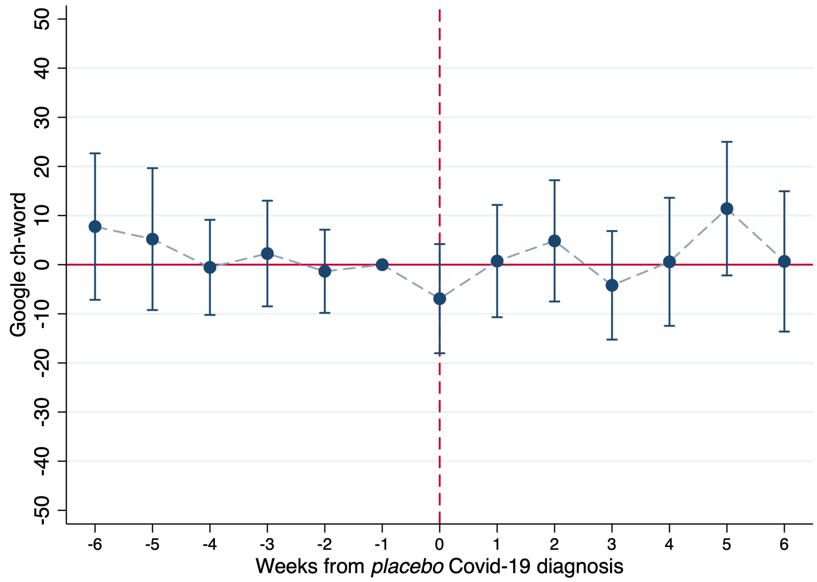
C: Twitter N-word



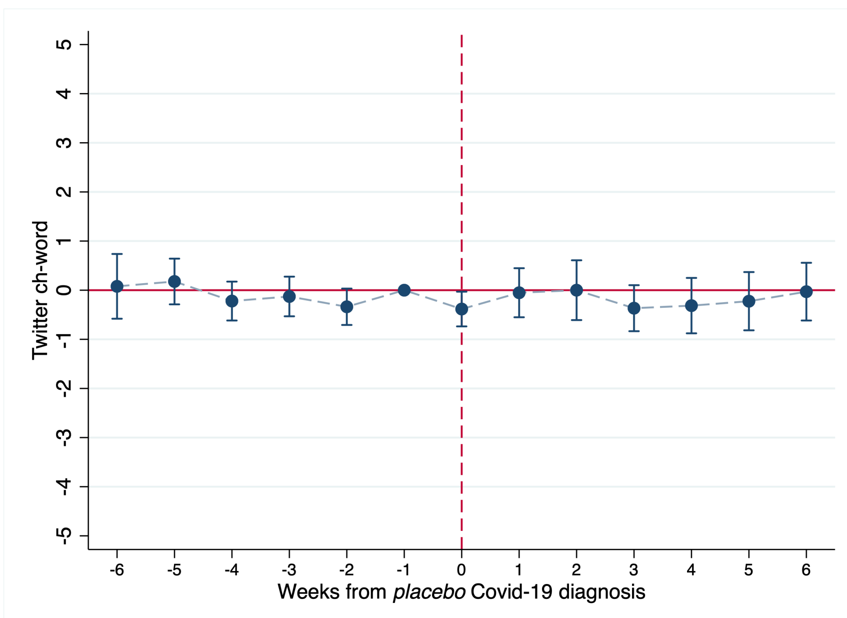
D: N-word trend

Notes: The figure presents the effect of the first local COVID-19 diagnosis on racial animus against the Hispanic, Jewish, and African American populations, using the Google search index and Twitter post index for"wetback(s)", "kike(s)", and the n-word as proxies. The indexes are defined following the method outlined in section 1.2.1. Regression samples for the n-word, k-word, and w-word Google search indexes contain 203, 78, and 27 media markets (panel A). Regression samples for the w-word and k-word Twitter post indexes contain 599 counties (panel B). Estimates of the coefficients and the 95 percent confidence intervals of the event dummies are from estimating equation 3.3 using the above indexes as outcomes. We include an indicator for the week of January 26, 2020 in the regression for the n-word to control for a spike in its use due to Kobe Bryant's death and MSNBC's anchor using the n-word while reporting the news. We include an indicator for the week of February 23, 2020 in the regression for the k-word to control for a spike in its use due to the Los Angeles Dodges player Enrique ("Kiké") Hernandez's performance in that week. All other specifications in panels A and B mirror those in column (1) of Table 1.2 and column (1) of Table 1.3, respectively. Panel C plots the estimates and 95 percent confidence intervals of the coefficients on event dummies in equation 3.3 using the Twitter post index for the n-word between November 2019 and April 2020 (blue line) and that between November 2018 and April 2019 (red line) as the outcomes. For the regression using the 2018-2019 data, we replace the date of the first local COVID-19 diagnosis with a placebo date which shares the same day and month as the actual date in 2020 but with the year as 2019. For the regression using 2019-2020 data, we include an indicator for the week of January 26, 2020 to control for Kobe Bryant's death on January 26, 2020 and an indicator for the week of February 9, 2020 to control for an extremely viral video tweet *unrelated* to COVID-19 but mentioning the n-word on February 10, 2020. Panel D plots time trends for the Twitter post index for the n-word in 2020 (blue line) and in 2019 (red line).

Figure A.6: The Effect of the First Local COVID-19 Diagnosis on Racial Animus against *Non-Asian* Minorities



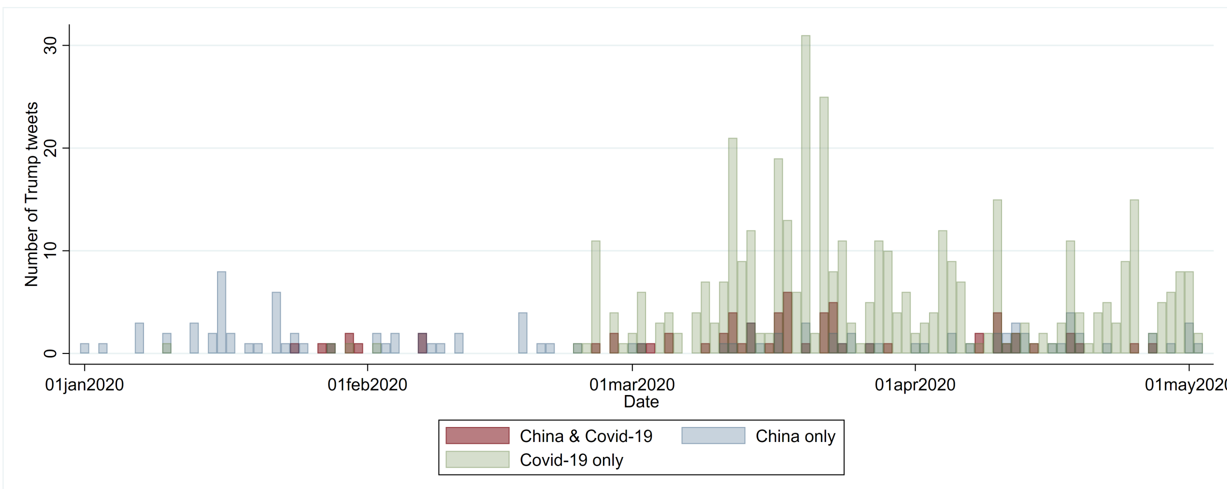
A: Google search index



B: Twitter post index

Note: The figure presents a placebo test for the effect of the first local COVID-19 diagnosis on the racially charged Google search index and Twitter post index. We replace the date of the first local COVID-19 diagnosis with a placebo date using the same calendar day and month of the actual diagnosis date but changing the year from 2020 to 2019. Panels A and B plot the estimates and 95 percent confidence intervals of the coefficients on the event dummies in equation 3.3 using the Google search index and the Twitter post index as the outcome, respectively. Specifications in panels A and B mirror those in column (1) of Table 1.2 and column (1) of Table 1.3, respectively.

Figure A.7: The Effect of the First Local COVID-19 Diagnosis on Racial Animus Placebo Test



Notes: This figure plots the number of President Trump's tweets by category on each day between January 1, 2020 and May 2, 2020. We categorize the president's tweets that include "china", "chinese", "huawei", "xi", "COVID", "COVID-19", "corona", "coronavirus", "virus", "epidemic", or "pandemic" into three categories: those mentioning only China (China only), only COVID-19 (COVID only), and both China and COVID-19 (China-and-COVID).

Figure A.8: Number of President Trump's Tweets about China or COVID-19

Table A.1: Sample Selection - Media Markets and Counties with Google and Twitter data

VARIABLES	(1) Google sample	(2) Twitter data	(3) Twitter sample
Log(pop)	0.193***	0.142***	0.142***
	(0.036)	(0.007)	(0.007)
% Asian	0.059**	0.007	0.007
	(0.030)	(0.009)	(0.009)
% Asian^2	-0.002**	-0.001**	-0.001**
	(0.001)	(0.000)	(0.000)
% Male	0.001	-0.001	-0.001
	(0.037)	(0.002)	(0.002)
% 65+	-0.013	-0.002	-0.003
	(0.015)	(0.002)	(0.002)
% BA+	0.009	0.002	0.002*
	(0.007)	(0.001)	(0.001)
% Unemp	-0.003	0.002	0.002
	(0.020)	(0.005)	(0.005)
% Vote share dem-rep	-0.001	0.001***	0.001***
	(0.002)	(0.000)	(0.000)
Hate crime/1m	-0.087***	0.001	0.001
	(0.026)	(0.005)	(0.005)
Intl airport enplanement	0.006*	0.006**	0.006**
	(0.003)	(0.002)	(0.002)
Observations	205	3,111	3,111
R-squared	0.695	0.341	0.343
Outcome mean	0.292	0.205	0.202

Notes: The table presents the sample selection in Google and Twitter data. The data are at the media market level in column (1) and at the county level in columns (2) and (3). The outcome is an indicator for having Google data in column (1), an indicator for having Twitter data in column (2), and an indicator for being in the Twitter regression sample in column (3). Note that all media markets with Google data are in the Google regression sample. "Log(pop)" is the natural log of local population estimates in 2018 from Census Bureau. "%Asian", "% Male", "% 65+", and "% BA+" are the percentage of Asians, males, population 65 years old or over, and population with Bachelor's or above degree in the local area from American Community Survey 2014-2018 five-year average. "%Unemp" is the average monthly local unemployment rate between 2014 and 2018 from the BLS. "% Vote share Dem-Rep" is the difference between the Democratic and the Republican vote shares in the 2012 presidential election from Harvard Dataverse. "Hate crime/1m" is the average annual number of anti-Asian hate crimes per million population between 2014-2018 from UCR. "Intl airport enplanement" is the international airport enplanements in 2016 according to the Federal Aviation Administration. The number of media markets and counties is less than 210 and 3141 due to missing covariates. All regressions control for state fixed effects. Standard errors in parentheses are clustered by media market in column (1) and by county in columns (2) and (3).
** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.2: Summary Statistics

VARIABLE	(1) Google Sample	(2) Twitter Sample
Ch-word index		
Weekly	25.213	0.591
	(29.008)	(2.572)
Daily	6.965	0.309
	(20.316)	(2.215)
Hourly	-	0.357
	-	(0.676)
Other indexes (weekly)		
N-word	31.781	33.448
	(25.042)	(65.104)
K-word	34.183	2.970
	(26.750)	(12.705)
W-word	29.452	0.436
	(29.702)	(2.708)
Asian(s)	79.305	138.026
	(10.815)	(199.423)
Other animus measures		
Anti-Asian hate crime/1m	0.037	0.003
	(0.099)	(0.036)
Chinese restaurant visits/1m	26353	23846
	(13148)	(12328)
Total restaurant visits/1m	698341	606620
	(189770)	(220331)
Geographic unit	Media market	County
Unique geo-units	60	641

Notes: The table presents summary statistics for our main regression samples. See section 1.2.1 for the definitions of Google search index and Twitter post index. “Anti-Asian hate crime/1m” is the monthly anti-Asian hate crimes per million population in a media market between January 2014 and December 2018. “Chinese (or total) restaurant visits/1m” is the monthly visits to Chinese (or all) restaurants per million population in a median market between January 2018 and December 2019. All other variables are measured at the media market×time level in column (1) and at the county×time level in column (2).

