UNIVERSITY OF CALIFORNIA

Los Angeles

Essays in Health Economics

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

by

Johnny Huynh

2024

ABSTRACT OF THE DISSERTATION

Essays in Health Economics

by

Johnny Huynh

Doctor of Philosophy in Economics
University of California, Los Angeles, 2024
Professor Adriana Lleras-Muney, Co-Chair
Professor Kathleen M. McGarry, Co-Chair

This dissertation studies the effectiveness of public policies aimed at improving population health and narrowing health disparities. It is divided into three chapters. Chapter one provides new evidence on the health effects of cash transfers for people with disabilities. The chapter evaluates the U.S. Department of Veterans Affairs' Disability Compensation program, one of the nation's largest cash transfer programs. To accomplish this, a large administrative dataset is compiled by combining disability claims, medical records, and death records for over 800,000 veterans. The empirical strategy leverages the quasi-random assignment of disability claims to examining caseworkers and physicians, comparing veterans whose disability claims were assigned to examiners with varying levels of leniency. The results indicate that receiving $300 per month lowers five-year mortality by 0.8 percentage points, or 9.5% of the mean rate.

The second chapter analyzes the prescribing behavior of nurse practitioners (NPs) and primary care physicians in states that allow NPs to prescribe medications independently. Inappropriate prescriptions are defined as drugs that are generally considered inappropriate for adults aged 65 years or older, according to the American Geriatrics Society's Beers

Criteria. The mean rates of inappropriate prescribing by NPs and primary care physicians are nearly the same. However, NPs are overrepresented among clinicians with the highest and lowest rates of inappropriate prescribing.

The third chapter examines the adverse motivational effects of monetary incentives for vaccines. A survey experiment was conducted among 513 vaccine-hesitant adults between May and September 2021, priming half with the idea of monetary incentives for the COVID-19 vaccine. The findings show that monetary incentives discourage 14% of participants from getting vaccinated, a significant reduction since only one in four considered getting vaccinated at the outset. Participants in the primed group are more likely to believe that vaccines are unsafe and more likely to believe that people do not have a responsibility to get vaccinated, compared to participants in the control group.

The dissertation of Johnny Huynh is approved.

Martha Jane Bailey

Felipe M. Goncalves

Adriana Lleras-Muney, Committee Co-Chair

Kathleen M. McGarry, Committee Co-Chair

University of California, Los Angeles

2024

Contents

List of Figures

List of Tables

ACKNOWLEDGMENTS

I express deep gratitude to my committee members, Adriana Lleras-Muney, Kathleen McGarry, Martha Bailey, and Felipe Goncalves, whose unwavering support, guidance, and feedback were pivotal throughout the course of my doctoral research. I am also grateful to my mentors, Kitt Carpenter, David Chan, Keshia Pollack Porter, and David Studdert. Special thanks to my colleagues and friends, whose camaraderie and intellectual discussions provided both motivation and perspective. Finally, I am indebted to my family for their boundless patience, understanding, and love.

This research benefited from support from the Robert Wood Johnson Foundation (77833), the National Science Foundation (2034835), the California Center for Population Research at UCLA (P2CHD041022), and the National Bureau of Economic Research (RRC08098400).

VITA

Johnny Huynh

EDUCATION

UNIVERSITY OF CALIFORNIA, LOS ANGELES

M.A. in Economics 2018-2019

POMONA COLLEGE

B.A. in Mathematics and Economics 2010-2014

PUBLICATIONS

"Accuracy of Valuations of Surgical Procedures in the Medicare Fee Schedule" (with David Chan and David Studdert). *New England Journal of Medicine*, 2019

"Understanding the Gap Between Efficacy in RCTs and Effectiveness in Real-World Use of GLP-1RA and DPP4 in Patients with Type II Diabetes" (with Ginger Carls, Steven Edelman, William Polonsky, Ruo-Ding Tan, Edward Tuttle, and John Yee). *Diabetes Care*, 2017

"Achievement of Glycated Hemoglobin Goals in the US Remains Unchanged Through 2014" (with Ginger Carls, Steven Edelman, Edward Tuttle, and John Yee). *Diabetes Therapy*, 2017

RESEARCH IN PROGRESS

"Do Cash Transfers Narrow Health Disparities? Evidence from Veterans with Disabilities"

"Vaccine Incentives Hurt Intrinsic Motivation: Evidence from a Survey Experiment"

"Achieving Race-Based Outcomes with Race-Blind Policies: Evidence from Col-

lege Admissions"

"Washed Away: Lasting Effects of the Ohio Flood of 1913"

"Did PrEP Reduce HIV Infections? Endogenous Take-up and Efficacy"

Relevant Positions

University of California Los Angeles

Student Affiliate, California Center for Population Research	2020-2024
Graduate Student Researcher, California Policy Lab	2021

Palo Alto Veterans Institute for Research

Research Analyst	2016-2018

Fellowships, Honors, and Awards

Fellowship in Retirement and Disability Policy Research, NBER.	2022-2024
Health Policy Research Scholars, Robert Wood Johnson Foundation.	2020-2024
Graduate Research Fellowship Program, National Science Foundation.	2020-2023
Best Paper Award, Proseminar in Applied Microeconomics, UCLA.	2023
Diversity Scholarship, American Society of Health Economists.	2023
Outstanding Teaching Award, UCLA.	2020
Outstanding Remote Teaching Award, UCLA.	2020
Morris B. Pendleton Prize, Pomona College.	2014
College Match Scholarship, QuestBridge.	2010-2014

Introduction

The first chapter of this dissertation studies the impact of cash transfer programs on health disparities. Cash transfer programs provide financial payments to individuals or families, typically aimed at reducing poverty and improving living standards. In addition, these programs can influence health production by enhancing access to health care services, improving nutrition, and reducing stress related to financial instability. In *Do Cash Transfers Narrow Health Disparities? Evidence from Veterans with Disabilities*, I investigate whether cash transfers narrow health disparities among veterans with disabilities, leveraging two natural experiments and novel administrative data. I specifically evaluate the U.S. Department of Veterans Affairs' Disability Compensation program, one of the nation's largest cash transfer programs, which assists veterans with service-connected disabilities. To accomplish this, I compile a large administrative dataset combining disability claims, medical records, and death records for over 800,000 veterans.

Veterans receiving cash transfers from Veterans Affairs differ from nonrecipients in many dimensions, generating selection bias in correlational estimates. By design, the program awards larger cash transfers to veterans with more severe disabilities, giving the impression that transfers harm health because individuals with severe disabilities generally have worse health outcomes in the absence of compensating transfers. My empirical strategy leverages the quasi-random assignment of disability claims to examining caseworkers and physicians, comparing veterans whose disability claims were assigned to more or less lenient examiners.

I find that receiving $300 per month lowers five-year mortality by 0.8 percentage points, or 9.5% of the mean rate.

I operationalize the roles of caseworkers and physicians using two instruments and investigate why my results differ depending on the instrument used. Using the caseworker instrument, I find that receiving $300 per month lowers five-year mortality by 1.0 pp. However, the physician instrument recovers a significantly smaller effect of 0.4 pp. I reconcile this discrepancy using a framework of heterogeneous treatment effects, in which differences in treatment effects stem from differences between compliers. I then expand on these results by modeling selection into treatment. I find that veterans in the poorest health receive larger cash transfers but have the least capacity to convert these transfers into health gains. In this scenario, positive selection on levels leads to reverse selection on gains.

Veterans with disabilities suffer from worse health compared to non-disabled veterans. I find that disability compensation improves health for the typical recipient, thus narrowing health disparities between veterans with and without disabilities. However, veterans in the poorest health gain the least from disability compensation, exacerbating health disparities among veterans with disabilities and underscoring an important equity-efficiency tradeoff in cash transfer policies.

There is a shortage of primary care physicians in the United States, and there is considerable potential for nurse practitioners (NPs) and other nonphysician providers to expand access to primary care. However, restrictive scope of practice laws and quality of care concerns limit their roles. Many states, including 16 in the last decade, have legislated to allow NPs to independently prescribe drugs. Critics contend these moves will adversely affect the quality of care.

The second part of this dissertation contributes to our understanding of this debate by comparing rates of inappropriate prescribing among NPs and primary care physicians. In *Inappropriate Prescribing to Older Patients by Nurse Practitioners and Primary Care*

Physicians, I calculate and compare rates of inappropriate prescribing for 23,669 NPs and 50,060 primary care physicians who wrote prescriptions for $\geq$ 100 patients per year, adjusting for practice experience, patient volume and risk, clinical setting, year, and state. Our data include Medicare Part D beneficiaries ages 65 and older in 2013 to 2019. Inappropriate prescriptions are defined as drugs that typically should not be prescribed for adults aged 65 years or older, according to the American Geriatrics Society's Beers Criteria.

I find that mean rates of inappropriate prescribing by NPs and primary care physicians are virtually identical. The adjusted odds ratio is 0.99. However, NPs are overrepresented among clinicians with the highest and lowest rates of inappropriate prescribing. Indeed, there are substantial differences in the distribution of inappropriate prescribing, with NPs more likely to be among both the most compliant and least compliant prescribers.

For both types of practitioners, discrepancies in inappropriate prescribing rates across states tended to be larger than discrepancies between these practitioners within states. Since NPs are no more likely than physicians to prescribe inappropriately to older patients, this study underscores the importance of addressing performance deficiencies among all clinicians rather than focusing solely on whether nonphysician providers should be allowed to prescribe. Policymakers are encouraged to use clinician-level performance measures and implement technological interventions and guidelines adherence initiatives to improve prescribing practices.

Vaccines are an integral part of controlling the COVID-19 pandemic and preventing future outbreaks. However, vaccine hesitancy, or the reluctance or refusal to get vaccinated despite the availability of vaccines, threatens their effectiveness in the real world. Vaccine hesitancy can stem from various factors, including misinformation, distrust in pharmaceutical companies or government agencies, fear of potential side effects, and cultural or religious beliefs. Financial incentives have been proposed as a potential intervention to overcome vaccine hesitancy. By providing monetary rewards or other financial benefits for getting vaccinated, these incentives can motivate individuals to reconsider their stance on vaccination, but the effectiveness of such incentives can vary depending on the underlying reasons for hesitancy and the demographic characteristics of the hesitant population.

The third chapter of this dissertation, *Vaccine Incentives Hurt Intrinsic Motivation: Evidence from a Survey Experiment*, examines the adverse motivational effects of monetary incentives. During the COVID-19 pandemic, some economists suggested financial incentives to boost vaccination rates among hesitant adults, arguing that the social benefits would justify the costs. Despite this, critics warned that monetary incentives might backfire by undermining intrinsic motivation, potentially decreasing the likelihood of vaccination among some individuals.

Monetary incentives exert two opposing effects: the price effect, which makes the incentivized behavior more attractive by reducing its relative cost, and the psychological effect, which can undermine intrinsic motivation by introducing external rewards. This dual nature creates ambiguity about their overall impact. To empirically test these effects, I conduct a survey experiment involving 513 vaccine-hesitant adults, employing behavioral priming to distinguish the psychological effect from the price effect. The findings imply that monetary incentives deter 14% of participants who might have otherwise considered vaccination, with deterred individuals exhibiting greater baseline vaccine skepticism.

The channels through which payments discourage vaccination include perceptions of vaccine safety and intrinsic motivations. Financial incentives might signal that vaccines are unsafe or crowd out prosocial motivations by making individuals doubt their prosocial intentions. The analysis shows that participants exposed to incentives are more likely to perceive vaccines as unsafe and to question their responsibility to get vaccinated. This study contributes to understanding vaccine demand determinants and suggests that monetary incentives may not always be the best strategy for increasing vaccination coverage.

Chapter 1

Do Cash Transfers Narrow Health Disparities? Evidence from Veterans with Disabilities

1. Introduction

People with disabilities generally experience worse health compared to those without disabilities.[1] In part, disability-related health disparities reflect structural barriers that disproportionately affect this population, such as poverty, inadequate access to medical care, and housing precarity (World Health Organization 2022). Recognizing that the root causes of health disparities may be economic in nature, governments around the world have implemented cash transfers to lessen the economic hardship faced by people with disabilities.[2] Do

[1]People with disabilities are twice as likely to die prematurely, resulting in a reduction in life expectancy of 10 to 20 years. In the United States, for instance, adults with disabilities are more likely than those without disabilities to report having poor or fair health (40.3% vs. 9.9%), have elevated rates of heart disease, diabetes, and certain cancers, and are less likely to receive preventive health care (Kuper et al. 2022; Forman-Hoffman et al. 2015; Krahn et al. 2015; Reichard et al. 2011). Disability-related health disparities are equally pronounced in low- and middle-income countries (Smythe and Kuper 2023).