Table A.3: Timing of the First Local COVID-19 Diagnosis - Weeks from Jan 19, 2020

VARIABLES	(1) Google sample Weeks from Jan192020	(2) Twitter sample Weeks from Jan192020
Log(pop)	-1.499***	-0.673***
	(0.474)	(0.053)
% Asian	0.156	-0.018
	(0.212)	(0.038)
% Asian^2	-0.004	-0.001
	(0.006)	(0.001)
% Male	-1.158***	0.019
	(0.414)	(0.033)
% 65+	-0.031	0.004
	(0.067)	(0.014)
% BA+	0.048	-0.005
	(0.040)	(0.007)
% Unemp	0.623***	0.061
	(0.212)	(0.044)
% VS dem-rep	-0.022**	-0.000
	(0.009)	(0.002)
Hate crime/1m	-0.850	-0.023
	(0.790)	(0.034)
Intl airport enplanement	0.035**	-0.028
	(0.015)	(0.017)
Observations	60	630
R-squared	0.984	0.646
Outcome mean	5.983	8.12

Notes: The table presents the relationship between the timing of the first local COVID-19 diagnosis and the characteristics of the local area. The data are at the media market level in column (1) and at the county level in column (2). The outcome is the number of weeks from the week of the first diagnosis in the United States, i.e., the week of January 19, 2020. See note to Table A.1 for variable definitions. The number of observation in column (2) is smaller than 641 due to missing covariates. All regressions control for state fixed effects. Standard errors are clustered by media market in column (1) or by county in column (2).
** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.4: Number of Media Markets and Counties by Week of the First COVID-19 Local Diagnoses

Date of Sunday	(1) Media markets	(2) Counties
Jan 19, 2020	2	
Jan 26, 2020	4	5
Feb 9, 2020	1	1
Feb 16, 2020	1	2
Feb 23, 2020	1	1
Mar 1, 2020	20	28
Mar 8, 2020	31	148
Mar 15, 2020		229
Mar 22, 2020		139
Mar 29, 2020		58
Apr 05, 2020		16
Apr 12, 2020		7
Apr 19, 2020		4
Total	60	641

Notes: The table presents the number of media markets and counties in our main regression samples by the week of their first local COVID-19 diagnoses.

Table A.5: Characteristics of First-Time Ch-word Users and Control Users

	Ch-word users		Control users	
	Mean	SD	Mean	SD
Panel A: User characteristics				
Account years	6.309	3.652	5.708	3.220
Followers	4103.629	57537.343	1330.260	9672.603
Followings	1440.742	3509.394	788.713	1920.440
Panel B: Prob. (re)tweet/reply/mention				
During pandemic:				
COVID conspiracy	0.017	0.131	0.004	0.063
Before pandemic:				
Anti-minority content	0.164	0.371	0.011	0.104
Anti-Asian user	0.866	0.341	0.228	0.419
Trump	0.561	0.496	0.157	0.364
McCarthy	0.137	0.344	0.004	0.063
McConnell	0.024	0.154	0.002	0.045
Pelosi	0.227	0.419	0.015	0.122
Schumer	0.178	0.383	0.010	0.101
Fox	0.282	0.450	0.027	0.161
CNN	0.388	0.487	0.045	0.207
CBS	0.115	0.319	0.009	0.094
N users	3,033		3,000	

Notes: This table presents the characteristics of first-time ch-word users and control users. Panel A reports information from Twitter user profiles. Panel B reports a user's likelihood of mentioning certain keywords in their tweets or interacting with certain users. "During pandemic" and "Before pandemic" refer to the period between January 21, 2020 and May 2, 2020 and that before January 21, 2020, respectively. See note to Figure 1.4 for definitions of the remaining variables.

Table A.6: Predictors of First-time Ch-word Users - Twitter Activity

VARIABLES	(1) First-time ch-word user
Anti-Asian user	0.558***
	(0.013)
Anti-minority	0.200***
	(0.012)
COVID consp.	0.004
	(0.046)
Trump	0.055***
	(0.014)
McCarthy	0.056***
	(0.016)
McConnell	-0.078**
	(0.031)
Pelosi	0.050***
	(0.019)
Schumer	0.010
	(0.020)
CBS	0.039**
	(0.017)
CNN	0.123***
	(0.018)
Fox	0.042**
	(0.019)
Account years	-0.017***
	(0.002)
Log(followers)	-0.001
	(0.004)
Log(followings)	-0.022***
	(0.005)
Observations	6,033
R-squared	0.465
Outcome mean	.502

Notes: This table reports the regression coefficients plotted in Figure 1.4 panel A. See note to the figure for variable definitions and regression specifications.
** p <0.01, ** p <0.05, * p <0.1.

Table A.7: Predictors of First-time Ch-word Users - User Profile Keywords

VARIABLES	(1) First-time ch-word user
artist	-0.171***
	(0.043)
business	-0.088*
	(0.052)
can	0.085*
	(0.050)
dad	-0.028
	(0.053)
enthusiast	-0.084
	(0.054)
fan	-0.087**
	(0.034)
father	-0.057
	(0.051)
games	-0.247***
	(0.054)
good	0.057
	(0.052)
hehim	0.034
	(0.038)
husband	-0.096*
	(0.054)
just	-0.043
	(0.035)
life	-0.058
	(0.036)
like	-0.067
	(0.045)
love	-0.060*
	(0.032)
lover	-0.092**
	(0.041)
mom	-0.055
	(0.046)
music	-0.044
	(0.044)
new	0.142***
	(0.046)
one	-0.038
	(0.053)
opinions	0.011
	(0.051)
politics	0.143***
	(0.051)
proud	-0.009
	(0.050)
retired	-0.018
	(0.050)
sheher	0.025
	(0.037)
sports	-0.076
	(0.049)
things	0.012
	(0.047)
time	-0.088*
	(0.048)
trump	0.291***
	(0.049)
twitch	-0.311***
	(0.041)
wife	-0.163***
	(0.049)
world	-0.018
	(0.048)
writer	0.008
	(0.042)
Account years	0.009***
	(0.002)
Log(followers)	0.036***
	(0.005)
Log(followings)	0.020***
	(0.007)
Observations	5,266
R-squared	0.072
Outcome mean	.502

Notes: This table reports the regression coefficients plotted in Figure 1.4 panel B. See note to the figure for variable definitions and regression specifications.
** p <0.01, ** p <0.05, * p <0.1.

Table A.8: Examples of President Trump's Tweets about China or COVID-19

Category	Post	Date
China-only	"Years from now, when we look back at this day, nobody's going to remember nancy's cheap theatrics, they will remember though how president trump brought the Chinese to the bargaining table and delivered achievements few ever thought were possible." @ingrahamangle @foxnews	1/17/20
China-only	The Wall Street Journal editorial board doesn't have a clue on how to fight and win. Their views on tariffs & trade are losers for the U.S., but winners for other countries, including China. If we followed their standards, we'd have no country left. They should love sleepy joe!	4/11/20
COVID-only	The coronavirus is very much under control in the USA. we are in contact with everyone and all relevant countries. CDC & World Health have been working hard and very smart. Stock market starting to look very good to me!	2/24/20
COVID-only	I am fully prepared to use the full power of the federal government to deal with our current challenge of the coronavirus!	3/11/20
China-and-COVID	Just received a briefing on the Coronavirus in china from all of our great agencies, who are also working closely with china. we will continue to monitor the ongoing developments. We have the best experts anywhere in the world, and they are on top of it 24/7!	1/30/20
China-and-COVID	I will be having a news conference today to discuss very important news from the FDA concerning the Chinese Virus!	3/18/20
China-and-COVID	Just finished a very good conversation with President Xi of China. Discussed in great detail the Coronavirus that is ravaging large parts of our planet. China has been through much & has developed a strong understanding of the virus. We are working closely together. Much respect!	3/22/20

Notes: This table presents examples of President Trump's tweets mentioning China and/or COVID-19. We manually categorize all President Trump's tweets between January 1, 2020 and May 2, 2020 that contain any of the words "china", "chinese", "huawei", "xi", "covid", "covid-19", "corona", "coronavirus", "virus", "epidemic", or "pandemic" into three categories: those mentioning only China (China-only), only COVID-19 (COVID-only), and both China and COVID-19 (China-and-COVID).

Table A.9: Relationship between Racial Animus Nationwide and Tweets from Politicians and National News Outlets

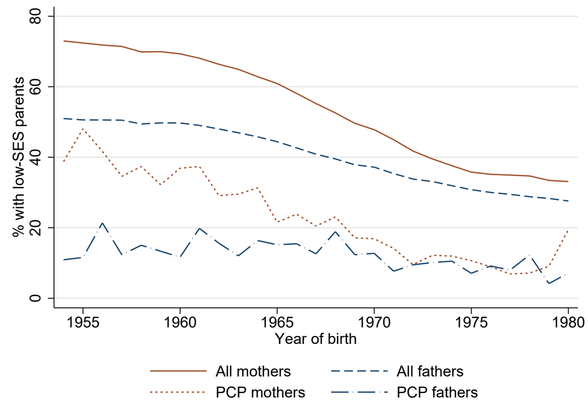
VARIABLES	(1) McCarthy	(2) McConnell	(3) Pelosi	(4) Schumer	(5) CBS	(6) CNN	(7) Fox
Panel A: Twitter post index							
China-and-COVID(t)	0.0328				-0.0131	0.0278*	0.0283
	(0.0594)				(0.0237)	(0.0155)	(0.1858)
China only(t)	0.0076			0.0446	-0.0203	0.0027	
	(0.0201)			(0.0391)	(0.0205)	(0.0132)	
Covid only(t)	0.0170	-0.0213*	-0.0087	-0.0304	0.0063***	-0.0012	0.0235
	(0.0109)	(0.0128)	(0.0226)	(0.0236)	(0.0018)	(0.0015)	(0.0223)
Observations	123	123	123	123	123	123	123
R-squared	0.5118	0.5088	0.4916	0.5057	0.5383	0.5090	0.5074
Outcome mean	.3445	.3445	.3445	.3445	.3445	.3445	.3445
Panel B: Log(hate incidents)							
China-and-COVID(t)	-0.5740*				-0.0175	-0.1768	0.2794
	(0.3287)				(0.1326)	(0.1365)	(0.2438)
China only(t)	-0.0090				0.2764	0.4743**	
	(0.0734)				(0.1714)	(0.2041)	
Covid only(t)	0.0426	-0.0081	0.0077	-0.0745	0.0031	0.0075	-0.0310
	(0.0553)	(0.0265)	(0.1218)	(0.0932)	(0.0052)	(0.0059)	(0.0690)
Observations	45	45	45	45	45	45	45
R-squared	0.8522	0.8198	0.8196	0.8236	0.8348	0.8611	0.8264
Outcome mean	3.1932	3.1932	3.1932	3.1932	3.1932	3.1932	3.1932

Notes: The table presents the relationship between the number of the tweets of major politician and news outlets about COVID-19 or China and racial animus nationwide. The outcome variable in panel A is the daily number of ch-word tweets per 100,000 "the" tweets nationwide between January 1, 2020 and May 2, 2020. The outcome variable in panel B is the natural log of the daily number of anti-Asian hate incidents nationwide from AP3CON Stop AAPI Hate Reporting system between March 19 and May 2, 2020. See note to Table 1.5 for definitions of the independent variables. All regressions control for year-week fixed effects and day-of-week fixed effects. Standard errors are clustered by date.
** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