[2]Most countries (170 out of the 186 for which information is available) offer disability-specific cash benefits to their citizens (United Nations 2018). Cash transfers for people with disabilities typically come in

cash transfers narrow health disparities? The evidence on the effectiveness of cash transfers in improving health is far from unified, and comparing estimates across studies is complicated by differences in their methodology, settings (e.g., disability insurance schemes in the U.S. (Gelber et al. 2023) versus the Netherlands (García-Gómez and Gielen 2018)) and the size of cash transfers (ranging from 0.5% to 4,000% of median annual income).[3] Furthermore, the differences across causal estimates could suggest that the effect of an identical transfer varies across individuals within the same population — a concept known as essential heterogeneity (Heckman et al. 2006). However, evidence on heterogeneity in the health effects of cash transfers is scarce, particularly for individuals with disabilities.

This paper examines the heterogeneous effects of cash transfers for veterans with disabilities. The experience of few groups is more germane to this issue than that of veterans with disabilities, as they are among the largest beneficiaries of redistributive policies in the U.S.[4] Veterans also have access to comprehensive health insurance, whether or not they receive cash transfers, presenting a useful contrast with populations for whom transfers greatly expand access to medical care. Our setting is the Department of Veteran Affairs' (VA) Disability Compensation program, which provides financial support for veterans whose disabilities resulted from their military service. We use a large, administrative dataset combining disability claims, medical records, and death records for over 800,000 veterans to estimate the causal effect of cash transfers on mortality and other health outcomes and to quantify heterogeneity in this relationship. The richness of our data lets us analyze medical conditions, healthcare utilization, and migration to shed light on the mechanisms driving these treatment effects.

To isolate the causal effect of cash transfers on health, we use an instrumental variables (IV) strategy that leverages the quasi-random assignment of disability claims to two agents:

the form of social insurance or compensatory benefits.

[3] See Deaton (2002), Currie (2009), and Cutler et al. (2012) for an in-depth overview of the debate regarding the causal relationship between income and health, with an emphasis on confounding and reverse causation in estimating this relationship.

[4] In 2021, federal outlays on cash transfers for veterans with disabilities totaled $112 billion, benefiting 5.6 million veterans and their families.

caseworkers and physicians. Following the literature relying on quasi-random appointment of judges for identification, we make use of the fact that veterans seeking disability compensation undergo assessments from a caseworker and a physician. VA policy prohibits veterans from choosing their assessors; the matching process is dictated by availability. We validate this research design by showing that a rich vector of veteran characteristics cannot predict caseworker or physician assignments. We further demonstrate that both caseworkers and physicians vary in their leniency. Veterans assigned to caseworkers and physicians in the 95^{th} percentile of leniency receive approximately $300 more per month (in 2018 dollars) than those assigned to agents in the 5^{th} percentile.

Our IV strategy compares health outcomes for otherwise similar veterans who received more (less) generous cash transfers because their claim was assigned to relatively lenient (stringent) caseworkers and physicians. Using the relative leniency of caseworkers and physicians as instruments, we find that receiving $300 per month lowers 5-year mortality by 0.8 percentage points (pp), or 9.5% of the mean. In addition, more generous cash transfers lead to increased utilization of preventive medical care, reduced reliance on homelessness services, and no change in utilization of emergency room or inpatient services. There is also a modest reduction in the incidence of metabolic diseases among recipients.

However, not all veterans benefit equally from the same transfer. We uncover patterns of heterogeneity consistent with results in the health disparities literature, and our findings provide new insights into the mechanisms driving these differences. First, for veterans claiming a mental health impairment, receiving $300 per month reduces 5-year mortality by 1.1 pp, but the same transfer has precisely zero effect for veterans claiming a musculoskeletal impairment. Understanding treatment-effect heterogeneity by impairment type is important. Given the rising prevalence of impairments associated with lower mortality (e.g., mental health and musculoskeletal impairments) in VA and non-VA disability rolls, there is widespread concern that providing cash transfers for these types of impairments reflects a growing misuse of disability programs (Autor and Duggan 2006; Liebman 2015).[5]

The second form of heterogeneity that we uncover is geographic. We find that the health effect of cash transfers varies substantially across commuting zones, with smaller estimates observed in regions with a large manufacturing presence. In fact, cash transfers appear to worsen mortality in some commuting zones. Our place-based results add to emerging research on the causal effect of place on health (Finkelstein et al. 2021; Deryugina and Molitor 2020; Baum et al. 2020). Our finding that cash transfers provide minimal benefit in manufacturing commuting zones also relates to research on the detrimental health effects of industrial decline in "rust belt" regions (Lang et al. 2019; Pierce and Schott 2020; Autor et al. 2019) and suggests that cash transfers alone may be insufficient to offset the adverse effects of deteriorating economic conditions in these regions.

A novel feature of our study is that we leverage the fact that both caseworkers and physicians help determine the cash transfer that veterans ultimately receive. We operationalize their roles in the determination process using two instruments and investigate why our results differ depending on the instrument used. Using only the caseworker instrument, we find that receiving $300 per month reduces 5-year mortality by 1.0 pp. The physician instrument, however, recovers a significantly smaller effect of 0.4 pp. We reconcile this discrepancy using a framework featuring heterogeneous treatment effects and heterogeneous choice behavior. In this framework, caseworkers and physicians have different objectives and information, with their assessments leading to different types of veterans receiving larger transfers (i.e., compliers), such that we identify distinct local average treatment effects. We find that caseworker compliers exhibit lower medical acuity than physician compliers.

We conclude by exploring patterns of treatment selection for caseworkers and physicians,

[5]The purpose of disability insurance schemes, such as Social Security Disability Insurance (SSDI) and Supplemental Security Income (SSI) in the U.S., is to assist people who cannot work due to an impairment. The practice of providing cash transfers to individuals with less severe health conditions may dilute the overall value of these programs. On the other hand, Deshpande and Lockwood (2022) show that insurance against nonhealth risk accounts for half of the value of disability programs because recipients with less severe conditions frequently experience nonhealth shocks.

in the style of Heckman and Vytlacil (2005) and Mogstad et al. (2018a). We do this by estimating marginal treatment effects, which relate veterans' propensity to receive cash transfers to the health effect of these transfers. For instance, physicians, due to their expertise in diagnosing health conditions, might be more inclined than caseworkers to target transfers based on medical acuity. This is indeed what we find. However, while veterans with greater medical acuity receive more generous cash transfers, they have the least capacity to convert transfers into health gains. In this scenario, positive selection on "levels" leads to reverse selection on "slopes", otherwise known as reverse-Roy selection.[6]

Our study contributes to three important strands of literature. First, it deepens our understanding of income and poverty as social determinants of health.[7] Vulnerable populations such as people with disabilities have vast material needs, and addressing health disparities requires a deeper understanding of the effectiveness of policies aimed at meeting those needs. A wide array of studies has estimated the causal effect of income on health, with many of them analyzing publicly funded cash transfer programs like the VA's disability program.[8] However, the existing literature provides no clear explanation for why the health effects of cash transfers vary so much (Lleras-Muney 2022). Our findings underscore the complexity of this causal relationship; we demonstrate within a single study that cash transfers can have both beneficial and nonbeneficial effects. Because we do this within the same setting, using parallel identification strategies that differ only in the agent influencing the decision, we rule

[6]Selection on gains from treatment, and reverse selection on gains in particular, has been observed in numerous other settings, such as studies on the returns to education (Carneiro et al. 2011; Walters 2018; Nybom 2017), early childhood programs (Cornelissen et al. 2018; Kline and Walters 2016), racial discrimination (Arnold et al. 2018; Arnold et al. 2022), housing and migration (Chyn 2018), electricity markets (Ito et al. 2023), and health care (Einav et al. 2022; Depalo 2020; Kowalski 2023).

[7]The social determinants of health refer to nonmedical factors that influence health and can be affected by social policies. Dating back to the Whitehall studies (Marmot et al. 1978), which found a link between occupational rank and heart disease, income has been widely recognized as a social determinant of health. Despite having a strong theoretical basis (Grossman 1972), the causal pathways connecting income and health are not empirically well understood (Braveman and Gottlieb 2014).

[8]Previous studies have estimated their effect on adults' health using a wide range of natural experiments, such as lotteries (Cesarini et al. 2016; Lindahl 2005; Apouey and Clark 2015; Lindqvist et al. 2020; Kim and Koh 2021) and pension reforms (Miglino et al. 2023; Malavasi and Ye 2023; Snyder and Evans 2006; Jensen and Richter 2004; Case 2004; Cheng et al. 2018; Aizer et al. 2023; Feeney 2018; Salm 2011).

out methodological differences as an explanation. Instead, essential heterogeneity seems to be driving these variations.

The study most closely related to ours is by Silver and Zhang (2023), who study the health effect of disability compensation for veterans with mental health disorders, leveraging the quasi-random assignment of claims to physician examiners. They find no effect on mortality outcomes. We replicate their findings and further show that the population they study (veterans with mental health disorders) and the identifying variation they use (physician leniency) lead them to their specific conclusion. Using richer variation and a broader sample, we show that the causal effect of disability compensation is heterogeneous, with only certain veterans benefiting from the cash transfer. To our knowledge, our finding that individuals in the poorest health sometimes benefit the least from cash transfers is new. At the end of our paper, we discuss the reasons why this is the case; one's ability to use income beneficially may be constrained by the severity of their impairment or economic circumstances.

Second, this paper contributes to the "judges design" literature, which identifies causal effects by exploiting quasi-random assignment to agents with varying treatment propensities.[9] While previous studies have focused on one type of agent (e.g., court judges, child welfare investigators), we innovate by analyzing two types of agents involved in the VA's disability determination process: caseworkers and physicians. This determination process mirrors other high-stakes settings in which decision-makers receive signals from more than one agent, such as judges hearing testimonies from multiple witnesses or journal editors reconciling reports from multiple referees. We empirically demonstrate that treatment effects are "local" to each type of agent because their corresponding sets of compliers do not fully overlap — a result that builds upon recent theoretical research on partial monotonicity in the presence of multiple instruments (Mogstad et al. 2020, 2021).

[9]The "judges design" has been employed in a variety of settings, such as incarceration (Kling 2006; Aizer and Doyle Jr 2015; Arteaga 2023; Bhuller et al. 2020; Eren and Mocan 2021; Norris et al. 2021, 2022), foster care (Doyle Jr 2007, 2008; Gross and Baron 2022; Bald et al. 2022), and health care (Blæhr and Søgaard 2021; Bakx et al. 2020; Chan et al. 2022). See Frandsen et al. (2023) for an overview of this literature.

Our final contribution is to document that the process for adjudicating disability claims can be enhanced and, more broadly, to demonstrate that targeting cash transfers in disability programs is crucial for maximizing their health gains. Our study relates to a significant body of literature evaluating the effectiveness of U.S. disability programs, including studies estimating their impact on labor market outcomes (Maestas et al. 2013; Gruber and Kubik 1997; French and Song 2014; Gelber et al. 2017; Autor and Duggan 2003, 2006). Another strand of literature has focused on the long-term outcomes of children with disabilities (Deshpande 2016; Deshpande and Mueller-Smith 2022; Guldi et al. 2022). Our paper centers on veterans with disabilities, relating to research by Autor et al. (2016), Silver and Zhang (2023), Eli et al. (2023), Coile et al. 2021, and Duggan et al. (2010). Although disability programs have drawn criticism for inadequate targeting of benefits, we find that cash transfers do improve the health of the typical recipient, thus narrowing health disparities between veterans with and without disabilities. However, cash transfers are less effective at reducing mortality for recipients in the poorest health and, consequently, might exacerbate health disparities among veterans with disabilities. Indeed, if the VA's goal is to maximize veterans' overall health, then its disability compensation program can be better targeted. However, increasing efficiency by targeting transfers could undermine the program's broader goal of assisting veterans with the most severe service-connected disabilities, underscoring an important equity-efficiency tradeoff in this policy decision.

This paper is organized as follows. Section 2. provides an overview of the VA's disability compensation program. Section 3. describes the data and sample, and Section 4. outlines our empirical strategy. Section 5. presents the main results for mortality and other outcomes. Section 6. analyzes treatment effect heterogeneity. Section 7. interprets our causal estimates using a behavioral choice framework. Section 8. discusses the implications of our results for health disparities.

2. Background

The VA's disability compensation program provides financial support for veterans with service-connected disabilities, and the generosity of this program varies according to recipients' disability rating. This section provides a primer on the compensatory program and the process used by the VA to determine disability ratings.