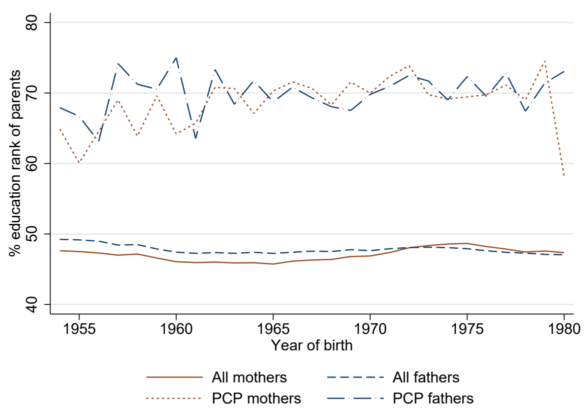
Appendix B

Appendix For Chapter Two

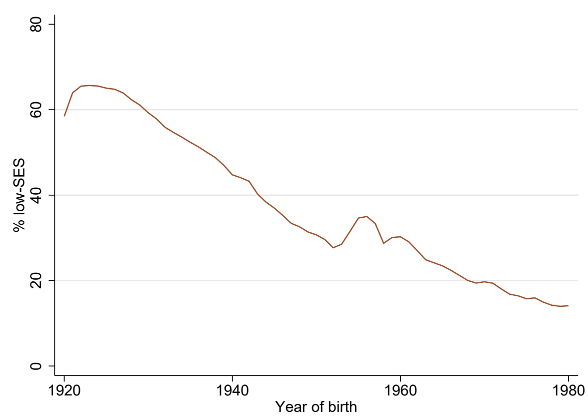
B.1 Additional Figures & Tables



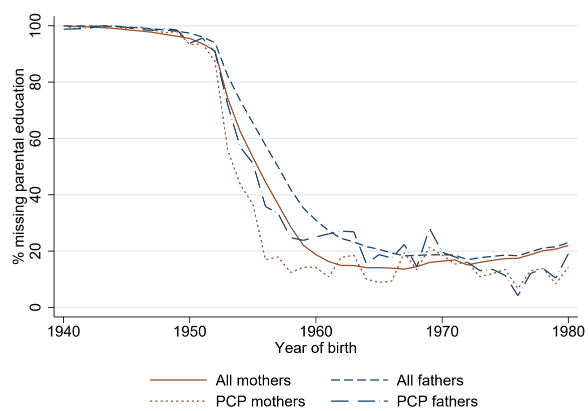
A: Share with a parent with primary school education



B: Parents' educational rank



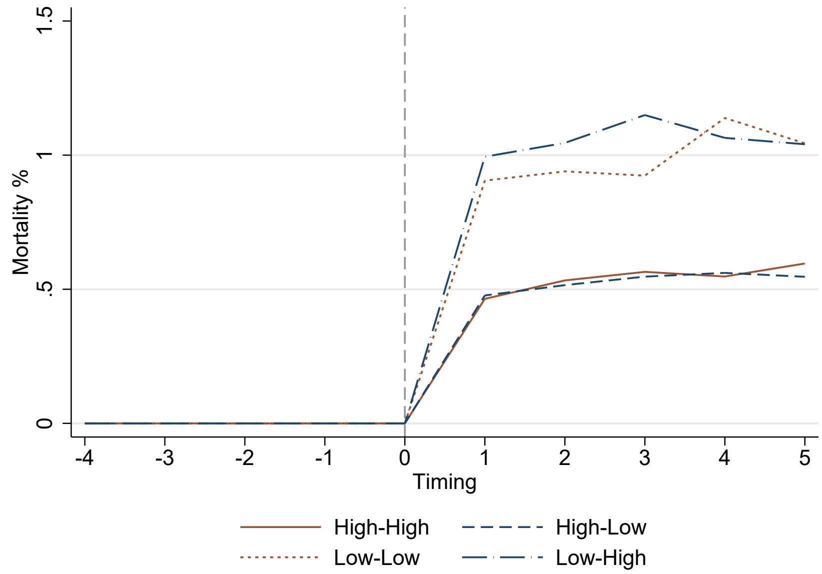
C: Share of population with primary school



D: Share with missing parent information

Note: The figure plots sample means by sub-populations.

Figure B.1: Summary Statistics on the Total Population and Physicians (PCP) by Birth Cohort



Note: The figure presents the raw correlation between physician-patient SES concordance on mortality.

Figure B.2: The correlations between Physician-patient SES Concordance and Mortality

Table B.1: ICD-10 Codes Used to Identify Cause of Death

	ICD-10 codes
Cardiovscular conditions	I
Cancer	C
Diabetes	E10-E14
COPD	J44

Table B.2: ATC Codes Used to Identify Health Behaviors Related to Chronic Conditions

	ATC
Statins	C10AA
ACE	C09
Metformin	A10
COPD	R03AC18
	R43AC19
	R43AL02
	R43AL03
	R43AL04
	R43AL05
	R43AL07
	R43AL09
	R03BB04
	R03BB05
	R03BB06
	R03BB07
	R03DX07

Table B.3: Test for Selection in Patient-Physician Reassignment by patient characteristics

	low SES physician (1)
Male	-0.00132
	(0.00103)
Age	-0.00003
	(0.00007)
Non-ethnic Danish	-0.00113
	(0.00209)
Married	0.00408***
	(0.00130)
Low SES	-0.00071
	(0.00112)
Observations	474614

Notes: Standard errors are clustered by patient ID.
** p <0.01, ** p <0.05, * p <0.1.

Table B.4: Test for Selection in Patient-Physician Reassignment by patient pre-closure conditions

	Any conditions	Statins	Metformin	Diabetes check	Diabetes	COPD meds	ACSC COPD	COPD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Selection	-0.00223	0.00026	-0.00328*	0.00029	-0.00070	-0.00207	-0.00419	-0.00223
	(0.00148)	(0.00148)	(0.00193)	(0.00218)	(0.00176)	(0.00151)	(0.00337)	(0.00148)
Observations	474614	474614	474614	474614	474614	474614	474614	474614

Notes: Standard errors are clustered by patient ID.
** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B.5: The Effect of Having a Low-SES Physician (PCP) for High-SES patients by event time

VARIABLES	(1) Death	(2) Number of visits	(3) Total reimbursement GP	(4) Statins	(5) Diabetes control	(6) ACSC COPD
t=-4 × Low-SES physician	0.00003**	0.02109	0.48292	-0.00032	-0.00177	0.00045*
	(0.00001)	(0.01937)	(0.48345)	(0.00077)	(0.00231)	(0.00026)
t=-3 × Low-SES physician	0.00002	0.00177	-0.07046	0.00024	0.00357	0.00039
	(0.00001)	(0.01845)	(0.46475)	(0.00069)	(0.00225)	(0.00026)
t=-2 × Low-SES physician	0.00003**	0.01583	-0.30077	0.00036	-0.00002	0.00018
	(0.00001)	(0.01678)	(0.41777)	(0.00056)	(0.00217)	(0.00026)
t=0 × Low-SES physician	-0.00000	-0.06464***	-0.49320	-0.00005	0.01235***	0.00003
	(0.00001)	(0.01748)	(0.47071)	(0.00062)	(0.00239)	(0.00028)
t=1 × Low-SES physician	0.00011	0.00455	-0.46567	0.00003	0.01756***	0.00084***
	(0.00026)	(0.02094)	(0.58564)	(0.00083)	(0.00290)	(0.00031)
t=2 × Low-SES physician	-0.00023	0.04158*	-0.16477	-0.00015	0.00316	0.00025
	(0.00028)	(0.02241)	(0.62151)	(0.00101)	(0.00295)	(0.00032)
t=3 × Low-SES physician	-0.00019	0.04598*	0.56152	-0.00074	0.00743**	0.00071**
	(0.00030)	(0.02400)	(0.66384)	(0.00118)	(0.00306)	(0.00035)
t=4 × Low-SES physician	0.00007	0.01321	-0.36429	0.00041	-0.00175	0.00009
	(0.00032)	(0.02579)	(0.72095)	(0.00135)	(0.00306)	(0.00037)
t=5 × Low-SES physician	-0.00060*	0.00374	-0.43763	0.00120	0.00376	0.00014
	(0.00034)	(0.02742)	(0.77570)	(0.00151)	(0.00349)	(0.00039)
Observations	3,099,012	2,581,696	3,099,012	3,099,012	1,206,639	3,099,012
Patient Characteristics	Y	Y	Y	Y	Y	Y
Old PCP FE	Y	Y	Y	Y	Y	Y
Patient ID FE	N	N	N	N	N	N
Old x new PCP FE	N	N	N	N	N	N

Notes: The table presents the effect of physician-patient SES concordance on selected outcomes, see column heading. All columns report the estimates of the coefficient on the event dummies relative to $t = -1$ from equation 2.1. All regressions control for year fixed effects, new physician characteristics (mean age, share of male physicians, share of ethnic danish physicians, solo clinic dummy, number of physicians in the clinic, and physicians' graduating institution), and patient characteristics. Patient characteristics include age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education level higher than primary school. Standard errors are clustered by patient ID.
** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B.6: The Effect of Having a Low-SES Physician (PCP) for Low-SES patients by event time

VARIABLES	(1) Death	(2) Number of visits	(3) Total reimbursement GP	(4) Statins	(5) Diabetes control	(6) ACSC COPD
t=-4 × Low-SES physician	-0.00005*	0.08493**	0.55141	0.00022	-0.01029**	0.00028
	(0.00003)	(0.03534)	(0.94801)	(0.00138)	(0.00408)	(0.00063)
t=-3 × Low-SES physician	-0.00005*	0.05717*	1.26329	-0.00022	-0.00014	0.00005
	(0.00003)	(0.03314)	(0.88106)	(0.00122)	(0.00393)	(0.00062)
t=-2 × Low-SES physician	-0.00005*	0.06770**	0.04047	-0.00035	-0.00869**	0.00009
	(0.00003)	(0.02968)	(0.76605)	(0.00097)	(0.00368)	(0.00063)
t=0 × Low-SES physician	-0.00008***	-0.08064***	0.01903	0.00207*	0.01176***	-0.00028
	(0.00003)	(0.03083)	(0.85734)	(0.00111)	(0.00397)	(0.00068)
t=1 × Low-SES physician	-0.00099*	0.20104***	2.98480***	0.00328**	0.02531***	-0.00013
	(0.00056)	(0.03818)	(1.08580)	(0.00151)	(0.00473)	(0.00074)
t=2 × Low-SES physician	-0.00106*	0.18076***	2.31234*	0.00130	0.02160***	-0.00101
	(0.00060)	(0.04128)	(1.19960)	(0.00184)	(0.00488)	(0.00079)
t=3 × Low-SES physician	-0.00236***	0.17850***	4.03913***	0.00320	0.01962***	-0.00158**
	(0.00063)	(0.04438)	(1.28102)	(0.00213)	(0.00507)	(0.00081)
t=4 × Low-SES physician	0.00067	0.10691**	2.24364	0.00248	0.01704***	-0.00086
	(0.00071)	(0.04742)	(1.37804)	(0.00241)	(0.00508)	(0.00088)
t=5 × Low-SES physician	-0.00017	0.06918	3.38732**	0.00388	0.02736***	0.00034
	(0.00073)	(0.05027)	(1.47647)	(0.00268)	(0.00574)	(0.00094)
Observations	1,243,377	1,066,244	1,243,377	1,243,377	487,346	1,243,377
Patient Characteristics	Y	Y	Y	Y	Y	Y
Old PCP FE	Y	Y	Y	Y	Y	Y
Patient ID FE	N	N	N	N	N	N
Old x new PCP FE	N	N	N	N	N	N

Notes: The table presents the effect of physician-patient SES concordance on selected outcomes, see column heading. All columns report the estimates of the coefficient on the event dummies relative to $t = -1$ from equation 2.1. All regressions control for year fixed effects, new physician characteristics (mean age, share of male physicians, share of ethnic danish physicians, solo clinic dummy, number of physicians in the clinic, and physicians' graduating institution), and patient characteristics. Patient characteristics include age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education level higher than primary school. Standard errors are clustered by patient ID.
** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B.7: The Effect of Physician-patient SES Concordance on Mortality by Population Demographics

VARIABLES	(1) Male	(2) Female	(3) yob<1958	(4) yob>=1958	(5) Ethnic Danish	(6) Not ethnic Danish
PCP low SES x Patient low SES x Post	-0.00184***	-0.00083*	-0.00079**	-0.00171***	-0.00133***	-0.00003
	(0.00062)	(0.00048)	(0.00039)	(0.00061)	(0.00040)	(0.00200)
PCP low SES x Patient low SES	0.00005	0.00019**	0.00002	0.00011	0.00014**	-0.00014
	(0.00010)	(0.00008)	(0.00007)	(0.00010)	(0.00006)	(0.00032)
Patient low SES x Post	0.00622***	0.00460***	0.00281***	0.00577***	0.00497***	0.00334***
	(0.00036)	(0.00028)	(0.00023)	(0.00036)	(0.00023)	(0.00098)
PCP low SES x Post	0.00001	-0.00019	0.00002	-0.00028	-0.00015	0.00018
	(0.00025)	(0.00020)	(0.00014)	(0.00031)	(0.00018)	(0.00042)
Patient low SES	-0.00044	-0.00126***	0.00012	0.00170	-0.00228***	0.00029
	(0.00093)	(0.00022)	(0.00033)	(0.00147)	(0.00036)	(0.00069)
Post	0.00602***	0.00354***	0.00155***	0.00821***	0.00510***	0.00329***
	(0.00016)	(0.00014)	(0.00009)	(0.00020)	(0.00012)	(0.00025))
Outcome mean	.00281	.00185	.00086	.00377	.00244	.00131
Gradient for high-SES physicians	.00641	.00495	.0028	.00586	.00527	.00355
Effect %	-28.7	-16.8	-28.2	-29.2	-25.2	-0.8
Observations	1,914,426	1,835,228	1,847,399	1,902,255	3,405,243	344,407
Patient Characteristics	Y	Y	Y	Y	Y	Y
Old x new PCP FE	Y	Y	Y	Y	Y	Y

Notes: The table presents the effect of physician-patient SES concordance on mortality for different sub-populations, see column heading.All columns report the estimates from the triple differences equation 3.3. Patient characteristics include age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education level higher than primary school. Gradient for high-SES physician is the difference in the outcome variable between high and low-SES patients who have high-SES physicians in the post period, calculated as $(lowSESoutcome - highSESoutcome)$. The effect in percentage is calculated as $(Tripledifferenceestimate/gradientforhigh - SESphysician) \times 100$. Standard errors are clustered by patient ID. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table B.8: The Effect of Physician(PCP)-patient SES Concordance on Mortality Caused by Chronic Conditions and Gender

VARIABLES	(1) Death	(2) CVC	(3) Cancer	(4) Diabetes	(5) COPD
Panel A: Female					
PCP low SES x Patient low SES x Post	-0.00083*	-0.00019	-0.00059*	0.00006	-0.00003
	(0.00048)	(0.00018)	(0.00033)	(0.00008)	(0.00013)
Outcome mean	.00185	.00025	.00092	.00004	.00012
Gradient for high-SES physicians	.00495	.00088	.00206	.00013	.00056
Effect %	-16.8	-21.6	-28.6	46.2	-5.4
Observations	1,835,228	1,835,228	1,835,228	1,835,228	1,835,228
Panel B: Male					
PCP low SES x Patient low SES x Post	-0.00184***	-0.00067**	-0.00027	0.00006	-0.00005
	(0.00062)	(0.00028)	(0.00037)	(0.00012)	(0.00013)
Outcome mean	.00281	.00058	.00104	.0001	.00011
Gradient for high-SES physicians	.00641	.0013	.00164	.00025	.00039
Effect %	-28.7	-51.5	-16.5	24.0	-12.8
Observations	1,914,426	1,914,426	1,914,426	1,914,426	1,914,426
Patient Characteristics	Y	Y	Y	Y	Y
Old x new PCP FE	Y	Y	Y	Y	Y

Notes: The table presents the effect of physician-patient SES concordance on mortality by causes of death. All columns report the estimates from the triple differences equation 3.3. Patient characteristics include age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education. Gradient for high-SES physician is the difference in the outcome variable between high and low-SES patients who have high-SES physicians in the post period, calculated as $(lowSESoutcome - highSESoutcome)$. The effect in percentage is calculated as $(Tripledifferenceestimate/gradientforhigh - SESphysician) \times 100$. Standard errors are clustered by patient ID.
** p <0.01, ** p <0.05, * p <0.1.

Table B.9: The Effect of Physician(PCP)-patient SES Concordance on Mortality Caused by Chronic Conditions by Birth Cohort

VARIABLES	(1) Death	(2) CVC	(3) Cancer	(4) Diabetes	(5) COPD
Panel A: Younger sample, year of birth>1957					
PCP low SES x Patient low SES x Post	-0.00079**	-0.00011	-0.00014	0.00003	-0.00010*
	(0.00039)	(0.00014)	(0.00022)	(0.00006)	(0.00006)
Outcome mean	.00086	.00012	.0003	.00002	.00002
Gradient for high-SES physicians	.0028	.00035	.0007	.00006	.00014
Effect %	-28.2	-31.4	-20.0	50.0	-71.4
Observations	1,847,399	1,847,399	1,847,399	1,847,399	1,847,399
Panel B: Older sample, year of birth<=1957					
New PCP low SES x Patient low SES x Post	-0.00171***	-0.00061**	-0.00084**	0.00009	0.00002
	(0.00061)	(0.00024)	(0.00037)	(0.00011)	(0.00014)
Outcome mean	.00377	.0007	.00164	.00012	.00021
Gradient for high-SES physicians	.00586	.00121	.00197	.0002	.00065
Effect %	-29.2	-50.4	-42.6	45.0	3.1
Observations	1,902,255	1,993,260	1,993,260	1,993,260	1,993,260
Patient Characteristics	Y	Y	Y	Y	Y
Old x new PCP FE	Y	Y	Y	Y	Y

Notes: The table presents the effect of physician-patient SES concordance on mortality by causes of death. All columns report the estimates from the triple differences equation 3.3. Patient characteristics include age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education. Gradient for high-SES physician is the difference in the outcome variable between high and low-SES patients who have high-SES physicians in the post period, calculated as $(lowSESoutcome - highSESoutcome)$. The effect in percentage is calculated as $(Tripledifferenceestimate/gradientforhigh - SESphysician) \times 100$. Standard errors are clustered by patient ID.
** p <0.01, ** p <0.05, * p <0.1.

Table B.10: The Effect of Physician-patient SES Concordance on Cancer Mortality

CONDITIONS	All (1)	lung (2)	Mouth (3)	Digestive system (4)	lung etc (5)	Bones (5)	Skin (6)	Breast (7)	genital organs (8)	Kidney (9)
Panel A: Female										
PCP low SES x Patient low SES x Post	-0.00059*	-0.00019	-0.00006*	-0.00011	-0.00017	-0.00001	0.00002	-0.00002	-0.00000	-0.00004
	(0.00033)	(0.00019)	(0.00004)	(0.00015)	(0.00019)	(0.00001)	(0.00004)	(0.00013)	(0.00010)	(0.00006)
Outcome mean	.00092	.00026	.00001	.00019	.00027	0	.00002	.00016	.0001	.00003
Gradient for high-SES physicians	.00206	.00088	.00006	.00043	.00088	.00001	-.00001	.00016	.00006	.00011
Observations	1,835,228	1,835,228	1,835,228	1,835,228	1,835,228	1,835,228	1,835,228	1,835,228	1,835,228	1,835,228
Panel B: Older sample, year of birth<=1957										
PCP low SES x Patient low SES x Post	-0.00070*	-0.00022	-0.00012	-0.00016	-0.00022	-0.00001	0.00004	0.00000	-0.00019	0.00013
	(0.00040)	(0.00023)	(0.00000)	(0.00021)	(0.00023)	(0.00000)	(0.00000)	(0.00011)	(0.00012)	(0.00000)
Outcome mean	.00164	.00047	.00005	.00045	.0005	0	.00003	.00012	.00015	.00008
Gradient for high-SES physicians	.00197	.00099	.00007	.00029	.00099	.00001	-.00005	.00015	.00013	.00009
Observations	1,902,255	1,902,255	1,902,255	1,902,255	1,902,255	1,902,255	1,902,255	1,902,255	1,902,255	1,902,255
Patient Characteristics	Y	Y	Y	Y	Y	Y	Y	Y		
Old x new PCP FE	Y	Y	Y	Y	Y	Y	Y	Y		

Notes: The table presents the effect of physician-patient SES concordance on health behaviors related to the four most common and unequal chronic conditions. All columns report the estimates from the triple differences equation 3.3. Patient characteristics include age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education level higher than primary school. Gradient for high-SES physician is the difference in the outcome variable between high and low-SES patients who have high-SES physicians in the post period, calculated as $(lowSESoutcome - highSESoutcome)$. The effect in percentage is calculated as $(Tripledifferenceestimate/gradientforhigh - SESphysician) \times 100$. Standard errors are clustered by patient ID. ** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B.11: The Effect of Physician(PCP)-patient SES Concordance on Primary Care Reimbursement

VARIABLES	(1) PCP	(2) Specialist	(3) Total
PCP low SES x Patient low SES x Post	2.75673***	4.08300**	6.83973***
	(0.97688)	(2.01551)	(2.35264)
Outcome mean	122.3	210.3	332.6
Gradient for high-SES physicians	32.3	-12.2	20.1
Effect %	8.5	-33.5	34.0
Observations	3,749,654	3,749,654	3,749,654
Patient Characteristics	Y	Y	Y
Old x new PCP FE	Y	Y	Y

Notes: The table presents the effect of physician-patient SES concordance on physician fee-for-service reimbursements in US dollars. All columns report the estimates from the triple differences equation 3.3. Patient characteristics include age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education level higher than primary school. Gradient for high-SES physician is the difference in the outcome variable between high and low-SES patients who have high-SES physicians in the post period, calculated as $(lowSESoutcome - highSESoutcome)$. The effect in percentage is calculated as $(Triple difference estimate/gradient for high - SES physician) \times 100$. Standard errors are clustered by patient ID. ** p <0.01, ** p <0.05, * p <0.1.

Table B.12: The Effect of Physician-patient SES Concordance on Healthcare Utilization by Gender

VARIABLES	(1) PCP visit (Dummy)	(2) PCP visit (N)	(3) Services per visit (N)	(4) Specialist visit (Dummy)
Panel A: Female				
PCP low SES x Patient low SES x Post	0.00055	0.13823***	0.02129***	0.00020
	(0.00204)	(0.04484)	(0.00751)	(0.00336)
Outcome mean	.89983	6.92182	1.57678	.39520
Gradient for high-SES physicians	.0122	1.57709	.04725	-.02907
Effect %	4.5	9.2	45.1	-0.7
Observations	1,835,228	1,650,269	1,835,228	1,835,228
Panel B: Male				
PCP low SES x Patient low SES x Post	-0.00300	0.11030**	0.00513	0.00694**
	(0.00289)	(0.04878)	(0.00830)	(0.00323)
Outcome mean	.78002	5.48716	1.31885	.26916
Gradient for high-SES physicians	.02112	1.11479	.03746	-.02022
Effect %	-14.2	9.9	13.7	-34.3
Observations	1,914,426	1,490,387	1,914,426	1,914,426
Patient Characteristics	Y	Y	Y	Y
Old x new PCP FE	Y	Y	Y	Y

Notes: The table presents the effect of physician-patient SES concordance on healthcare utilization. All columns report the estimates of coefficients on the event dummies in equation 3.3. Patient characteristics include age fixed effects, gender, dummy for being ethnic Danish, dummy for being married, and education. Gradient for high SES patients is the difference of the outcome variable between high and low SES patients who have high SES physicians, calculated as $(lowSESoutcome - highSESoutcome)$. Standard errors are clustered by patient ID.
** p <0.01, ** p <0.05, * p <0.1.