2.1 Disability compensation program

The U.S. Department of Veterans Affairs administers one of the nation's largest cash transfer programs, disbursing over $112 billion to 5.6 million veterans (1 in 4 living veterans) and their families in 2021 (Government Accountability Office 2022).[10] The program was created to compensate veterans for disabilities that were caused or aggravated by their active-duty military service. While the origins of the VA's disability compensation program can be traced to the American revolutionary war, the modern-era program that we study was established alongside the creation of the VA as a cabinet-level agency in 1989. In the past two decades, the proportion of veterans receiving disability compensation has tripled, a trend driven by expanded eligibility criteria for veterans who served in Vietnam, Cambodia, Laos, and the surrounding waters (Coile et al. 2021).

The VA's disability compensation program provides stable and long-lasting income assistance to eligible veterans. The generosity of this program is a deterministic function of a veteran's service-connected disability rating, ranging from 0% to 100% and increasing in increments of 10%, as shown in Appendix Section 4.1. In 2018, an unmarried veteran with a 10% disability rating received $136 per month, while a veteran with a 100% disability rating received $2,974 per month. The VA's gradation system stands in contrast to the systems in

[10] Fewer than 1 in 10 American adults are veterans, but total expenditure on the VA's disability compensation program exceeds federal spending on either the Supplemental Nutrition Assistance Program or the Earned Income Tax Credit. It equals approximately 85% of spending on Social Security Disability Insurance.

other disability programs such as SSDI, which uses an all-or-nothing system to determine whether a claimant receives the benefit. For veterans with multiple service-connected disabilities, the combined disability rating, which is a concave function of individual ratings, determines the cash transfer amount.[11] Transfers are annually adjusted for inflation.

Disability compensation payments are disbursed monthly, are not means tested, and modestly scale up for veterans with a spouse or child dependent. The payments are exempt from federal and state income taxes, so $1 in benefits equates to approximately $1.20 to $1.30 in earned income, depending on the veteran's marginal tax rate. Once awarded, disability compensation is seldom revoked and typically continues beyond retirement age. Unlike SSDI beneficiaries, most VA disability compensation beneficiaries may engage in gainful employment.[12] The labor force participation rate for this group was 36% in 2022. In this regard, compared to disability insurance schemes that restrict gainful employment, the VA's program more closely resembles an unconditional cash transfer.

2.2 Disability rating process

The VA employs a comprehensive review process to determine service-connected disability ratings and, as a result, the amount of compensation that each veteran receives. To apply for disability compensation, a veteran submits a disability claim to one of 57 regional Veterans Benefits Administration offices across the country. The office assigns the disability claim to a Veterans Service Representative (henceforth, caseworker), who assists the applicant in navigating the benefits system and manages the claim throughout the application process. The

[11] The combined disability rating is calculated based on "remaining efficiency", which recognizes the compounding nature of coexisting disabilities. Suppose that a veteran has two disabilities: 40% rating and 20% rating. First, the veteran is considered 40% disabled and has 60% remaining efficiency. Then, the subsequent 20% rating applies to the remaining efficiency, leading to a 12% reduction (= 20% $\times$ 60%). Both ratings sum to 52%, which rounds to a 50% combined disability rating.

[12] One exception is the VA's individual unemployability rule, which compensates veterans with a 100% disability rating. Eligible veterans must demonstrate they cannot engage in gainful employment due to their service-connected disability, with either one disability rated $\geq$ 60% or multiple disabilities that combine to $\geq$ 70%.

caseworker is tasked with interviewing the veteran and collecting relevant documents, such as medical records, military service records, and any evidence needed to substantiate whether the claimed impairment was caused or aggravated by military service, giving consideration to the VA's national criteria.

During the disability rating process, a medical examination is scheduled with a physician, who assesses the severity of the claimed impairment.[13] This visit is purely investigative, and VA regulations do not allow the veteran and physician to interact outside of their scheduled visit. The physician is tasked with reviewing the veteran's medical history, performing diagnostic tests, and giving their opinion regarding the veracity of the claimed impairment, which is collected through the VA's disability benefit questionnaire. Importantly, the questionnaire assesses whether the preponderance of the evidence (at least 50% chance) indicates that the impairment has a known medical origin and was caused or aggravated by the veteran's military service. Claims related to disabilities not linked to military service, however severe, are denied.

All of the evidence generated by the caseworker and physician is forwarded to a Ratings Veterans Service Representative (henceforth, adjudicator), who uses the body of evidence to assign the veteran a combined disability rating. This rating determines the amount of compensation that the veteran receives. On average, the rating process takes 152 days to complete, although this time varies by the type of impairment and across regional offices.

Two features of the VA's disability rating process are noteworthy. First, veterans cannot choose their caseworker or physician. The VA's objective is to conduct assessments within a reasonable time frame, and the availability of workers at a specific time and location largely dictates the assignment process. Second, both caseworkers and physicians have considerable

[13] Although referred to as "physicians" by the VA, medical examiners can be non-physician practitioners, such as physician assistants or other licensed clinicians. Medical examiners are employed or contracted by the VA specifically to conduct medical examinations, assessing only those impairments for which they have received clinical training. For example, an audiologist would assess a tinnitus claim, and a psychologist would assess a mental health claim.

discretion in performing their roles and, as we subsequently show in Section 4.3, can influence the amount of compensation that each veteran receives. Although neither makes the final decision to approve or reject the disability claim, the evidence that they generate is a key input in the adjudicator's decision. These two institutional features set up an ideal setting for studying the impact of the VA's disability compensation program on mortality and other health outcomes.

3. Data and Sample Selection

This section describes our administrative data from the VA and our mortality data, which were aggregated across multiple VA and non-VA sources. We proceed to define the key variables used in this study. Finally, we explain the methodology behind the construction of the sample and present summary statistics for the sample.

3.1 Data sources and variables

We analyze administrative benefits data from the Veterans Benefits Administration, which contain the universe of disability claims submitted to the VA between 2000 and 2018, including the claim's submission date, regional office, impairment type (Appendix Section 4.1 presents the most prevalent impairments in 2018), service-connected disability rating, and identifiers for the claimant and caseworker. We do not observe the identity of the adjudicator in the data provided to us.

We link the benefits data to health data from the Veterans Health Administration, which contain demographic information, residential zip codes, and detailed records for all medical care furnished or reimbursed by the Veterans Health Administration.[14] The medical records

[14]The federally-run Veterans Health Administration is the nation's largest integrated healthcare system, overseeing more than 170 medical centers and 1,000 outpatient clinics and serving 9 million veterans annually. Approximately 60% of veterans, and an even higher percentage of veterans with disabilities, utilize VA health

cover the period from 2000 to 2023 and include physician identifiers, diagnostic codes (International Classification of Diseases), procedure codes (Current Procedural Terminology), and the location of inpatient and outpatient encounters (VA stop codes). We transform veterans' diagnostic codes into a set of 31 Elixhauser comorbidity indicators (Elixhauser et al. 1998), which are widely used by health services researchers to summarize disease burden in patients. In addition, because the VA provides integrated health care to veterans, we can observe the utilization of supportive housing services for veterans who are at risk of or currently experiencing homelessness. Appendix Section 4.2 presents the list of VA stop codes used to create our health care utilization variables.

The primary outcome variable is all-cause mortality, measured 1 to 10 years after the rating process.[15] The exact date of death is obtained from multiple sources: the VA Beneficiary Identification and Records Locator System, VA inpatient facilities, the VA National Cemetery Administration, the Centers for Medicare & Medicaid Services Vital Status File, and the Social Security Administration Death Master File. Veteran mortality data compiled from multiple sources have been shown to be more accurate than data from any single source (Sohn et al. 2006). We link the benefits, health, and mortality data using veterans' Social Security numbers.

We construct two indices for the veterans in our sample: medical acuity and social deprivation. The medical acuity index is derived from a logistic regression model that predicts 5-year mortality based on the 31 Elixhauser comorbidity indicators, which were measured prior to the disability rating process. The social deprivation index, which was developed by Butler et al. (2013), is a zip code-level measure derived from seven demographic characteristics from the American Community Survey, including poverty, unemployment, and lack

care.

[15]Mortality is commonly recognized as a key indicator of health, as it reflects not only the incidence of diseases but also the cumulative effects of environmental, behavioral, social, and genetic influences on health (Tulchinsky et al. 2022). Its primary strength as a health outcome is its definitive nature: death is an unambiguous outcome. This clarity allows for straightforward comparisons of mortality rates across different locations, populations, and time periods.

of access to education and essential services. Both indices, measured prior to the disability rating process, are standardized to have a mean of zero and a standard deviation of one. Higher values signify greater medical acuity or social deprivation.

3.2 Sample and summary statistics

Our analytic sample includes 812,281 veterans, selected from the universe of veterans who submitted a disability claim to the VA between 2001 and 2018 (N = 3,963,288). The study period begins in 2001 to ensure that health data covering at least 1 year before the date of claim submission are available. In cases where a veteran applied for disability compensation multiple times in the study period, we analyze the first claim. We narrow our sample to veterans whose claims were assessed by a caseworker and a physician, each with a record of assessing at least 200 claims. This is done to reduce noise in the constructed measures of their leniency, which we use as instruments. Finally, we include only veterans in an impairment type × regional office × year with at least 2 caseworkers and 2 physicians available. This final restriction sets up our natural experiment because disability claims are as good as randomly assigned to caseworkers and physicians only within an impairment, regional office, and year. For instance, a mental health claim originating from Los Angeles has zero chance of being assigned to an oncologist in New York. This sample restriction guarantees that we compare only cases in which there are comparable individuals who had the same set of caseworkers and physicians available at the time of the application, given their impairment.

Table 1.1 presents summary statistics for our sample. Healthcare utilization and comorbidity variables are measured in the year *before* the rating process. To benchmark these summary statistics, we compare our sample with two other relevant populations: nondisabled veterans who enrolled with the Veterans Health Administration between 2001 and 2018 (column 2), and recipients of SSDI (column 3).[16]

[16] To calculate statistics for nondisabled veterans who enrolled with the Veterans Health Administration,

Our sample is predominantly male (91%), reflecting the broader demographics of veterans in the general population, who are 90% male. Approximately 75% are non-Hispanic White, 16% Black, 8% Hispanic, and 4% Asian or Pacific Islander. The average age is 50.2 years, noticeably older than the post-9/11 veterans studied by Bruhn et al. (2022). Half of the veterans in our sample utilized VA health care in the year before submitting their disability claim, with 11% accessing mental health services. Among those utilizing VA health care, the most prevalent comorbidities are hypertension (35%), depression (23%), and diabetes (17%). Appendix Section 4.3 presents summary statistics for all 31 Elixhauser comorbidities.

Veterans who apply for VA disability compensation experience significantly worse health compared to nondisabled veterans. This pattern mirrors health disparities observed between disabled and nondisabled people in the civilian population (Forman-Hoffman et al. 2015). Approximately 8.0% of veterans die within five years of the disability rating process. This number rises to 17.9% within ten years. In contrast, the 5- and 10-year mortality rates for nondisabled veterans are 4.3% and 9.4%, respectively. Such disparities underscore the urgency of addressing the health and well-being of disabled veterans, who not only face greater health challenges but also increased mortality risks. Furthermore, given their complex health needs and economic vulnerabilities, veterans with disabilities represent an extremely relevant population for studying the health effect of cash transfers, offering insights into broader implications for social support systems like the VA's disability compensation program.

4. Empirical Strategy

To isolate the causal impact of VA disability compensation, we employ an IV approach that circumvents selection bias associated with nonrandom treatment take-up. We begin by

---

we construct a subsample of 58,613 veterans who had a 0% disability rating and received no income support from the VA between 2001 and 2018. The statistics on SSDI recipients come from annual reports from the Social Security Administration.

describing the potential sources of endogeneity in our setting. Next, we develop two instruments based on the relative leniency of quasi-randomly assigned caseworkers and physicians, and conclude by presenting exercises testing the validity of these instruments.

4.1 Potential sources of endogeneity

Veterans receiving cash transfers from the VA's disability compensation program differ from nonrecipients in numerous observable and unobservable ways, potentially creating selection bias in correlational estimates. The disability program's design, which awards larger cash transfers to veterans with more severe disabilities, might give the impression that transfers harm health because individuals with severe disabilities generally have worse health outcomes in the absence of compensating transfers. Thus, cross-sectional comparisons between veterans with varying transfer amounts could understate the causal impact of cash transfers on health. In addition, estimates based on comparisons over time can be biased because disability application patterns fluctuate nonrandomly. For instance, SSDI applications tend to increase during economic downturns, likely driven by workers seeking disability compensation as a buffer against income losses and uncertainty (Black et al. 2002; Maestas et al. 2021). If disability applications correlate positively with poor economic conditions, then ordinary least squares (OLS) estimates will be biased upward in the short term because downturns are associated with decreased population-level mortality risk, due to reductions in accidents and pollution (Ruhm 2000; Cutler et al. 2016; Finkelstein et al. 2023). However, these estimates might be biased downward in the long term, capturing the detrimental effect of economic recessions over a longer time horizon (Schwandt and von Wachter 2020).

Indeed, we observe that veterans in poorer health tend to receive more generous cash transfers, consistent with the VA's objective of awarding greater compensation to veterans with more severe disabilities. Appendix Section 4.4 presents summary statistics for the veterans in our sample stratified by their disability rating (above vs. below 50%). In the year prior to the rating process, veterans subsequently assigned a disability rating $\geq$ 50% had a higher incidence of hypertension (36.0 vs. 33.6%) and depression (30.0 vs. 17.0%), conditions linked to common causes of death among veterans (Maynard et al. 2018). These veterans were also more likely to utilize inpatient services (5.2 vs. 3.0%) and the emergency department (4.3 vs. 3.2%). Indeed, if the impairments that qualify a veteran for VA disability compensation also independently increase their risk of mortality, then OLS estimates will be biased toward finding a null or negative effect of cash transfers.

4.2 Caseworker and physician instruments

Our empirical strategy leverages the quasi-random assignment of disability claims to both caseworkers and physicians, as well as the influence that each agent has on the cash transfers that veterans receive. Our main specification is:

$$Y_i = \beta D_i + \mu S_i + \epsilon_i \tag{1.1}$$

$$D_i = \alpha^c Z_i^c + \alpha^p Z_i^p + \gamma S_i + \nu_i \tag{1.2}$$

The outcome for individual i is Y_i, which depends on the multivalued treatment $D_i \in \{d_1, \ldots, d_k\}$. In our setting, the treatment is veteran i's disability compensation measured in 2018 dollars. Veteran i is assessed by caseworker c and physician p. The vector S_i contains dummies for the conditioning set (impairment type $\times$ regional office $\times$ year) within which the assignment of claims to caseworkers and physicians is plausibly unrelated to other determinants of Y_i; there are 34,160 conditioning sets in our sample.

The main coefficient of interest is β, or the causal effect of D_i on Y_i. As Section 4.1 discusses, the OLS coefficient estimated by equation (1.1) alone may suffer from selection bias. To isolate the causal effect, we instrument D_i using caseworkers' and physicians' relative leniency. The caseworker instrument is Z^c, defined as the leave-out mean of D calculated using every claim assigned to c other than i's. The leave-out mean is a standard approach for measuring leniency in the "judges design" literature because it prevents overfitting the first-stage equation, which would otherwise bias our result toward OLS.[17]

$$Z_i^c \equiv \frac{1}{|I_c|-1} \sum_{i' \neq i} \mathbb{1}(i' \in I_c) D_i \tag{1.3}$$

We construct the physician instrument Z^p in the same manner:

$$Z_i^p \equiv \frac{1}{|I_p|-1} \sum_{i' \neq i} \mathbb{1}(i' \in I_p) D_i \tag{1.4}$$

where I_c and I_p are the sets of veterans assigned to caseworker c and physician p, respectively.

We estimate our IV model using two-stage least squares (2SLS). Since D_i is a positive ordered treatment, under the assumption that equation (1.2) is a good approximation of the conditional mean, β^{2SLS} can be interpreted as an average causal response (ACR) of Y_i with respect to shifts in D_i (Blandhol et al. 2022; Garin et al. 2023). For exposition of the ACR, suppose that F_Z is the CDF for an arbitrary instrument Z with support $[\underline{z}, \bar{z}]$. Define:

$$\beta_z \equiv \frac{\overbrace{E[Y(D(z))|Z=z] - E[Y(D(\underline{z}))|Z=\underline{z}]}^{\text{reduced form}}}{\underbrace{E[D|Z=z] - E[D|Z=\underline{z}]}_{\text{first stage}}} \tag{1.5}$$

which is the Wald estimator associated with instrument value z relative to the lower bound of $\underline{z}$. Angrist and Imbens (1995) show that the 2SLS estimate can be written as a weighted average of Wald estimates with weights that depend on how $\underline{z} \to z$ shifts individuals across

[17] An alternative approach is to treat caseworker and physician indicators as instruments. However, IV estimators with many weak instruments are biased toward the OLS estimator (Bound et al. 1995) and are inconsistent when the number of instruments grows in proportion to the sample size (Bekker 1994). Consequently, most studies employing a "judges design" opt for the leave-out mean as their instrument. Simulations indicate that leave-out mean estimators and their standard errors are reliable when the number of cases per "judge" is large (Bhuller et al. 2020).

different levels of D:

$$\beta^{2SLS} = \int_{\underline{z}}^{\bar{z}} \omega(z)\beta_z dF_Z(z) \tag{1.6}$$

where $\omega(z) = \frac{z(E[D|Z=z]-E[D|Z=\underline{z}])}{\int z(E[D|Z=z]-E[D|Z=\underline{z}])dF_Z(z)}$ are weights summing to one.[18] Notice in the numerator of $\omega(z)$ that only individuals whose treatment status is moved by variations in the instrument contribute to the 2SLS estimate; these individuals are compliers. The causal effect for compliers is the local average treatment effect (LATE).

Equation (1.2) relates the endogenous treatment D to the caseworker and physician leave-out mean instruments, and we also present results from an alternate specification that includes an interaction term between both instruments. In addition, since veterans can apply for multiple disabilities concurrently, we conduct the analysis using claim-level data and reweight the estimates such that all veterans are equally weighted. Finally, standard errors are clustered at the caseworker and physician level.

4.3 Instrument validity

This section evaluates the assumptions needed for instrument validity (Angrist and Imbens 1995). The first identifying assumption is relevance: instruments Z^c and Z^p must be strongly correlated with the endogenous treatment D. Figure 1.1 plots the variation in the caseworker and physician leave-out mean instruments. Panel A shows the histogram of the residualized (on impairment type $\times$ regional office $\times$ year) variation in Z^c, and Panel B shows the histogram of the residualized variation in Z^p. For both caseworkers and physicians, there are meaningful variations in their leniency, with veterans assigned to more lenient agents ultimately receiving more generous transfers. The solid black line is the fractional polynomial

[18]This ACR result can be generalized to cases with multiple instruments and models with covariates. In such cases, the weights are also a function of the conditional (on X) variance of Z and the distance between the Wald estimate and the center of the distribution of Z.

regression of the residualized treatment variable on the residualized instrument, which shows a first-stage relationship that is very predictive, monotonic, and close to linear for both instruments. Based on the VA's 2018 payment schedule, being assigned to a caseworker in the 95^{th} percentile of the leave-out mean distribution increases transfers by \$240 per month more than being assigned to one in the 5^{th} percentile. Being assigned to a physician in the 95^{th} percentile increases cash transfers by \$230 more per month than being assigned to one in the 5^{th} percentile. A joint F-statistic of 664 ($p < 0.001$), which is much larger than the rule of thumb of 100 (Lee et al. 2022), indicates that the instruments are not weak.

The second identifying assumption is conditional independence: the potential outcomes for Y must be independent of Z^c and Z^p, conditional on impairment type $\times$ regional office $\times$ year. This assumption is implied by the quasi-random assignment of disability claims to caseworkers and physicians, and has been previously vetted by Silver and Zhang (2023). We further empirically evaluate this assumption by testing the correlation between Z^c and Z^p and veteran characteristics that are themselves correlated with D. To do this, we estimate predicted disability compensation $\hat{D}$ using a vector of veteran demographics (age, sex, race and ethnicity), their prior year's utilization of outpatient and inpatient care, 31 Elixhauser comorbidity indicators, indicators for Agent Orange or radiation exposure, and the social deprivation index based on residential zip code. $\hat{D}$ is strongly correlated with 5-year mortality ($F = 7268$; $p < 0.001$), making it a compelling test for instrument independence. In Figure 1.1, the dashed line is the fractional polynomial regression of residualized $\hat{D}$ on residualized Z^c (Panel A) and on residualized Z^p (Panel B), which shows no relationship between $\hat{D}$ and either instrument. The t-statistics for their OLS coefficients are -1.2 ($p = 0.24$) and 0.6 ($p = 0.53$), respectively.

Although the conditional independence assumption is sufficient for a causal interpretation of the reduced-form effect of being assigned to a more lenient caseworker or physician, interpreting the 2SLS estimates as the causal effect of cash transfers requires the exclusion restriction to be satisfied. This assumption states that Z^c and Z^p affect Y only by shifting i's disability compensation. The defense of the exclusion restriction for Z^c is straightforward. Caseworkers interact with veterans for the sole purpose of gathering evidence that will be used in the disability rating process. One potential violation of the exclusion restriction for Z^p is that lenient physicians may be more likely to treat veterans during or after their visit or somehow influence their subsequent treatment decisions. However, the chance of this violation occurring is minimal, as VA policy requires the visit to be purely investigative: the physician is prohibited from treating the veteran during or after the formal examination, which rarely lasts longer than two hours.

In addition to cash compensation, veterans with service-connected disabilities are eligible for nonmonetary benefits such as vocational rehabilitation, adaptive housing grants, and automobile grants (Appendix Section 4.5). The take-up of nonmonetary benefits is low, however. While all veterans with a service-connected disability rating $\geq 10\%$ receive the cash transfer, only 2.5% take advantage of vocational rehabilitation, and fewer than 1% receive the adaptive housing or automobile grant. Therefore, we maintain that the primary effect we are estimating comes from increased cash transfers, although the precise causal interpretation pertains to the bundled treatment.

If the causal effect β is constant across individuals, then the instruments need to satisfy only the relevance, conditional independence, and exclusion assumptions for 2SLS to recover the average treatment effect (ATE). However, if the treatment effects are heterogeneous, then 2SLS with multiple instruments can be interpreted as a positively weighted average of local average treatment effects (LATE), insofar as partial monotonicity is satisfied (Mogstad et al. 2021): $D_i(z_c^{'}, z_p) \geq D_i(z_c, z_p)$ and $D_i(z_c, z_p^{'}) \geq D_i(z_c, z_p)$ $\forall i$. Partial monotonicity requires that, with either instrument held fixed, all individuals are weakly induced to greater treatment by increasing the other instrument.[19] To assess the partial monotonicity assump-

[19]By contrast, the standard monotonicty condition formalized by Imbens and Angrist (1994) requires that if $z^{'} > z$, then $D(z^{'}) \geq D(z)$ (or vice versa) for all individuals. Heckman and Vytlacil (2005) observe that

tion, we adopt the standard practice in the "judges design" literature of testing whether the first-stage relationship between D and Z remains positive for subgroups defined by different observable characteristics (Arnold et al. 2018; Bhuller et al. 2020). In all subgroups stratified by age (18–34, 35–49, 50–64, 65+), race (White, Black), and sex (male, female), we reject at conventional significance levels the hypothesis that the first-stage coefficient is negative for either instrument (Appendix Section 4.6).

5. Causal Effect of Disability Compensation

This section presents our main results. First, we report OLS and 2SLS estimates for mortality measured 1 to 10 years following the disability rating process. Since our administrative dataset go until 2018 but our health and death records go until 2023, we observe mortality for up to 5 years after the rating process for the entire sample.[20] For healthcare utilization and morbidity variables, we report estimates for outcomes measured 5 years after the rating process.

5.1 Mortality

We begin by estimating the impact of cash transfers on all-cause mortality. Unless stated otherwise, the OLS and 2SLS coefficients that we report are scaled to represent the effect of receiving \$300 per month, which, based on the VA's 2018 payment schedule (Appendix Section 4.1), is commensurate with a 10% disability rating increase for the median claimant. The outcome variables are binary variables indicating whether the veteran was deceased 1 to 10 years after the rating process. The primary outcome is 5-year mortality. Figure 1.2

standard monotonicity cannot be satisfied with multiple instruments without restricting choice behavior to be homogeneous. Under a random-utility framework, this restriction does not allow individuals to differ in their behavioral responses to various instruments.

[20]For years 6 through 10, the mortality outcomes are analyzed in subsamples with a sufficient time horizon for us to measure mortality within that timeframe. For example, our analysis of 10-year mortality excludes veterans whose claims were processed after 2012.

plots the OLS and 2SLS coefficients for 1- to 10-year mortality and their 95% confidence intervals.

The OLS coefficient for 5-year mortality is 0.75 pp (s.e. = 0.02), indicating that veterans receiving larger cash transfers are more likely to die within 5 years of the rating process than veterans receiving smaller transfers. The OLS coefficients increase linearly over time (1.4 pp by year 10), implying a constant hazard rate for mortality at this margin. However, as discussed in Section 4.1, the OLS estimates likely suffer from selection bias because the types of impairments that qualify a veteran for more generous transfers independently increase their risk of mortality.