Table B.13: The Effect of Physician-patient SES Concordance on Healthcare Utilization by Birth Cohort

VARIABLES	(1) PCP visit (Dummy)	(2) PCP visit (N)	(3) Services per visit (N)	(4) Specialist visit (Dummy)
Panel A: Young sample				
PCP low SES x Patient low SES x Post	-0.00021	0.05617	0.00065	0.00595*
	(0.00278)	(0.04820)	(0.00820)	(0.00346)
Outcome mean	.82432	5.53897	1.39339	.29216
Gradient for high-SES physicians	.02068	1.28907	.03479	-.0193
Effect %	-1.0	4.4	1.9	-30.8
Observations	1,847,399	1,520,555	1,847,399	1,847,399
Panel B: Old sample				
PCP low SES x Patient low SES x Post	-0.00204	0.13201***	0.02665***	-0.00016
	(0.00225)	(0.04536)	(0.00765)	(0.00318)
Outcome mean	.85258	6.89963	1.4953	.36842
Gradient for high-SES physicians	.01943	1.28458	.04365	-.03153
Effect %	-10.5	10.3	61.1	0,5
Observations	1,902,255	1,620,140	1,902,255	1,902,255
Patient Characteristics	Y	Y	Y	Y
Old x new PCP FE	Y	Y	Y	Y

Notes: The table presents the effect of physician-patient SES concordance on healthcare utilization. All columns report the estimates of coefficients on the event dummies in equation 3.3. Patient characteristics include age fixed effects, gender, dummy for being ethnic Danish, dummy for being married, and education. Gradient for high SES patients is the difference of the outcome variable between high and low SES patients who have high SES physicians, calculated as $(lowSESoutcome - highSESoutcome)$. Standard errors are clustered by patient ID.
** p <0.01, ** p <0.05, * p <0.1.

Table B.14: The Effect of Physician-patient SES Concordance on Primary Care Reimbursement by Gender

VARIABLES	(1) PCP	(2) Specialist	(3) Total
Panel A: Female			
PCP low SES x Patient low SES x Post	4.55223***	3.67556	8.22779**
	(1.36678)	(2.94248)	(3.39929)
Outcome mean	142.0	251.7	393.7
Gradient for high-SES physicians	34.19915	-24.79902	9.40013
Effect %	13.3	-14.8	87.5
Observations	1,835,228	1,835,228	1,835,228
Panel B: Male			
PCP low SES x Patient low SES x Post	0.82749	3.99810	4.82559
	(1.40418)	(2.74350)	(3.24872)
Outcome mean	103.5	170.6	274.1
Gradient for high-SES physicians	24.33833	-12.7548	11.58353
Effect %	3.4	-31.3	41.7
Observations	1,914,426	1,914,426	1,914,426
Patient Characteristics	Y	Y	Y
Old x new PCP FE	Y	Y	Y

Notes: The table presents the effect of physician-patient SES concordance on physician fee-for-service reimbursement in US dollars. All columns report the estimates from the triple differences equation 3.3. Patient characteristics include age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education level higher than primary school. Gradient for high-SES physician is the difference in the outcome variable between high and low-SES patients who have high-SES physicians in the post period, calculated as $(lowSESoutcome - highSESoutcome)$. The effect in percentage is calculated as $(Tripledifferenceestimate/gradientforhigh - SESphysician) \times 100$. Standard errors are clustered by patient ID. ** p <0.01, ** p <0.05, * p <0.1.

Table B.15: The Effect of Physician-patient SES Concordance on Primary Care Reimbursement by Birth Cohort

VARIABLES	(1) PCP	(2) Specialist	(3) Total
Panel A: Younger sample, year of birth>1957			
PCP low SES x Patient low SES x Post	4.55223***	3.67556	8.22779**
	(1.36678)	(2.94248)	(3.39929)
Outcome mean	109.87715	193.30564	303.1828
Gradient for high-SES physicians	28.59009	-1.56773	27.02235
Effect %	15.9	-234.5	30.4
Observations	1,847,399	1,847,399	1,847,399
Panel B: Older sample, year of birth<=1957			
PCP low SES x Patient low SES x Post	3.46626**	3.47760	6.94387**
	(1.38461)	(2.76099)	(3.24055)
Outcome mean	134.43148	226.83002	361.26151
Gradient for high-SES physicians	26.82045	-31.01981	-4.19936
Effect %	12.9	-11.2	-165.4
Observations	1,902,255	1,902,255	1,902,255
Patient Characteristics	Y	Y	Y
Old x new PCP FE	Y	Y	Y

Notes: The table presents the effect of physician-patient SES concordance on physician fee-for-service reimbursement in US dollars. All columns report the estimates from the triple differences equation 3.3. Patient characteristics include age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education. Gradient for high-SES physician is the difference in the outcome variable between high and low-SES patients who have high-SES physicians in the post period, calculated as $(lowSESoutcome - highSESoutcome)$. The effect in percentage is calculated as $(Tripledifferenceestimate/gradientforhigh - SESphysician) \times 100$. Standard errors are clustered by patient ID.*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B.16: The Effect of Physician-patient SES Concordance on Health Behaviors Related to Chronic Conditions by Gender

CONDITIONS	CVC		Cancer		Diabetes		COPD	
	Statins	ACE	Lung scan	Radiologist	Metformin	Office visit	Medication	Hospitalization
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Female								
PCP low SES x Patient low SES x Post	0.00228	0.00138	0.00221	0.00035	0.00042	0.01460***	-0.00066	-0.00020
	(0.00231)	(0.00232)	(0.00150)	(0.00084)	(0.00122)	(0.00354)	(0.00069)	(0.00165)
Outcome mean	.09325	.1142	.03284	.0225	.03415	.09541	.00598	.06515
Gradient for high-SES physicians	.06817	.05991	.00945	-.00475	.02406	.03542	.00962	.0418
Effect %	3.3	2.3	23.4	-7.4	1.7	41.2	-6.9	-0.5
Observations	1,835,228	1,835,228	1,591,982	1,835,228	1,835,228	1,030,128	1,835,228	1,835,228
Panel B: Male								
PCP low SES x Patient low SES x Post	0.00362	0.00075	-0.00050	0.00001	0.00030	0.00740**	-0.00192***	-0.00162
	(0.00248)	(0.00255)	(0.00156)	(0.00081)	(0.00154)	(0.00356)	(0.00070)	(0.00152)
Outcome mean	.11459	.13641	.03275	.01735	.0517	.09504	.0053	.0466
Gradient for high-SES physicians	.0274	.02858	.00841	-.0018	.02596	.01168	.0076	.02014
Effect %	13.2	2.6	5.9	-0.6	1.2	63.4	-8.0	-25.3
Observations	1,914,426	1,914,426	1,669,987	1,914,426	1,914,426	1,093,202	1,914,426	1,914,426
Patient Characteristics	Y	Y	Y	Y	Y	Y	Y	Y
Old x new PCP FE	Y	Y	Y	Y	Y	Y	Y	Y

Notes: The table presents the effect of physician-patient SES concordance on health behaviors related to the four most common and unequal chronic conditions. All columns report the estimates from the triple differences equation 3.3. Patient characteristics include age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education level higher than primary school. Gradient for high-SES physician is the difference in the outcome variable between high and low-SES patients who have high-SES physicians in the post period, calculated as $(lowSESoutcome - highSESoutcome)$. The effect in percentage is calculated as $(Tripledifferenceestimate/gradientforhigh - SESphysician) \times 100$. Standard errors are clustered by patient ID. ** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table B.17: The Effect of Physician-patient SES Concordance on Health Behaviors Related to Chronic Conditions by Birth Cohort

CONDITIONS	CVC		Cancer		Diabetes		COPD	
	Statins	ACE	Lung scan	Radiologist	Metformin	Office visit	Medication	Hospitalization
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Younger sample, year of birth>1957								
PCP low SES x Patient low SES x Post	0.00127	-0.00025	-0.00083	0.00120	-0.00004	0.00565*	-0.00041	-0.00009
	(0.00194)	(0.00209)	(0.00149)	(0.00092)	(0.00121)	(0.00293)	(0.00053)	(0.00159)
Outcome mean	.04747	.06231	.02718	.02101	.02565	.05446	.00223	.04709
Gradient for high-SES physicians	.02852	.02899	.0088	-.00286	.01606	.01441	.00427	.02341
Effect %	4.5	-0.9	-9.4	-42.0	-0.2	39.2	-0.4	-9.6
Observations	1,847,399	1,847,399	1,711,402	1,847,399	1,847,399	1,168,080	1,847,399	1,847,399
Panel B: Older sample, year of birth<=1957								
PCP low SES x Patient low SES x Post	0.00359	0.00227	0.00262*	-0.00051	0.00053	0.01105***	-0.00196**	-0.00111
	(0.00259)	(0.00260)	(0.00156)	(0.00076)	(0.00145)	(0.00395)	(0.00077)	(0.00158)
Outcome mean	.15919	.18695	.03899	.01876	.06006	.14504	.00893	.06402
Gradient for high-SES physicians	.03038	.02086	.00603	-.00316	.02001	.01388	.01038	.03632
Effect %	11.8	10.9	43.4	-16.1	2.6	79.6	-3.1	-18.9
Observations	1,902,255	1,902,255	1,550,770	1,902,255	1,902,255	955,507	1,902,255	1,902,255
Patient Characteristics	Y	Y	Y	Y	Y	Y	Y	Y
Old x new PCP FE	Y	Y	Y	Y	Y	Y	Y	Y

Notes: The table presents the effect of physician-patient SES concordance on health behaviors related to the four most common and unequal chronic conditions. All columns report the estimates from the triple differences equation 3.3. Patient characteristics include age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education. Gradient for high-SES physician is the difference in the outcome variable between high and low-SES patients who have high-SES physicians in the post period, calculated as $(lowSESoutcome - highSESoutcome)$. The effect in percentage is calculated as $(Triple difference estimate/gradient for high - SES physician) \times 100$. Standard errors are clustered by patient ID.
** p <0.01, ** p <0.05, * p <0.1.

Table B.18: Robustness Check: Adherence (A) vs. detection (D) effects

	(1)	(2)	(3)	(4)	(5)	(6)
	Death		Death from CVC		Statins	
	A	D	A	D	A	D
PCP low SES x Patient low SES x Post	-0.00218***	-0.00081**	-0.00071	-0.00038**	0.01357*	0.00018
	(0.00083)	(0.00041)	(0.00071)	(0.00015)	(0.00713)	(0.00149)
Outcome mean	.00373	.00169	.00117	.00031	.70931	.01662
Gradient for high-SES physicians	.00728	.00389	.00138	.00084	.01681	.01888
Effect %	-29.9	-20.8	-51.4	-45.2	80.7	1.0
Observations	1,189,938	2,559,716	473,814	3,275,840	473,814	3,275,840
Patient Characteristics	Y	Y	Y	Y	Y	Y
Old x new GP FE	Y	Y	Y	Y	Y	Y

Notes: The table presents the effect of physician-patient SES concordance on main outcomes, see column heading. All columns report the estimates from the triple differences equation 3.3 by patient types. Column 1 shows the effect for patients with pre-existing diagnosis before clinic closures, column 2 shows the effect for new patients of the respective disease. E.g., columns 3 and 5 show the effect for patients prescribed statins in the pre-period, columns 4 and 6 for patients not prescribed statins in the pre-period. Column 7 (9) shows the effect for patients being treated for diabetes or (COPD) in the pre-period, and column 8 (10) shows the effect for patients not treated in the pre-period. Patient characteristics include age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education level higher than primary school. Gradient for high-SES physician is the difference in the outcome variable between high and low-SES patients who have high-SES physicians in the post period, calculated as $(lowSESoutcome - highSESoutcome)$. The effect in percentage is calculated as $(Triple difference estimate/gradient for high - SES physician) \times 100$. Standard errors are clustered by patient ID.
** p <0.01, ** p <0.05, * p <0.1.

Table B.19: Internal Validity: The Role of Other Physician Characteristics in Reducing the SES-gradient in Mortality

VARIABLES	(1) Most experience	(2) Male	(3) Ethnic Danish	(4) UCPH
PCP X x Patient low-SES x Post	0.00021	0.00009	0.00005	0.00047
	(0.00039)	(0.00037)	(0.00041)	(0.00038)
Outcome mean	.00234	.00234	.00234	.00234
Gradient for other physicians	.00491	.00495	.00496	.00476
Effect %	4.3	1.8	1.0	9.9
Observations	3,749,654	3,749,654	3,749,654	3,749,654
Patient Characteristics	Y	Y	Y	Y
Old x new PCP FE	Y	Y	Y	Y

Notes: The table tests for the role of other physician characteristics on the health-SES gradient. All columns report the estimates from the triple differences equation 3.3, replacing physician SES by the respective physician characteristic. Patient characteristics include age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education level higher than primary school. Gradient for high-SES physician is the difference in the outcome variable between high and low-SES patients who have high-SES physicians in the post period, calculated as $(lowSESoutcome - highSESoutcome)$. The effect in percentage is calculated as $(Tripledifferenceestimate/gradientforhigh - SESphysician) \times 100$. Standard errors are clustered by patient ID. ** p <0.01, ** p <0.05, * p <0.1.

Table B.20: Internal Validity: The Role of Other Physician Characteristics in Reducing the SES-gradient in Mortality

VARIABLES	Most experienced (1)	Least experienced (2)	Most male (3)	Least male (4)	Non-ethnic Danish (5)	Ethnic Danish (6)	UCPH (7)	Non UCPH (8)
PCP low SES x Patient low SES x Post	-0.00154***	-0.00110**	-0.00169***	-0.00094*	-0.00093	-0.00147***	-0.00198***	-0.00090*
	(0.00055)	(0.00054)	(0.00056)	(0.00053)	(0.00083)	(0.00044)	(0.00065)	(0.00048)
Outcome mean	.00234	.00234	.00234	.00234	.00234	.00234	.00234	.00234
Gradient for high-SES physicians	.00474	.00522	.00495	.00497	.00504	.00489	.00473	.00546
Observations	1,876,444	1,873,210	1,699,672	2,049,982	868,550	2,881,104	1,499,790	2,249,864
Patient Characteristics	Y	Y	Y	Y	Y	Y	Y	Y
Old x new PCP FE	Y	Y	Y	Y	Y	Y	Y	Y

Notes: The table presents the effect of physician-patient SES concordance on selected outcomes, see column heading. All columns report the estimates from the triple differences equation 3.3. All columns report the estimates from the triple differences equation 3.3, but replacing physician being low-SES with another characteristics. Patient characteristics include age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education level higher than primary school. Gradient for high-SES physician is the difference in the outcome variable between high and low-SES patients who have high-SES physicians in the post period, calculated as $(lowSESoutcome - highSESoutcome)$. The effect in percentage is calculated as $(Triple difference estimate/gradient for high - SES physician) \times 100$.

Table B.21: Robustness Check: The Effect of Physician Academic Performance

VARIABLES	(1) Death	(2) Death from CVC	(3) Number of visits	(4) Total reimbursement	(5) Statins	(6) Hospitalization COPD	(7) Diabetes control
PCP high GPA x Patient low-SES x Post	-0.00008	-0.00009	-0.00707	0.52153	-0.00086	-0.00008	-0.00054
	(0.00044)	(0.00018)	(0.03634)	(2.65880)	(0.00191)	(0.00056)	(0.00262)
Outcome mean	.00234	.00042	6.2	332.6	.10415	.00563	.09522
Gradient for high-SES physicians	.00541	.00101	1.4598	20.10358	.04643	.00866	.0243
Effect %	-1.5	-8.9	-0.5	2.6	-1.9	-0.9	-2.2
Observations	3,749,654	3,749,654	3,140,867	3,749,654	3,749,654	3,749,654	2,123,957
Patient Characteristics	Y	Y	Y	Y	Y	Y	Y
Old x new PCP FE	Y	Y	Y	Y	Y	Y	Y

Notes: The table presents the effect of physician-patient SES concordance on selected outcomes, see column heading. All columns report the estimates from the triple differences equation **??**, defining a physician as having high academic performance (measured by high-school GPA) if they are among the 30 percent highest. Patient characteristics includes age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education. Gradient for high-SES physician is the difference in the outcome variable between high and low-SES patients who have high-SES physicians in the post period, calculated as $(lowSESoutcome - highSESoutcome)$. The effect in percentage is calculated as $(Triple difference estimate/gradient for high - SES physician) \times 100$. Standard errors are clustered by patient ID.*** p <0.01, ** p <0.05, * p <0.1.

Table B.22: Mechanisms: The Effect of Physicians' Parents' Illness on The SES Gradient in Mortality

VARIABLES	All cause mortality All conditions	All cause mortality CVC	All cause mortality of Cancer	CVC mortality CVC	Cancer mortality Cancer
	(1)	(2)	(3)	(4)	(5)
PCP high GPA x Patient low-SES x Post	-0.00114***	-0.00137***	-0.00046	-0.00024	-0.00051**
	(0.00043)	(0.00041)	(0.00040)	(0.00017)	(0.00025)
Outcome mean	.00234	.00234	.00234	.00042	.00098
Gradient for high-SES physicians	.00541	.00541	.00541	.00101	.00182
Observations	3,654,767	3,749,654	3,749,654	3,749,654	3,749,654
Patient Characteristics	Y	Y	Y	Y	Y
Old x new PCP FE	Y	Y	Y	Y	Y

Notes: The table presents the effect of physician-patient SES concordance on selected outcomes, see column heading. All columns report the estimates from the triple differences equation **??**, defining a physician as having high academic performance (measured by high-school GPA) if they are among the 30 percent highest. Patient characteristics includes age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education. Gradient for high-SES physician is the difference in the outcome variable between high and low-SES patients who have high-SES physicians in the post period, calculated as $(lowSESoutcome - highSESoutcome)$. The effect in percentage is calculated as $(Tripledifferenceestimate/gradientforhigh - SESphysician) \times 100$. Standard errors are clustered by patient ID.*** p <0.01, ** p <0.05, * p <0.1.

Table B.23: Robustness Check: Define Physician SES Using Parents' Educational Rank

VARIABLES	(1) Death	(2) Death from CVC	(3) Number of visits	(4) Total reimbursement	(5) Statins	(6) Hospitalization COPD	(7) Diabetes control
PCP low-SES x Patient low-SES x Post	-0.00080*	-0.00038**	0.05378	5.59302**	0.00257	-0.00090*	0.00532**
	(0.00042)	(0.00017)	(0.03594)	(2.55400)	(0.00187)	(0.00054)	(0.00269)
Outcome mean	.00234	.00042	6.24079	332.64699	.10415	.00563	.09522
Gradient for high-SES physicians	.00517	.00101	1.4598	20.10358	.04643	.00866	.0243
Effect %	-15.5	-37.6	3.7	27.8	5.5	-10.4	21.9
Observations	3,749,654	3,749,654	3,140,867	3,749,654	3,749,654	3,749,654	2,123,957
Patient Characteristics	Y	Y	Y	Y	Y	Y	Y
Old x new PCP FE	Y	Y	Y	Y	Y	Y	Y

Notes: The table presents the effect of physician-patient SES concordance on main outcomes. All columns report the estimates of coefficients from the triple-difference equation 3.3. Patient characteristics include age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education level higher than primary school. Gradient for high-SES physician is the difference in the outcome variable between high and low-SES patients who have high-SES physicians in the post period, calculated as $(lowSESoutcome - highSESoutcome)$. The effect in percentage is calculated as $(Triple difference estimate/gradient for high - SES physician) \times 100$. Standard errors are clustered by patient ID.*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B.24: Robustness Check: Define SES Using Other Education Levels

VARIABLES	(1) Primary school (baseline result)	(2) Vocational education	(3) College education
PCP education level x Patient education level x Post	-0.00134***	-0.00036	-0.00000
	(0.00039)	(0.00030)	(0.00032)
Outcome mean	.00234	.00234	.00234
Gradient for other physicians type	.00541	.00505	.00498
Effect %	-1.5	-8.9	-0.5
2.6	-1.9	-0.9	-2.2
Observations	3,749,654	3,749,654	3,749,654
Patient Characteristics	Y	Y	Y
Old x new PCP FE	Y	Y	Y

Notes: The table presents the effect of physician's parents and patient education concordance on mortality. All columns report the estimates from the triple differences equation 3.3. Patient characteristics include age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education level higher than primary school. Gradient for high-SES physician is the difference in the outcome variable between high and low-SES patients who have high-SES physicians in the post period, calculated as $(lowSESoutcome - highSESoutcome)$. The effect in percentage is calculated as $(Triple difference estimate/gradient for high - SES physician) \times 100$. Standard errors are clustered by patient ID.
** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table B.25: Robustness Check: Restricted Choice Sample

VARIABLES	(1) Death	(2) Death from CVC	(3) Number of visits	(4) Total reimbursement	(5) Statins	(6) Hospitalization COPD	(7) Diabetes control
PCP low SES x Patient low-SES x Post	-0.00144*	-0.00041	0.16621**	16.46544***	0.00070	-0.00078	0.00126
	(0.00085)	(0.00035)	(0.07191)	(5.73123)	(0.00367)	(0.00118)	(0.00347)
Outcome mean	.00195	.00032	6.2	391.9	.12511	.00606	.05107
Gradient for high-SES physicians	.00607	.00112	1.46923	27.13185	.0646	.01044	.01143
Effect %	-23.7	-36.6	11.3	60.7	1.1	-7.5	11.0
Observations	1,028,570	1,028,570	871,287	1,028,570	1,028,570	1,028,570	766,414
Patient Characteristics	Y	Y	Y	Y	Y	Y	Y
Old x new PCP FE	Y	Y	Y	Y	Y	Y	Y

Notes: The table presents the effect of physician-patient SES concordance on selected outcomes, see column heading. All columns report the estimates from the triple differences equation 3.3. Patient characteristics include age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education level higher than primary school. Gradient for high-SES physician is the difference in the outcome variable between high and low-SES patients who have high-SES physicians in the post period, calculated as $(lowSESoutcome - highSESoutcome)$. The effect in percentage is calculated as $(Tripledifferenceestimate/gradientforhigh - SESphysician) \times 100$. The restricted choice sample is defined as experiencing a clinic closure in a municipality and year where the average physician has more than 1600 patients. Standard errors are clustered by patient ID.*** p <0.01, ** p <0.05, * p <0.1.

Table B.26: Robustness check: Aggregating Physician SES to the Clinic Level

VARIABLES	(1) Death	(2) Death from CVC	(3) Number of visits	(4) Total reimbursement	(5) Statins	(6) Hospitalization COPD	(7) Diabetes control
Panel A: Min							
PCP low SES x Patient low-SES x Post	-0.00115**	-0.00031	0.17232***	3.79079	0.00855***	-0.00186***	0.00979***
	(0.00052)	(0.00023)	(0.04476)	(3.20352)	(0.00232)	(0.00066)	(0.00377)
Effect %	-21.3	-30.7	11.8	18.9	18.4	-21.5	40.3
Observations	3,749,654	3,749,654	3,140,867	3,749,654	3,749,654	3,749,654	2,123,957
Panel B: Mean							
PCP low SES x Patient low-SES x Post	-0.00138***	-0.00041*	0.16751***	7.55329**	0.00729***	-0.00191***	0.01441***
	(0.00050)	(0.00022)	(0.04314)	(3.08251)	(0.00223)	(0.00064)	(0.00352)
Effect %	-25.5	-40.6	11.5	37.6	15.7	-22.1	59.3
Observations	3,749,654	3,749,654	3,140,867	3,749,654	3,749,654	3,749,654	2,123,957
Patient Characteristics	Y	Y	Y	Y	Y	Y	Y
Old x new PCP FE	Y	Y	Y	Y	Y	Y	Y

Notes: The table presents the effect of physician-patient SES concordance on the main outcomes, see column heading. All columns report the estimates from the triple differences equation 3.3. Patient characteristics include age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education. Panel A ("min") defines a clinic to be low-SES of all physicians are low-SES. Panel B ("mean") uses the proportion of physicians that are low-SES. Gradient for high-SES physician is the difference in the outcome variable between high and low-SES patients who have high-SES physicians in the post period, calculated as $(lowSESoutcome - highSESoutcome)$. The effect in percentage is calculated as $(Tripledifferenceestimate/gradientforhigh - SESphysician) \times 100$. Standard errors are clustered by patient ID.
** p <0.01, ** p <0.05, * p <0.1.