In contrast, the 2SLS coefficient for 5-year mortality is negative and indicates that receiving $300 per month reduces mortality by 0.76 pp (s.e. = 0.19). This causal estimate is meaningful, amounting to nearly one-tenth of the mean 5-year mortality rate in our sample. Comparing the OLS and 2SLS estimates suggests that the direction of the selection bias is positive, with higher-rated veterans facing an elevated risk of premature mortality, holding constant the compensatory effect of their cash transfers.

As Figure 1.2 illustrates, the protective effect of cash transfers rises over time, reaching its peak in year 6 before starting to descend. By year 9, the causal effect is statistically indistinguishable from zero. This pattern is logical if we follow the intuition and predictions from models of the relationship between health and mortality (Grossman 1972; Lleras-Muney and Moreau 2022). The protective effect grows initially because greater income enables veterans to accumulate health capital but eventually shrinks because all individuals die at some future point regardless of how much money they receive.

We evaluate the robustness of our main results to changes to the model's specification. First, we assess the stability of our estimates as we add covariates to the model, as suggested by Altonji et al. (2005). Second, we present the results after we control for more granular conditioning sets (impairment type × regional office × year) with quarter–year fixed effects and narrower impairment categories. Finally, we compare the 2SLS estimate against alternative IV estimators with different statistical properties, such as limited information maximum likelihood (LIML) and generalized method of moments (GMM).[21] The results from these robustness exercises are similar in magnitude and significance to our main results.

5.2 Healthcare utilization

One way that cash transfers can affect mortality is through healthcare utilization. Since health care is a normal good, cash transfers might increase the use of preventive care and timely medical interventions, both of which lower the risk of premature death. Conversely, if a veteran's health improves because of the transfer, the need for medical treatments could decrease, leading to reduced healthcare utilization. Therefore, the net effect of cash transfers on healthcare utilization is ambiguous and an empirical question.

Table 1.2 presents the 2SLS coefficients for an array of healthcare utilization outcomes. In the 5 years following the disability rating process, 90% of the veterans in our sample had at least 1 outpatient visit at a VA clinic. Preventive care (defined as primary care and mental health) was the most frequently accessed service. On average, veterans had a total of 15 primary care visits and 13 mental health visits over the 5-year period. The 2SLS coefficient for primary care utilization is 0.92 visits (s.e. = 0.17). For mental healthcare utilization, it is 2.5 visits (s.e. = 0.27). Receiving $300 per month significantly boosts veterans' utilization of preventive care, with mental health services driving most of this increase.

Copays for preventive care at VA clinics are extremely low or zero for most veterans, so it is unlikely that improved affordability fully accounts for the increase in preventive care utilization. One possible explanation for our finding is that cash transfers lead to a reduction

[21] These IV estimators have both advantages and drawbacks compared to 2SLS, depending on specific conditions. LIML is generally more robust than 2SLS in the presence of weak instruments, when the correlation between the instruments and the endogenous variable is low. GMM tends to be more precise when the IV model is overidentified (i.e., there are more instruments than endogenous regressors). In such scenarios, GMM offers a framework for using the available instruments more efficiently.

in veterans' labor supply, freeing up time to seek preventive care. Indeed, individuals who are busy with work are more likely to skip preventive visits, such as cancer screenings and dental check-ups (Yao et al. 2015). While we cannot observe labor force participation in our data, Autor et al. (2016) use Social Security records to show that a 2001 policy expanding eligibility for VA disability compensation significantly depressed veterans' labor force participation.

Cash transfers have little to no effect on veterans' utilization of inpatient or emergency department services. The 2SLS coefficients for these utilization measures are indistinguishable from zero, in contrast to the estimates for preventive care. This null effect for emergency department and inpatient utilization could be due to improved health decreasing the demand for critical, end-of-line medical services, which also tends to be price inelastic (Ellis et al. 2017).

The Veterans Affairs Supportive Housing (VASH) program offers unhoused veterans and their families housing vouchers combined with comprehensive supportive services. An estimated 5.0% of veterans in our sample had at least 1 VASH encounter in the 5 years after the rating process, totaling 0.9 VASH encounters per veteran. The 2SLS coefficient for VASH utilization is -0.21 encounters (s.e. = 0.07), indicating that cash transfers reduce recipients' reliance on homelessness services. This result is encouraging, as the VA's disability compensation program is not explicitly a homelessness program, and confirms prior research that has identified economic vulnerability as a determinant of homelessness for veterans and the broader population (Benjaminsen and Andrade 2015; Meyer et al. 2021; Evans et al. 2019).

5.3 Morbidity

We use medical diagnoses to proxy for cause of death, as this information is not available in the vital records available to us. Specifically, for the subset of veterans with at least 1 VA encounter within 5 years of the rating process (N = 733,250 veterans), we analyze whether they were diagnosed with a cardiovascular disease (hypertension, cardiac arrhythmia, or congestive heart failure), diabetes, or a mental health disorder (psychosis or depression). Table 1.3 presents mean incidence rates (per 100 veterans) and the corresponding 2SLS coefficients for these morbidity outcomes.

We find a modest reduction in both cardiovascular disease and diabetes attributed to cash transfers. Specifically, receiving $300 per month decreases the incidence of cardiovascular disease by 1.9 pp (s.e. = 0.59) and diabetes by 1.0 pp (s.e. = 0.38). Metabolic conditions such as cardiovascular disease and diabetes are among the leading causes of death for older men, who comprise most of our sample. Our finding suggests that improved metabolic health may be a key mediator in the protective effect of cash transfers.

Cash transfers do not appear to improve mental health and, in fact, increase the likelihood that a veteran is diagnosed with a mental health condition. The 2SLS coefficient of 5.6 pp (s.e. = 0.44) indicates that receiving $300 per month significantly raises the incidence of psychosis or depression. This result could reflect a true decline in mental health, a result that would contradict studies finding a beneficial effect on mental health in other populations (McGuire et al. 2022; Zimmerman et al. 2021; Lindqvist et al. 2020). On the other hand, the positive coefficient associated with psychosis and depression could result from enhanced detection of previously undiagnosed conditions due to the increased utilization of preventive care that we showed in Section 5.2. There is strong correlational evidence to support this interpretation: veterans with at least 5 preventive care visits are 6.7 more likely to be diagnosed with psychosis or depression than veterans with fewer than 5 preventive care visits. Given the nuances of how mental health conditions are diagnosed in healthcare settings and the observed positive relationship between cash transfers and healthcare utilization, we cannot definitively determine whether cash transfers benefit or harm mental health.

6. Treatment Effect Heterogeneity

This section explores whether veterans differentially benefit from the same cash transfer. We begin by analyzing heterogeneity in the health effects of cash transfers by impairment type, focusing on mental health and musculoskeletal conditions. Then, we analyze heterogeneity by commuting zone and correlate commuting zone–specific causal estimates with regional characteristics.

6.1 Heterogeneity by impairment type

The VA's disability compensation program provides financial support for veterans with a wide range of impairments resulting from their military service, including 833 that are currently eligible for compensation. The most frequently claimed impairments are tinnitus, limitation of flexion, hearing loss, post-traumatic stress disorder, and lumbosacral strain (Appendix Section 4.1). Over the past two decades, there has been a dramatic rise in musculoskeletal and mental health impairments in the program's rolls. The shift in the impairment composition of beneficiaries has attracted scrutiny, as musculoskeletal and mental health impairments are generally harder to verify and beneficiaries with these conditions tend to have higher survival rates. Consequently, there are questions about whether cash transfers actually benefit these recipients and whether this trend indicates ineffective targeting of transfers in the disability rating process (Autor and Duggan 2006; Meseguer 2021). Given these concerns, we investigate whether cash transfers improve the health of individuals with musculoskeletal and mental health impairments.

Table 1.4 reports first-stage and 2SLS coefficients for 5-year mortality stratified by impairment type: musculoskeletal, auditory, mental health, neurologic, cardiovascular, dermatologic, and respiratory (ordered by frequency). We adjust impairment-specific 2SLS estimates for sampling error using a standard empirical Bayes procedure, following the approach of

Chandra et al. (2016) and Finkelstein et al. (2021). The intuition behind this adjustment is that impairments for which the 2SLS coefficient is above (below) average are more likely to suffer from positive (negative) sampling error; this procedure brings their estimates closer to the empirical mean. Appendix Section 1.5 describes how we implement the empirical Bayes procedure.

Receiving $300 per month reduces 5-year mortality for veterans with mental health impairments by 1.1 pp (s.e. = 0.23) and for veterans with auditory impairments by 2.9 pp (s.e. = 0.24). While mental health and auditory conditions may seem unrelated, hearing loss is a known risk factors for psychological distress (Blazer and Tucci 2019; Geocze et al. 2013), and individuals with tinnitus have a heightened risk of suicide (Lewis et al. 1994). Cash transfers may be especially protective for this group because mental health and auditory impairments stemming from military service are severe due to exposure to heavy artillery and warzone deployment. Moreover, these conditions may exhibit more plasticity, and as we show in Section 5.2, cash transfers boost utilization of mental health services, which could serve as a pathway through which transfers confer health benefits on those with mental health impairments.

Veterans with a musculoskeletal impairment do not experience any improvement in mortality from cash transfers. For this group, the 2SLS coefficient for 5-year mortality is precisely zero. Why might cash transfers be less effective for veterans with impairments of this kind? Musculoskeletal conditions, while causing pain and limitations in mobility, are rarely life-threatening on their own and are difficult to verify. In our sample, the 5-year mortality rate for this group is 6.4%, which is the lowest among all of the impairments we analyze. In addition, previous studies of SSDI have shown that claimants with musculoskeletal conditions are more likely to find employment if their claims are rejected (Von Wachter et al. 2011), which potentially blunts the negative impacts of a benefit denial.

Notably, for veterans with hemic and infectious impairments, which are among the most acute impairments, with mean 5-year mortality rates of 19.3% and 17.5%, respectively, cash transfers do not appear to improve health. Given the relatively small sample size of these groups, their standard errors are wide. Nonetheless, this finding is part of a broader pattern across all of the impairments that we analyze of little to no correlation between the severity of an impairment and the health effect of transfers for that impairment.

6.2 Geographic heterogeneity

Commuting zones (CZs) in the U.S. vary significantly in their economic and social conditions. Cash transfers may interact with these conditions, leading to variations in their effect on health. For instance, regions with greater healthcare quality could enable individuals to more effectively convert cash into beneficial health outcomes. To investigate place-based heterogeneity in treatment effects, we calculate 2SLS coefficients for 5-year mortality for each of the 132 CZs with at least 1,000 claimants in our sample. As in our analysis of impairment-specific estimates, we adjust the CZ-specific estimates using a standard empirical Bayes procedure (Appendix Section 1.5).

Figure 1.3 plots the 2SLS coefficients by CZ, with the most pronounced protective effects situated on the left side. Dark gray shading indicates that the 2SLS coefficient is statistically significant at the 5% level, demonstrating significant place-based variation in health outcomes related to transfers. Receiving \$300 per month decreases 5-year mortality by ≥ 1 pp in 19 CZs. While there is no significant effect for the majority of CZs, transfers appear to be detrimental in three areas: Boston, Charlotte, and Pittsburgh.[22] The cross-CZ standard deviation is 0.76 pp, highlighting potentially important interactions between location and the health impacts of cash transfers.

We investigate the potential drivers of geographic heterogeneity by correlating the CZ-

[22] However, we caution against placing too much emphasis on any single estimate given the relatively small sample sizes for individual CZs.

specific treatment effects with regional characteristics. Our analysis focuses on the geographic determinants of health and mortality from Finkelstein et al. (2021) and Chetty et al. (2016), which include life expectancy, healthcare quality (30-day hospital mortality rate, Medicare spending per enrollee, uninsured rate), health behaviors (rates of smoking, obesity, and exercise), socioeconomic conditions (income segregation, racial segregation, poverty rate, household income, and share in manufacturing), and racial demographics. Given the volume of CZ-specific observables, we discipline our investigation by using LASSO regression to select the variables most predictive of cash transfers' health effect. LASSO allows us to remain neutral as to the sources of heterogeneity and uncover patterns in the data by searching over a high-dimensional set of characteristics. Finally, we note that the correlations from this exercise offer suggestive evidence and need not represent causal relationships.