Table B.27: Robustness check: Subsample of Physicians with Non-missing SES

VARIABLES	(1) Death	(2) Death from CVC	(3) Number of visits	(4) Total reimbursement	(5) Statins	(6) Hospitalization COPD	(7) Diabetes control
Panel A: Max							
PCP low SES x Patient low SES x Post	-0.00098*	-0.00044**	0.11248***	5.84811*	0.00389*	-0.00171***	0.01615***
	(0.00052)	(0.00022)	(0.04354)	(3.17264)	(0.00228)	(0.00066)	(0.00347)
Effect %	-18.7	-41.1	7.7	23.9	7.5	-20.2	56.2
Panel B: Min							
New PCP low-SES x Patient low-SES x Post	-0.00112**	-0.00040	0.14080***	1.15863	0.00853***	-0.00200***	0.00865**
	(0.00057)	(0.00025)	(0.04832)	(3.48727)	(0.00252)	(0.00073)	(0.00402)
Effect %	-21.4	-37.4	9.6	4.7	16.5	-23.6	30.1
Panel C: Mean							
New PCP low-SES x Patient low-SES x Post	-0.00115**	-0.00046*	0.12679***	3.65262	0.00683***	-0.00196***	0.01523***
	(0.00058)	(0.00025)	(0.04869)	(3.52553)	(0.00255)	(0.00074)	(0.00400)
Outcome mean	.00222	.00041	6.16178	340.17327	.10275	.00551	.0918
Gradient for high-SES physicians	.00523	.00107	1.461	24.5176	.05167	.00848	.02874
Effect %	-22.0	-43.0	8.7	14.9	13.2	-23.1	53.0
Observations	1,910,919	1,910,919	1,608,112	1,910,919	1,910,919	1,910,919	1,057,929
Patient Characteristics	Y	Y	Y	Y	Y	Y	Y
Old x new PCP FE	Y	Y	Y	Y	Y	Y	Y

Notes: The table presents the effect of physician-patient SES concordance on selected outcomes, see column heading. All columns report the estimates from the triple differences equation 3.3. Patient characteristics include age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education. The panels use different ways of aggregating physician SES to clinic SES. Panel A ("max") defines a clinic to be low-SES if at least one physician is low-SES. Panel B ("min") defines a clinic to be low-SES of all physicians are low-SES. Panel C ("mean") uses the proportion of physicians that are low-SES. Gradient for high-SES physician is the difference in the outcome variable between high and low-SES patients who have high-SES physicians in the post period, calculated as $(lowSESoutcome - highSESoutcome)$. The effect in percentage is calculated as $(Tripledifferenceestimate/gradientforhigh - SESphysician) \times 100$. Standard errors are clustered by patient ID.
** p <0.01, ** p <0.05, * p <0.1.

Table B.28: Robustness Check: Excluding Non-ethnic Danish patients

VARIABLES	(1) Death	(2) Death from CVC	(3) Number of visits	(4) Total reimbursement	(5) Statins	(6) Hospitalization COPD	(7) Diabetes control
PCP low SES x Patient low-SES x Post	-0.00133***	-0.00037**	0.10966***	7.52797***	0.00209	-0.00122**	0.01039***
	(0.00040)	(0.00016)	(0.03375)	(2.40024)	(0.00173)	(0.00050)	(0.00257)
Outcome mean	.00244	.00043	6.208341	327.24859	.1044	.00574	.09622
Gradient for high-SES physicians	.00527	.00095	1.58367	27.50425	.04843	.00884	.02503
Effect %	-25.2	-38.9	6.9	27.4	4.3	-13.8	41.5
Observations	3,405,243	3,405,243	2,845,634	3,405,243	3,405,243	3,405,243	1,913,919
Patient Characteristics	Y	Y	Y	Y	Y	Y	Y
Old x new PCP FE	Y	Y	Y	Y	Y	Y	Y

Notes: The table presents the effect of physician-patient SES concordance on main outcomes, see column heading. All columns report the estimates from the triple differences equation 3.3. Patient characteristics include age fixed effects, gender, dummy for being non-ethnic Danish, dummy for being married, and education level higher than primary school. Gradient for high-SES physician is the difference in the outcome variable between high and low-SES patients who have high-SES physicians in the post period, calculated as $(lowSESoutcome - highSESoutcome)$. The effect in percentage is calculated as $(Tripledifferenceestimate/gradientforhigh - SESphysician) \times 100$. Standard errors are clustered by patient ID.
** p <0.01, ** p <0.05, * p <0.1.

References

[1] Marcella Alsan, Owen Garrick, and Grant Graziani. Does Diversity Matter for Health? Experimental Evidence from Oakland. *American Economic Review*, 109(12):4071–4111, December 2019.

[2] D Mark Anderson, Benjamin Crost, and Daniel I Rees. Do economic downturns fuel racial animus? *Journal of Economic Behavior & Organization*, 175:9–18, 2020.

[3] Luvell Anderson and Ernie Lepore. Slurring words. *Noûs*, 47(1):25–48, 2013.

[4] Robert Warren Anderson, Noel D Johnson, and Mark Koyama. Jewish persecutions and weather shocks: 1100–1800. *The Economic Journal*, 127(602):924–958, 2017.

[5] Elisabeth Artmann, Hessel Oosterbeek, and Bas Klaauw. Do Doctors Improve the Health Care of Their Parents? Evidence from Admission Lotteries. *American Economic Journal: Applied Economics (Forthcoming)*, 2022.

[6] Philadelphia Bar Association. Resolution opposing use of racial slurs in business names, 2014.

[7] Christopher Avery, Paul Resnick, and Richard Zeckhauser. The market for evaluations. *American economic review*, 89(3):564–584, 1999.

[8] Abhijit V Banerjee and Esther Duflo. Reputation effects and the limits of contracting: A study of the indian software industry. *The Quarterly Journal of Economics*, 115(3):989–1017, 2000.

[9] Vojtech Bartos, Michal Bauer, Jana Cahlíková, and Julie Chytilová. Covid-19 crisis fuels hostility against foreigners. 2020.

[10] Gary S Becker. *The economics of discrimination*. University of Chicago press, 1957.

[11] Alan Benson, Aaron Sojourner, and Akhmed Umyarov. Can reputation discipline the gig economy? experimental evidence from an online labor market. *Management Science*, 66(5):1802–1825, 2020.

[12] Sylvie Blasco and Barbara Pertold-Gebicka. Employment policies, hiring practices and firm performance. *Labour Economics*, 25:12–24, 2013.

[13] Simon Board and Moritz Meyer-ter Vehn. Reputation for quality. *Econometrica*, 81(6):2381–2462, 2013.

[14] David Card, Alexandre Mas, and Jesse Rothstein. Tipping and the dynamics of segregation. *The Quarterly Journal of Economics*, 123(1):177–218, 2008.

[15] Amitabh Chandra, David Cutler, and Zirui Song. Who Ordered That? The Economics of Treatment Choices in Medical Care. volume 2 of *Handbook of Health Economics*, pages 397–432. Elsevier, 2011.

[16] Kerwin Kofi Charles and Jonathan Guryan. Prejudice and wages: an empirical assessment of becker's the economics of discrimination. *Journal of political economy*, 116(5):773–809, 2008.

[17] Yiqun Chen, Petra Persson, and Maria Polyakova. The Roots of Health Inequality and the Value of Intra-Family Expertise. *American Economic Journal: Applied Economics (Forthcoming)*, 2022.

[18] Raj Chetty, John N Friedman, Nathaniel Hendren, Michael Stepner, et al. How did covid-19 and stabilization policies affect spending and employment? a new real-time economic tracker based on private sector data. Technical report, National Bureau of Economic Research, 2020.

[19] Raj Chetty, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler. The Association Between Income and Life Expectancy in the United States, 2001-2014. *Jama*, 315(16):1750–1766, 2016.

[20] Stephen M Croucher, Thao Nguyen, and Diyako Rahmani. Prejudice toward asian americans in the covid-19 pandemic: the effects of social media use in the united states. *Frontiers in Communication*, 5:39, 2020.

[21] Janet Currie. Inequality at Birth: Some Causes and Consequences. *American Economic Review*, 101(3):1–22, May 2011.

[22] Janet M Currie and W Bentley MacLeod. Understanding Doctor Decision Making: The Case of Depression Treatment. *Econometrica*, 88(3):847–878, 2020.

[23] David Cutler, Jonathan S. Skinner, Ariel Dora Stern, and David Wennberg. Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending. *American Economic Journal: Economic Policy*, 11(1):192–221, February 2019.

[24] Gordon B Dahl and Anne C Gielen. Intergenerational spillovers in disability insurance. *American Economic Journal: Applied Economics*, 13(2):116–50, 2021.

[25] Amanda Dahlstrand. Defying distance? the provision of services in the digital age. *Working paper*, 2021.

[26] Dansk selskab for Almen Medicin. Iskæmiske hjerte-kar-sygdom. 2022.

[27] Dansk selskab for Almen Medicin. KOL. 2022.

[28] Dansk selskab for Almen Medicin. Type 2-diabetes. 2022.

[29] Steven J Davis and Brenda Samaniego de la Parra. Application flows. *Unpublished manuscript*, 2017.

[30] Angus Deaton. The Great Escape: Health, Wealth, and the Origins of Inequality. In *The Great Escape: Health, Wealth, and the Origins of Inequality*. Princeton University Press, 2013.

[31] Thomas S. Dee. A Teacher Like Me: Does Race, Ethnicity, or Gender Matter? *American Economic Review*, 95(2):158–165, May 2005.

[32] Chrysanthos Dellarocas. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management science*, 49(10):1407–1424, 2003.

[33] Huan Deng and Yujung Hwang. Structural analysis of xenophobia. *Available at SSRN 3958536*, 2021.

[34] Emilio Depetris-Chauvin. Fear of obama: An empirical study of the demand for guns and the us 2008 presidential election. *Journal of Public Economics*, 130:66–79, 2015.

[35] Joseph J Doyle Jr, Steven M Ewer, and Todd H Wagner. Returns to physician human capital: Evidence from patients randomized to physician teams. *Journal of health economics*, 29(6):866–882, 2010.

[36] James N Druckman and Justin W Holmes. Does presidential rhetoric matter? priming and presidential approval. *Presidential Studies Quarterly*, 34(4):755–778, 2004.

[37] Georgy Egorov, Ruben Enikolopov, Alexey Makarin, and Maria Petrova. Divided we stay home: Social distancing and ethnic diversity. Technical report, National Bureau of Economic Research, 2020.

[38] Itzik Fadlon and Jessica Van Parys. Primary care physician practice styles and patient care: Evidence from physician exits in medicare. *Journal of health economics*, 71:102304, 2020.

[39] ME Falagas, KZ Vardakas, and PI Vergidis. Under-diagnosis of common chronic diseases: prevalence and impact on human health. *International journal of clinical practice*, 61(9):1569–1579, 2007.

[40] Jason Faulkner, Mark Schaller, Justin H Park, and Lesley A Duncan. Evolved disease-avoidance mechanisms and contemporary xenophobic attitudes. *Group Processes & Intergroup Relations*, 7(4):333–353, 2004.

[41] Thomas Fujiwara, Karsten Müller, and Carlo Schwarz. The effect of social media on elections: Evidence from the united states. Technical report, National Bureau of Economic Research, 2021.

[42] Brad N. Greenwood, Seth Carnahan, and Laura Huang. Patient–physician gender concordance and increased mortality among female heart attack patients. *Proceedings of the National Academy of Sciences*, 115(34):8569–8574, 2018.

[43] Brad N. Greenwood, Rachel R. Hardeman, Laura Huang, and Aaron Sojourner. Physician–patient racial concordance and disparities in birthing mortality for newborns. *Proceedings of the National Academy of Sciences*, 117(35):21194–21200, 2020.

[44] Jennifer Fong Ha and Nancy Longnecker. Doctor-Patient Communication: A Review. *Ochsner Journal*, 10(1):38–43, 2010.