The LASSO regression identifies only one characteristic as predictive of CZ-level treatment effects: the proportion of workers in the manufacturing sector. In CZs where the manufacturing sector labor force exceeds 20%, cash transfers no longer significantly reduce 5-year mortality rates. While this result is not causal in nature, that cash transfers have minimal health effect in regions with a significant manufacturing presence connects to research on "deaths of despair" in the rust belt and parts of Appalachia that suffered disproportionately from trade liberalization, technological changes, and increased competition from Chinese manufacturers (Case and Deaton 2015; Pierce and Schott 2020).[23] Workers in these CZs have seen diminished earnings, but our findings suggest that, in terms of health risk, cash transfers do not compensate for this decline. Notably, opioid overdoses were a significant driver of mortality in these disadvantaged areas during our study period. However, using CZ-level data on opioid supply from Arteaga and Barone (2022), we find no correlation between opioid availability and CZ-specific treatment effects. The diminished health effect

[23]For example, manufacturing industries that were more exposed to trade liberalization with China in 2001 exhibited large declines in employment and uptake of welfare programs (Autor et al. 2013). Case and Deaton (2017) argue that social and economic deterioration associated with trade liberalization contributed to a rise in "deaths of despair" in many of these regions.

of cash transfers seems not to be driven by opioid addiction but rather is perhaps driven by the lack of opportunities available to individuals in these areas.

7. Model-Based Interpretation of Causal Estimands

This section provides a conceptual framework to interpret the causal estimands reported in the previous sections. We begin by presenting instrument-specific treatment effects and reconcile their differences by analyzing compliers. Then, we explore selection into treatment using a marginal treatment effects framework. We conclude by discussing patterns of treatment selection.

7.1 Instrument-specific treatment effects

Up to this point, the treatment effects that we have reported were derived from an overidentified IV model combining caseworker and physician leave-out mean instruments. The textbook advantage of combining multiple instruments is statistical efficiency, but this advantage comes at the expense of interpretability when treatment effects are heterogeneous. Under partial monotonicity (defined in Section 4.3), the overidentified 2SLS model identifies a positively weighted average of causal effects from each instrument (Mogstad et al. 2021). In this section, we parse the overidentified 2SLS estimand into its constituent caseworker and physician parts.

Figure 1.4 reports 2SLS coefficients for 1- to 10-year mortality, calculated separately using the caseworker instrument and the physician instrument. The caseworker-specific 2SLS estimate indicates that receiving $300 per month reduces 5-year mortality by 1.02 pp (s.e. = 0.22). In contrast, the physician-specific 2SLS estimate indicates that the same transfer reduces 5-year mortality by only 0.42 pp (s.e. = 0.19). This discrepancy is statistically significant (overidentified J-statistic $= 10.6$; p-value $= 0.001$).[24] That the overidentified 2SLS coefficient lies between the caseworker- and physician-specific 2SLS coefficients is consistent with the fact that the overidentified 2SLS estimate is a positively weighted mean of individual effects from each instrument.

It is noteworthy that the caseworker instrument and physician instrument yield distinct treatment effects, as this discrepancy mimics the wide range of causal estimates observed in previous studies on the effect of income on health. While the wide range of estimates in previous studies might reflect differences in study methodologies, our analysis using two parallel identification strategies that differ only in the agent influencing the treatment decision produces two markedly distinct estimates. We also use the same sample for both calculations, which lets us rule out compositional differences as the cause of this discrepancy.[25]

7.2 Compliers analysis

Why do our results differ depending on the instrument used? In equation (1.6), we express the 2SLS estimand as a weighted mean of individual causal effects, where the weights depend on how much each instrument shifts veterans across different treatment intensities. These weights are nonzero for compliers: $z \neq z' \implies D_i(z) \neq D_i(z')$. It follows that a gap between instrument-specific LATEs might emerge if (i) complier subgroups are disjointed or only partially overlap and (ii) treatment effects are heterogeneous between these complier subgroups. We find evidence to support both claims.

Understanding compliers in our institutional setting helps to contextualize our 2SLS estimates, particularly as they relate to the disability rating process. In our setting, caseworker compliers are veterans who receive more (less) generous cash transfers when their disabil-

[24]The textbook motivation for the Sargan–Hansen J test is to test the validity of overidentifying restrictions in IV models under *constant* treatment effects (Sargan 1958; Hansen 1982).

[25]We also rule out the possibility that the physician instrument is weak, which would bias its 2SLS estimate toward OLS (Bound et al. 1995). In Section 4.3, we showed that the caseworker instrument and physician instrument had nearly equally strong first-stage relationships; their F-statistics are 803 and 821.

ity claim is assigned to a relatively lenient (stringent) caseworker. Physician compliers are likewise defined in relation to physicians.[26] Based on this intuition, disability claims for compliers may be more challenging to adjudicate, causing caseworkers' and physicians' to disagree and their relative leniency to have greater weight in the disability rating process. Finally, we emphasize that a caseworker complier is not necessarily a physician complier and vice versa.[27]

We examine differences between caseworker and physician compliers using the κ theorem from Abadie (2003).[28] For tractability and ease of interpretation, we dichotomize the treatment variable: $\tilde{D}_i \equiv \mathbb{1}\{D_i \geq 1000\}$.[29] Although our compliers analysis could theoretically be extended to the 11-valued ordinal treatment in equation (1.2), estimates of characteristics for 110 ($= 20 + 18 + 16 + \ldots + 2$) types of compliers would be imprecise and difficult to interpret. We estimate that 30% of our sample are caseworker compliers and 19% are physician compliers, noting that these values likely underestimate the true complier shares since they are calculated with the dichotomized treatment.

Table 1.5 presents the characteristics of caseworker compliers and physician compliers. Compared to the overall sample, caseworker compliers and physician compliers are more likely to be under age 50, Black, and Hispanic. Notably, there is a significant difference

[26]Conversely, always-takers receive large transfers, and never-takers receive small transfers, irrespective of their assigned caseworker or physician. Consider the following scenarios. A veteran with two limb amputations that were unambiguously caused by military service will receive the maximum transfer regardless of the leniency of her assessor, which leads to her classification as an always-taker. A veteran who is unambiguously able-bodied receives no transfer regardless of the leniency of his assessor, which leads to his classification as a never-taker.

[27]Mogstad et al. (2021) formalize the concept of nonoverlapping complier subgroups by introducing four types of compliers in a setting with one binary treatment and two binary instruments. Eager compliers participate in treatment if either instrument activates; reluctant compliers only participate if both instruments activate; Z_1 compliers participate if and only if Z_1 activates; Z_2 compliers participate if and only if Z_2 activates. The 2SLS estimand can be expressed as a weighted combination of LATEs corresponding to the four complier subgroups.

[28]We implement the kappa method by running 2SLS where the outcome variable in the reduced-form equation is replaced with $X_i D_i$.

[29]To dichotomize the treatment variable, we use a cutoff value of $1,000, which corresponds to the cash transfer for the median claimant. In the appendix, we show that the 2SLS estimate for the ordered treatment and for the dichotomized treatment are comparable after we adjust for the mechanical difference in the size of the cash transfers.

in medical acuity between caseworker and physician compliers. Caseworker compliers have 0.05 s.d. lower medically acuity than physician compliers and 0.10 s.d. lower acuity than the overall sample. Caseworker and physician compliers reside in zip codes with no more or less social deprivation than the average across the overall sample.

7.3 Marginal treatment effects

In this section, we utilize the marginal treatment effect (MTE) framework (Björklund and Moffitt 1987; Heckman 1997; Heckman and Vytlacil 2005; Mogstad et al. 2018a) to explore patterns of treatment effect heterogeneity with respect to unobserved resistance to treatment. The central advantage of the MTE is that it can be aggregated in various ways to recover meaningful treatment-effect parameters, such as the LATE associated with the 2SLS estimate. For reasons explained in Section 7.2, we dichotomize the treatment variable for our MTE analysis.[30]

We begin by modeling potential outcomes $Y_i(0)$ and $Y_i(1)$ as linear projections on the conditioning set S_i, resembling the standard 2SLS specification in equation (1):

$$Y_i(d) = \mu S_i + U_{di}, \qquad d = 0, 1 \tag{1.7}$$

where d is the potential value of $\tilde{D}_i$. Selection into treatment $\tilde{D}_i$ is determined by the following latent index model:

[30] The MTE framework has been extended to settings with multivalued treatments when enough continuous instruments are available (Lee and Salanié 2018). However, estimating and interpreting multidimensional MTE models becomes exponentially more complex as the number of treatments increases due to the curse of dimensionality. Without dichotomization of the treatment, 11 distinct values of D_i and 2 instruments generates 110 MTEs.

$$\begin{aligned}\tilde{D}_i &= \mathbb{1}\{\beta_k Z_i^j \geq V_i^j\} \\ &= \mathbb{1}\{F(\beta_k Z_i^j) \geq F(V_i^j)\} \\ &\equiv \mathbb{1}\{P(Z_i^j) \geq U_i^j\}, \qquad\qquad j = c, p\end{aligned} \tag{1.8}$$

where Z_{ic} and Z_{ip} are the caseworker and physician leave-out mean instruments. Equation (1.8) models selection into treatment separately for caseworkers and physicians because the error term $V_{ij} \sim F$ represents the unobserved resistance to treatment (i.e., utility cost of treatment participation) and is specific to the instrument at hand. Taken together, veteran i receives treatment $\tilde{D}_i$ when the treatment propensity score $P(Z_j)$ exceeds her unobserved resistance to treatment U_j.

The MTE traces out the treatment effect for a veteran who is in percentile u^j of the V^j distribution. We estimate the MTE using the local IV estimator (Heckman et al. 2018; Carneiro et al. 2011). We implement this approach by estimating the propensity score p^j in equation (1.8) using a probit model and then estimating the outcome equation that flexibly includes p^j. The MTE is the derivative of the conditional expectation of Y with respect to the propensity score (Heckman and Vytlacil 2007).

$$\begin{aligned}\text{MTE}(U^j = u^j) &\equiv \mathbb{E}[Y_1 - Y_0 | U^j = u^j] \\ &= \frac{\partial \mathbb{E}[Y | P(Z^j) = p^j]}{\partial p^j}, \qquad j = c, p\end{aligned} \tag{1.9}$$

Our main specification assumes a joint-normal distribution for Y and p^j, but we find similar results when we use a semiparametric specification. Our estimation of the MTE closely follows the approach taken by Cornelissen et al. (2018).

Figure 1.5 presents the MTE curves corresponding to the caseworker instrument (Panel A) and physician instrument (Panel B). We caution that the propensity score distributions for p^c and p^p do not have full support (Appendix Section 4.3), so we primarily focus on percentiles u^j between 0.2 and 0.6, for which there is nonzero density and inference does not rely on overly strong parametric assumptions. Since higher values of U^c and U^p signify lower treatment propensities, a downward-sloping MTE curve indicates reverse selection on gains. Specifically, cash transfers have a minimal or even counterproductive effect on reducing mortality among the veterans most likely to receive large transfers.

We detect reverse selection on gains for both the caseworker and physician instruments, but the negative slope of the MTE curve is more pronounced for physicians. To illustrate, large cash transfers appear to harm health for the veterans most likely to be awarded them by physicians ($u^p < 0.3$), although this effect is only marginally significant. Conversely, large cash transfers have a precise null effect for the veterans most likely to receive transfers on the evaluation of caseworkers ($u^c < 0.3$).

We integrate over the MTE curves to compute estimates for other treatment-effect parameters, such as the ATE and policy relevant treatment effect (PRTE).[31] Because both caseworkers and physicians exhibit reverse selection on gains, the treatment effects on the treated are considerably smaller than the LATEs. We reiterate that the propensity score distributions for p^c and p^p do not have full support, so identification in the tails of the MTE curve relies on strong parametric assumptions. The main purpose of our MTE estimation is to distinguish between caseworker- and physician-specific LATEs, but for readers who are interested in parameters identified at the tails of the MTE, we also present results from a secondary analysis that utilizes both instruments and their interaction to expand the support of the overall propensity score distribution. In every specification, we can reject the existence of positive selection on gains, and in most cases, we find reverse selection on gains for caseworkers and physicians.

[31] The PRTE measures the average causal effect of transitioning from the status quo to a counterfactual policy that alters individuals' propensity for treatment without changing the returns to the treatment. For instance, the PRTE might estimate the effect of a policy that makes caseworkers or physicians more lenient in the rating process, which would induce the marginal claimant to receive greater transfers.

7.4 Interpreting reverse selection on gains

Given our observation that veterans with the least resistance to treatment minimally benefit from cash transfers, we explore how positive selection on levels might lead to reverse selection on gains. We analyze positive selection on levels through two exercises. First, we estimate untreated mortality rates $Y(0)$ across different levels of unobserved resistance. To do this, we estimate the MTE in equation (1.9) but substitute the outcome variable with $Y_i \times (\tilde{D}_i - 1)$. This approach is analagous to the control function estimator proposed by Brinch et al. (2017). Table 1.6 reports mean values of the untreated 5-year mortality rate for different percentiles of unobserved resistance: $\mathbb{E}[Y(0)|U_j = u_j]$. We calculate values separately for the caseworker instrument (column (1)) and physician instrument (column (2)), focusing on percentiles $u_j \in [0.2, 0.6]$, the region where both propensity score distributions have sufficient density. In the absence of cash transfers, veterans with lower levels of treatment resistance are at a higher risk of dying. For instance, the 5-year mortality rate for a veteran with $u_p = 0.3$ is approximately 9 pp higher than the rate for the entire sample.

In our second exercise, we examine differences in observable characteristics between always-takers and never-takers, abstracting away from the MTE framework. Always-takers, as described in Section 7.2, have the lowest levels of unobserved resistance because they receive generous levels of compensation regardless of their caseworker's and physician's leniency. Likewise, never-takers have the highest levels of unobserved resistance. To characterize these groups, we dichotomize the caseworker and physician instruments using cutoffs at the 50$^{\text{th}}$ percentile. As we noted regarding compliers earlier, always-takers (never-takers) for one instrument are not necessarily always-takers (never-takers) for the other.

Table 1.6 reports mean characteristics for always-takers and never-takers corresponding to each instrument and to both instruments jointly. For most observable characteristics, such as veterans' sex and race, there are minor differences between always-takers and never-takers. We also do not find a meaningful difference in the social deprivation associated with veterans' zip code. The major exception is veterans' medical acuity, which is 0.14 s.d. (s.e. = 0.006) higher for always-takers and 0.05 s.d. (s.e. = 0.005) lower for never-takers than for the overall sample. The fact that always-takers exhibit greater medical acuity mirrors our earlier finding that veterans with lower levels of resistance are at a higher risk of dying without the treatment.

7.5 Explaining positive selection on levels

Why do veterans in poorer health receive more generous cash transfers even when the transfers fail to yield significant health gains? One important consideration is that cash transfers may improve recipients' quality of life without necessarily reducing their risk of mortality. For example, cash transfers have been shown to improve happiness, with the strength of this relationship diminishing as income rises (Dwyer and Dunn 2022; Kangas et al. 2020). Caseworkers and physicians may target cash transfers to veterans who benefit in dimensions of well-being that are not apparent in the data available to us.

In addition, the pattern of reverse selection on gains that we observe does not imply that caseworkers and physicians target cash transfers with the intent of minimizing health outcomes, nor does it rule out the possibility that health gains positively enter their utility functions. Instead, reverse selection on gains indicates that the joint information and the decision rules used by caseworkers and physicians when awarding transfers are negatively correlated with the health gains derived from such transfers. This is sensible for several reasons. First, the VA's disability program is legally mandated to provide greater compensation to veterans with more severe impairments. A priori, we should expect those with a higher baseline mortality risk to be the least resistant to receiving larger transfers, evidence that the VA's disability program is functioning as intended.

Second, even if improving health is the primary objective of caseworkers and physicians, discerning who would benefit the most is challenging, as individual treatment effects are not directly observable. Caseworkers and physicians do not engage in long-term follow-ups with veterans, limiting their ability to learn this over time. In particular, visits with physicians are brief and focus exclusively on clinical information, constraining physicians' ability to target on gains. Our finding that reverse selection on gains is more pronounced for physicians aligns with this intuition.

We conclude by discussing three reasons why cash transfers might be ineffective for recipients in the poorest health. First, veterans with the most acute medical conditions may be at an advanced stage of their illness where additional resources or treatments do not lead to health improvements, reflecting diminishing health returns to cash. Many types of conditions, especially severe conditions, are also inherently unresponsive to cash transfers (e.g., no amount of money can replace a lost limb). Second, veterans with the most acute conditions might be more prone to increase their consumption of temptation goods, which do not contribute to, and may even offset, the health effect of cash transfers. In Section 5.2, for example, we find that cash transfers lead to a modest increase in the utilization of substance abuse services. As we noted previously, this increase could instead reflect greater detection and treatment of previously undiagnosed mental health conditions. Finally, structural barriers, such as living in an economically distressed neighborhood, can limit the plasticity of health for vulnerable populations and, consequently, the extent to which transfers improve health outcomes. Veterans in poorer health tend to live in more distressed neighborhoods. Indeed, in Section 6.2, we observe that the effectiveness of cash transfers varies considerably by commuting zone, with minimal efficacy observed in regions with a large manufacturing presence.

8. Discussion

An estimated 1.3 billion people live with a significant disability, representing 16% of the world's population (World Health Organization 2022). In every context that has been studied, people with disabilities die significantly earlier than their nondisabled counterparts, a health disparity that reflects the economic inequities faced by people with disabilities. Can economic interventions narrow such health disparities? This paper presents new evidence on the health effects of cash transfers for veterans with disabilities, a demographic group in which health and economic disparities are pronounced. To isolate the causal effect of cash transfers on health, we compile a dataset that combines disability claims, medical records, and death records for over 800,000 veterans applying for the VA's disability compensation program. Our empirical analysis leverages the quasi-random assignment of disability claims to caseworkers and physicians.

Our analysis produces three key empirical insights. First, for the typical veteran claimant, receiving $300 per month reduces 5-year mortality by 0.8 pp, equaling 9.5% of the mean rate. This reduction in mortality coincides with increased utilization of preventive health care and a moderate reduction in metabolic diseases. Second, cash transfers confer smaller health benefits for veterans in poorer health. Last, veterans in the poorest health receive the largest cash transfers despite their health improving the least from these transfers.

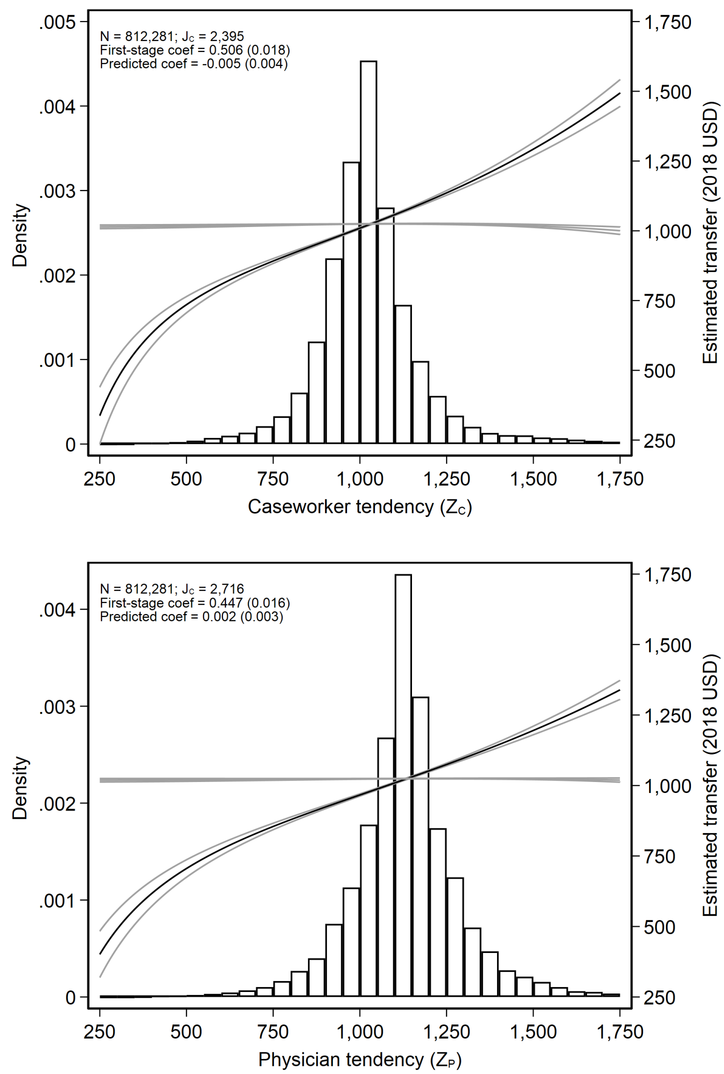
Compared to nondisabled veterans, veterans in our sample died at a staggeringly higher rate (4.3% versus 8.0% within 5 years of the disability rating process), mirroring disability-related health disparities in the civilian population. This paper highlights the potential of cash transfers as a tool to mitigate disability-related health disparities. Our finding that VA disability compensation lowers mortality risk for the typical recipient suggests that cash transfers narrow health disparities *between* veterans with and without disabilities. However, our secondary finding that the veterans in the poorest health benefit the least underscores the limitations of cash transfers in narrowing health disparities *within* the population of veterans with disabilities. Specifically, cash transfers can exacerbate existing health disparities among the population with disabilities, as veterans in better health experience the most significant gains from such transfers.

Building on these insights, we perform a back-of-the-envelope calculation using our causal estimates to understand how the VA's disability compensation program affects health disparities. We specifically evaluate the potential impact of two types of policies: increasing the generosity of cash transfers for current recipients and expanding the pool of eligible recipients. This analysis suggests that, with the set of recipients held constant, increasing transfers by $300 per month would narrow health disparities between disabled and nondisabled veterans by approximately one-quarter but would also widen disparities among current recipients by 12%. Alternatively, expanding benefits to include the sickest 10% of nondisabled veterans would increase health disparities between disabled and nondisabled veterans by one-fifth and would not affect disparities among veterans with disabilities. Overall, this counterfactual exercise highlights the trade-off between addressing between- versus within-group disparities.

This paper sheds new light on the economic and social determinants of health for vulnerable populations, particularly for people with disabilities. While the existing literature offers no clear explanation for why the health effect of cash transfers varies widely across studies, our study suggests that essential heterogeneity likely plays an important role. One policy implication of this type of heterogeneity is that cash transfers can be targeted to align with the goals of redistributive policies, recognizing the potential trade-offs between maximizing health outcomes and reducing health disparities. Last, while this paper directly informs cash transfer policies, such as the VA's disability compensation program, which augments income for individuals, our finding that their effect varies by commuting zone warrants further exploration. Given a growing body of evidence highlighting the place-based drivers of health, more can be done to understanding the interplay between cash transfers and place.

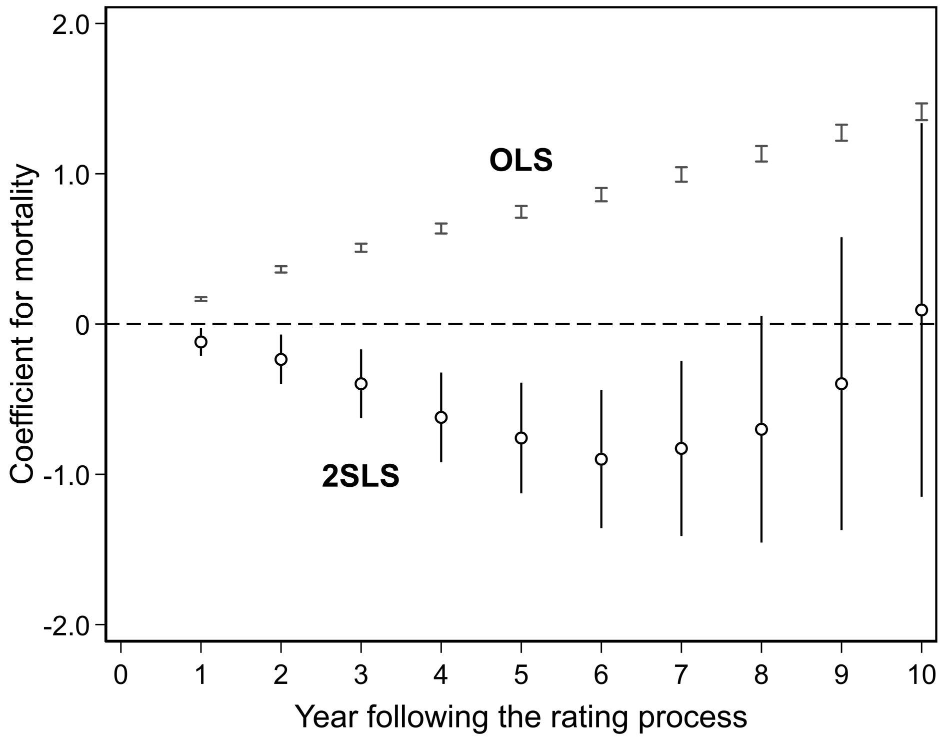
9. Figures

Figure 1.1: Validity Exercises for the Caseworker and Physician Leniency Instruments