[45] Emma Hanes and Stephen Machin. Hate crime in the wake of terror attacks: Evidence from 7/7 and 9/11. *Journal of Contemporary Criminal Justice*, 30(3):247–267, 2014.

[46] R. Brian Haynes, Heather P. McDonald, and Amit X. Garg. Helping Patients Follow Prescribed Treatment - Clinical Applications. *JAMA*, 288(22):2880–2883, 12 2002.

[47] Jack Healy and Dionne Searcey. Two crises convulse a nation: A pandemic and police violence. *The New York Times*, 2020.

[48] John Heritage and Douglas W Maynard. *Communication in medical care: Interaction between primary care physicians and patients*, volume 20. Cambridge University Press, 2006.

[49] Andrew Hill, Daniel Jones, and Lindsey Woodworth. Physician–Patient Race–Match Reduces Patient Mortality. *Working paper*, 2020.

[50] Ria Ivandic, Tom Kirchmaier, and Stephen J Machin. Jihadi attacks, media and local hate crime. 2019.

[51] Remi Jedwab, Noel D Johnson, and Mark Koyama. Negative shocks and mass persecutions: evidence from the black death. *Journal of Economic Growth*, 24(4):345–395, 2019.

[52] Ginger Zhe Jin and Phillip Leslie. Reputational incentives for restaurant hygiene. *American Economic Journal: Microeconomics*, 1(1):237–67, 2009.

[53] Neeraj Kaushal, Robert Kaestner, and Cordelia Reimers. Labor market effects of september 11th on arab and muslim residents of the united states. *Journal of Human Resources*, 42(2):275–308, 2007.

[54] Maiken Ina Siegismund Kjaersgaard, Peter Vedsted, Erik Thorlund Parner, Bodil Hammer Bech, Mogens Vestergaard, Kaare Rud Flarup, and Morten Fenger-Grøn. Algorithm linking patients and general practices in denmark using the danish national health service register. *Clinical epidemiology*, 8:273, 2016.

[55] Michal Kolesár and Christoph Rothe. Inference in regression discontinuity designs with a discrete running variable. *American Economic Review*, 108(8):2277–2304, 2018.

[56] Sergei Koulayev, Emilia Simeonova, and Niels Skipper. Can Physicians Affect Patient Adherence With Medication? *Health Economics*, 26(6):779–794, 2017.

[57] Astrid Kunze and Amalia R. Miller. Women Helping Women? Evidence from Private Sector Data on Workplace Hierarchies. *The Review of Economics and Statistics*, 99(5):769–775, 12 2017.

[58] Michael Luca. Reviews, reputation, and revenue: The case of yelp. com. *Com (March 15, 2016). Harvard Business School NOM Unit Working Paper*, (12-016), 2016.

[59] Thomas Lyngaa, Christian Fynbo Christiansen, Henrik Nielsen, Mette Asbjørn Neergaard, Anders Bonde Jensen, Kristina Grønborg Laut, and Søren Paaske Johnsen. Intensive care at the end of life in patients dying due to non-cancer chronic diseases versus cancer: a nationwide study in denmark. *Critical Care*, 19(1):1–9, 2015.

[60] Hanjia Lyu, Long Chen, Yu Wang, and Jiebo Luo. Sense and sensibility: Characterizing social media users regarding the use of controversial terms for covid-19. *IEEE Transactions on Big Data*, 2020.

[61] Alexandre Mas and Amanda Pallais. Valuing alternative work arrangements. *American Economic Review*, 107(12):3722–59, 2017.

[62] Justin McCrary. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics*, 142(2):698–714, 2008.

[63] Ellen R. Meara, Seth Richards, and David M. Cutler. The Gap Gets Bigger: Changes In Mortality And Life Expectancy, By Education, 1981–2000. *Health Affairs*, 27(2):350–360, 2008.

[64] Ministeriet for Sundhed og Forebyggelse. Ulighed i sundhed – kroniske og langvarige sygdomme. Technical report, Ministeriet for Sundhed og Forebyggelse, 2014.

[65] Karsten Müller and Carlo Schwarz. From hashtag to hate crime: Twitter and anti-minority sentiment. *Available at SSRN 3149103*, 2019.

[66] S. Mullis and H. Glenn. New site collects reports of racism against asian americans amid coronavirus pandemic, 2020.

[67] Thu T Nguyen, Hsien-Weng Meng, Sanjeev Sandeep, Matt McCullough, Weijun Yu, Yan Lau, Dina Huang, and Quynh C Nguyen. Twitter-derived measures of sentiment towards minorities (2015–2016) and associations with low birth weight and preterm birth in the united states. *Computers in human behavior*, 89:308–315, 2018.

[68] David Novgorodsky and Bradley Setzler. Practical guide to event studies, 2019.

[69] OECD. Health for Everyone? Social Inequalities in Health and Health Systems. *OECD Health Policy Studies*, 2019.

[70] Ady Oster and Andrew B Bindman. Emergency Department Visits for Ambulatory Care Sensitive Conditions: Insights Into Preventable Hospitalizations. *Medical care*, pages 198–207, 2003.

[71] Lars Osterberg and Terrence Blaschke. Adherence to medication. *New England journal of medicine*, 353(5):487–497, 2005.

[72] Paul Oyer and Scott Schaefer. Personnel economics: Hiring and incentives. Technical report, National Bureau of Economic Research, 2010.

[73] Brian A O'Shea, Derrick G Watson, Gordon DA Brown, and Corey L Fincher. Infectious disease prevalence, not race exposure, predicts both implicit and explicit racial prejudice across the united states. *Social Psychological and Personality Science*, 11(3):345–355, 2020.

[74] Praktiserende Lægers Organisation. 2/3 af landets læger har nu lukket for flere patienter. Technical report, 2017.

[75] Praktiserende Lægers Organisation. Antallet af ældre patienter hos de praktiserende læger stiger markant. Technical report, 2019.

[76] Wullianallur Raghupathi and Viju Raghupathi. An Empirical Study of Chronic Diseases in the United States: A Visual Analytics Approach to Public Health. *International Journal of Environmental Research and Public Health*, 15(3):431, 2018.

[77] Colin D. Rehm, José L. Peñalvo, Ashkan Afshin, and Dariush Mozaffarian. Dietary Intake Among US Adults, 1999-2012. *JAMA*, 315(23):2542–2553, 06 2016.

[78] Rafael Rob and Arthur Fishman. Is bigger better? customer base expansion through word-of-mouth reputation. *Journal of political Economy*, 113(5):1146–1162, 2005.

[79] Arlyss Anderson Rothman and Edward H Wagner. Chronic Illness Management: What Is the Role of Primary Care? *Annals of Internal Medicine*, 138(3):256–261, 2003.

[80] Naveed Sattar, David Preiss, Heather M Murray, Paul Welsh, Brendan M Buckley, Anton JM de Craen, Sreenivasa Rao Kondapally Seshasai, John J McMurray, Dilys J Freeman, J Wouter Jukema, Peter W Macfarlane, Chris J Packard, David J Stott, Rudi G Westendorp, James Shepherd, Barry R Davis, Sara L Pressel, Roberto Marchioli, Rosa Maria

Marfisi, Aldo P Maggioni, Luigi Tavazzi, Gianni Tognoni, John Kjekshus, Terje R Pedersen, Thomas J Cook, Antonio M Gotto, Michael B Clearfield, John R Downs, Haruo Nakamura, Yasuo Ohashi, Kyoichi Mizuno, Kausik K Ray, and Ian Ford. Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials. *The Lancet*, 375(9716):735–742, 2010.

[81] Scandinavian Simvastatin Survival Study Group. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the scandinavian simvastatin survival study (4s). *The Lancet*, 344(8934):1383–1389, 1994. Originally published as Volume 2, Issue 8934.

[82] Mark Schaller and Steven L Neuberg. Danger, disease, and the nature of prejudice (s). In *Advances in experimental social psychology*, volume 46, pages 1–54. Elsevier, 2012.

[83] Leonard Schild, Chen Ling, Jeremy Blackburn, Gianluca Stringhini, Yang Zhang, and Savvas Zannettou. " go eat a bat, chang!": An early look on the emergence of sinophobic behavior on web communities in the face of covid-19. *arXiv preprint arXiv:2004.04046*, 2020.

[84] Molly Schnell and Janet Currie. Addressing the Opioid Epidemic: Is There a Role for Physician Education? *American Journal of Health Economics*, 4(3):383–410, 2018.

[85] Smriti Sharma. Caste-based crimes and economic status: Evidence from india. *Journal of comparative economics*, 43(1):204–226, 2015.

[86] Laura Ruth Silver. China in the media: Effects on american opinion. 2016.

[87] Emilia Simeonova. Doctors, patients and the racial mortality gap. *Journal of Health Economics*, 32(5):895–908, 2013.

[88] Emilia Simeonova, Niels Skipper, and Peter Rønø Thingholm. Physician Health Management Skills and Patient Outcomes. 2022.

[89] Marianne Simonsen, Lars Skipper, Niels Skipper, and Peter Rønø Thingholm. Discontinuity in care: Practice closures among primary care providers and patient health care utilization. *Journal of Health Economics*, 80:102551, 2021.

[90] Jason Sockin and Aaron Sojourner. What's the inside scoop? challenges in the supply and demand for information on employers. Technical report, Technical Report 3596666, Social Science Research Network, 2020.

[91] Seth Stephens-Davidowitz. The cost of racial animus on a black candidate: Evidence using google search data. *Journal of Public Economics*, 118:26–40, 2014.

[92] Richard L. Street. Information-giving in medical consultations: The influence of patients' communicative styles and personal characteristics. *Social Science & Medicine*, 32(5):541–548, 1991.

[93] Liyang Sun and Sarah Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 2020.

[94] Sundhedsstyrelsen. Kræftprofil - lungekræft 2000-2006. Technical report, Sundhedsstyrelsen, 2009.

[95] Sundhedsstyrelsen. Sygdomsbyrden i Danmark. Technical report, Sundhedsstyrelsen, 2015.

[96] Steven Tadelis. The market for reputations as an incentive mechanism. *Journal of political Economy*, 110(4):854–882, 2002.

[97] The Association of the Nordic Cancer Registries. Breast cancer - Age-standardized relative survival (%), Females, diagnosed 2015-2019. 2022.

[98] The Association of the Nordic Cancer Registries. Lung cancer - Age-standardized relative survival (%), diagnosed 2015-2019. 2022.

[99] Rachel L. Johnson Thornton, Neil R. Powe, Debra Roter, and Lisa A. Cooper. Patient–physician social concordance, medical visit communication and patients' perceptions of health care quality. *Patient Education and Counseling*, 85(3):e201–e208, 2011.

[100] Daniel B Turban and Daniel M Cable. Firm reputation and applicant pool characteristics. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, 24(6):733–751, 2003.

[101] Brendesha M Tynes, Michael T Giang, David R Williams, and Geneene N Thompson. Online racial discrimination and psychological adjustment among adolescents. *Journal of adolescent health*, 43(6):565–569, 2008.

[102] Undervisningsministeriet. Uddannelsessystemet i tal gennem 150 år- Undervisningsministeriet 1848 - 1998. Technical report, Undervisningsministeriet, 1998.

[103] Edward H Wagner. The role of patient care teams in chronic disease management. *BMJ*, 320(7234):569–572, 2000.

[104] Di Wang. Xiuli wang, winning american hearts and minds: China's image building efforts in the 21st century. *International Journal of Communication*, 15:4, 2021.

[105] S. Willems, S. De Maesschalck, M. Deveugele, A. Derese, and J. De Maeseneer. Socio-economic status of the patient and doctor–patient communication: does it make a difference? *Patient Education and Counseling*, 56(2):139–146, 2005.

[106] Caleb Ziems, Bing He, Sandeep Soni, and Srijan Kumar. Racism is a virus: Anti-asian hate and counterhate in social media during the covid-19 crisis. *arXiv preprint arXiv:2005.12423*, 2020.

[107] Kelly B Haskard Zolnierek and M Robin DiMatteo. Physician Communication and Patient Adherence to Treatment: A Meta-analysis. *Medical care*, 47(8):826, 2009.