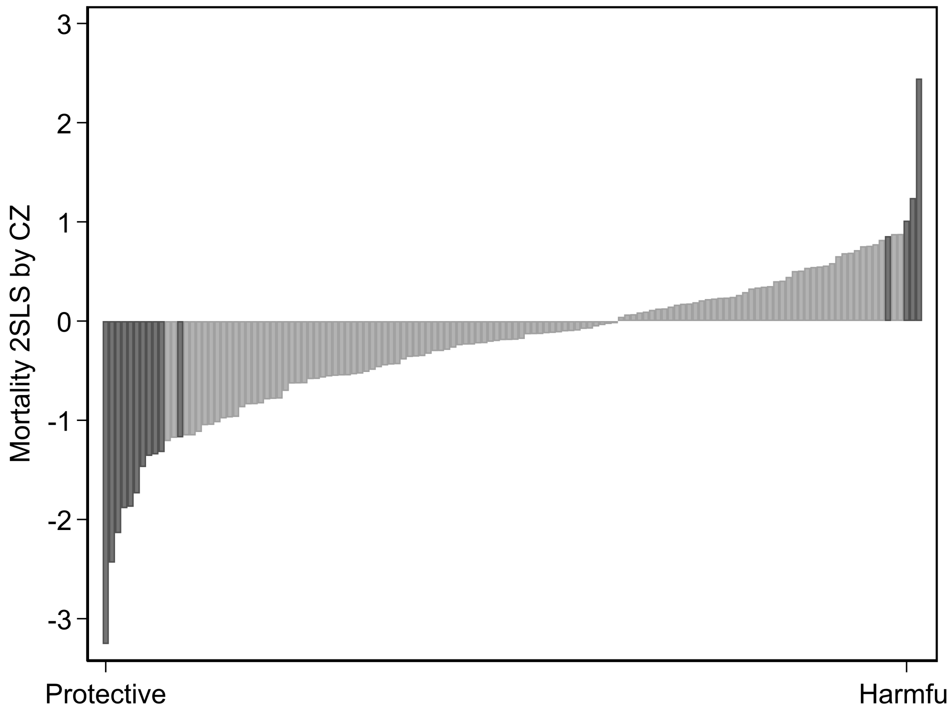
Notes: Panel A (top) plots the distribution of caseworkers' residualized leave-out treatment mean (disability compensation in 2018 dollars), from equation (1.3). Panel B (bottom) plots the distribution of physicians' residualized leave-out treatment mean, from equation (1.4). In both panels, the solid black line represents a fractional polynomial regression of the residualized treatment variable on the corresponding agent's residualized leave-out treatment mean. The dashed gray line represents a fractional polynomial regression of the residualized *predicted* treatment variable on the residualized leave-out treatment mean. The method to calculate the predicted treatment variable is described in Section 4.2. Ninety-five percent confidence intervals are shown. Sample sizes and regression coefficients (with standard errors in parentheses) for the first-stage and balance exercises are reported in the upper-left corners of each panel.

Figure 1.2: OLS and 2SLS Estimates for Mortality in Years 1 to 10



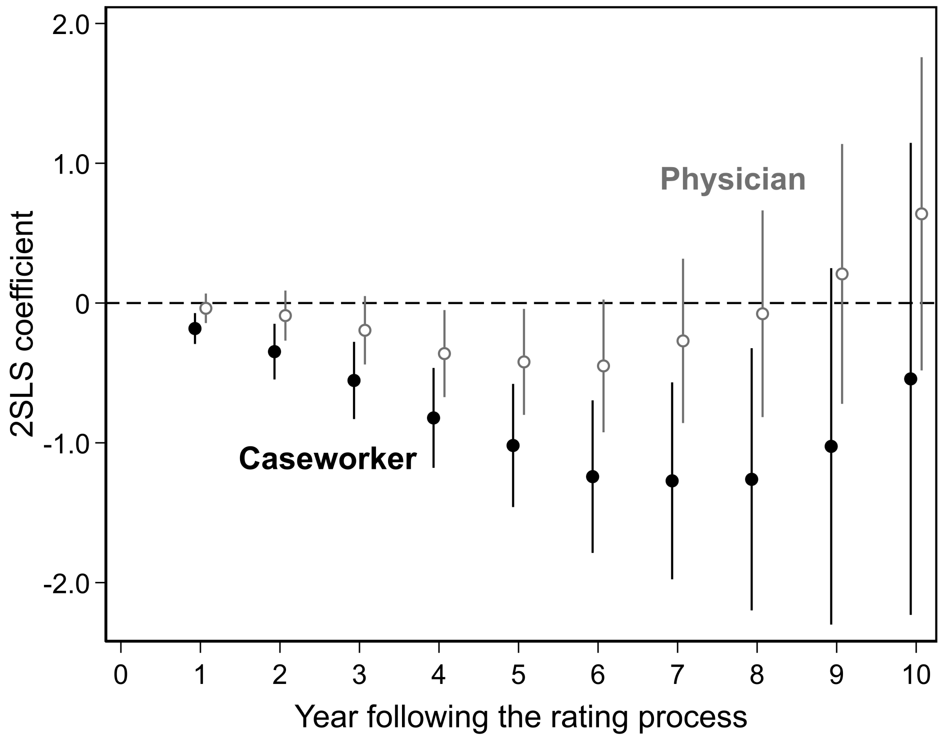
Notes: This figure shows ordinary least squares (OLS) and two-stage least squares (2SLS) estimates of the effect of disability compensation (scaled to \$300 per month in 2018 dollars) on mortality measured 1 to 10 years following the disability rating process. The 2SLS estimates are derived with both the caseworker and physician leniency instruments, respresented in equations (1.1) and (1.2). For the first 5 years, mortality outcomes are analyzed for the complete sample ($N = 812,281$). For years 6 to 10, mortality outcomes are analyzed in subsamples for which there is an adequate time horizon to assess mortality within that period. A negative value indicates a decrease in mortality in percentage points. Standard errors are clustered at the caseworker and physician level, with 95% confidence intervals shown.

Figure 1.3: 2SLS Estimates for 5-Year Mortality across Commuting Zones



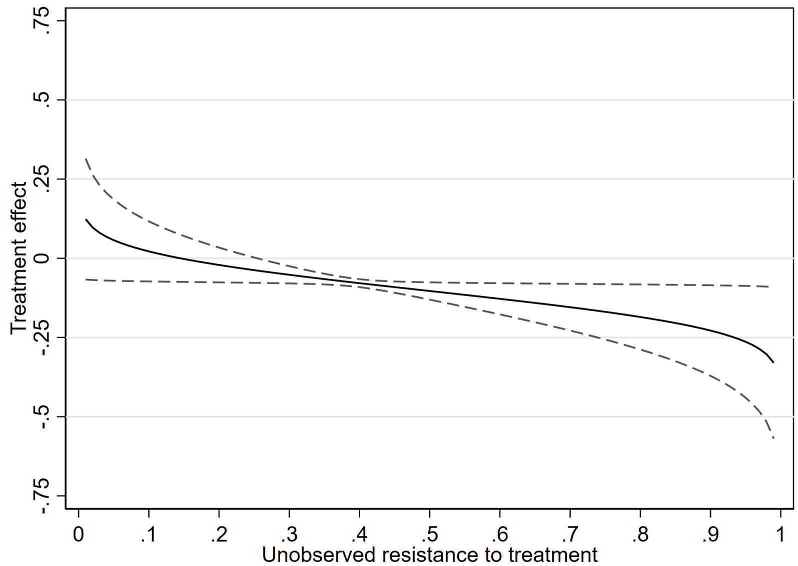
Notes: This figure shows empirical Bayes– (EB-) adjusted two-stage least squares (2SLS) estimates of the effect of disability compensation (scaled to $300 per month in 2018 dollars) on 5-year mortality, stratified by commuting zone (CZ). We separately analyze CZs with $\geq$ 1000 veterans from our sample. The 2SLS estimates are derived with both the caseworker and physician leniency instruments, respresented in equations (1.1) and (1.2). Standard errors are clustered at the caseworker and physician level, and estimates that are statistically significant at the 5% level are highlighted in dark gray. CZ-specific treatment effects, standard errors, and characteristics are provided in the appendix.

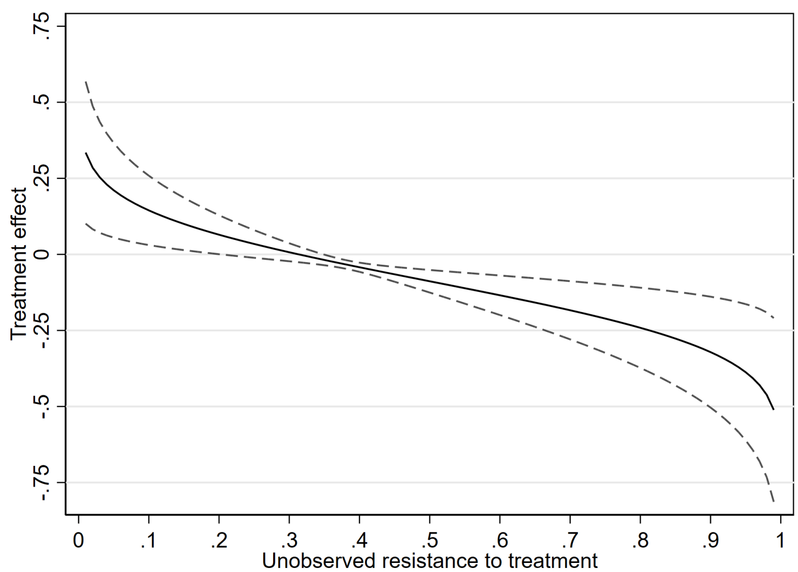
Figure 1.4: 2SLS Estimates for Mortality in Years 1 to 10, with Caseworker and Physician Instruments Separately



Notes: This figure displays two-stage least squares (2SLS) estimates of the effect of disability compensation (scaled to \$300 per month in 2018 dollars) on mortality measured 1 to 10 years following the disability rating process. The 2SLS estimates are derived separately with the caseworker and physician leniency instruments, as outlined in equations (1.1) and (1.2). For the first 5 years, mortality outcomes are analyzed for the complete sample ($N = 812,281$). For years 6 to 10, outcomes are analyzed in subsamples for which there is an adequate time horizon to assess mortality within that period. A negative value indicates a decrease in mortality in percentage points. Standard errors are clustered at the caseworker or physician level, depending on the analysis, with 95% confidence intervals shown.

Figure 1.5: Marginal Treatment Effects Curves for the Caseworker and Physician Leniency Instruments





Notes: Panel A (top) depicts the marginal treatment effect (MTE) curve for 5-year mortality associated with the caseworker leniency instrument, calculated with the local instrumental variables estimator in equation (1.9). The MTE curve plots the analytical derivatives at each value of unobserved resistance to treatment. The treatment variable is dichotomized (Section 7.2) to be either above or below $1,000 in 2018 dollars. Panel B (bottom) depicts the MTE curve associated with the physician leniency instrument. Dashed lines denote 95% confidence intervals. The distribution of propensity scores for each instrument can be found in the appendix.

10. Tables

Table 1.1: Characteristics of the Sample, Compared to Characteristics of Nondisabled Veterans and SSDI Recipients

	(1) Analytic sample	(2) Nondisabled veterans	(3) SSDI recipients
Demographics			
Age	50.2	49.3	55.0†
Male	90.7	90.6	50.0†
White	74.7	75.9	66.5†
Black	15.7	16.3	19.1†
Hispanic	7.5	5.6	-
Asian/Pacific Islander	3.9	2.7	-
Prior-year utilization			
Outpatient	51.7	46.8	-
Mental health services	11.0	6.3	-
Inpatient	4.0	1.9	-
Emergency room	3.7	2.8	-
Homelessness services	1.2	1.3	-
Comorbidities			
Hypertension	34.7	27.1	-
Depression	23.1	12.1	-
Diabetes mellitus	16.6	9.1	-
Obesity	10.6	8.4	-
Mortality rate			
1 year	1.4	0.6	-
5 year	8.0	4.3	14.0‡
10 year	17.9	9.4	-
Individuals (N)	812,281	58,613	9,218,080

Notes: This table provides summary statistics for individual demographics, healthcare utilization (any use in the year before the disability rating process), comorbidities, and mortality rates (measured 1, 5, and 10 years after the rating process). For year 10, mortality is analyzed in the subsample for which there is an adequate time horizon to assess mortality within that period ($N = 646,079$). Columns (1) and (2) display summary statistics for veterans in the analytic sample (who submitted a claim for disability compensation, as described in Section 3.2) and for nondisabled veterans (those never receiving disability compensation), respectively. Column (3) shows summary statistics for recipients of Social Security Disability Insurance (SSDI). †Demographics for SSDI recipients are sourced from tables 5.A1, 5.A5, and E-5.A1 in the "Annual Statistical Supplement to the Social Security Bulletin, 2022". ‡Mortality rates for SSDI recipients are derived by Gelber et al. (2023) with the 2006–10 actuarial estimates from Zayatz (2015). COPD = Chronic obstructive pulmonary disease.

Table 1.2: Effect of Disability Compensation on Mortality and Healthcare Utilization

	(1) Mean	(2) OLS	(3) 2SLS	(4) Interacted 2SLS
Mortality rate				
1 year	1.4	0.17 (0.005)	-0.12 (0.05)	-0.15 (0.04)
5 year	8.0	0.75 (0.01)	-0.76 (0.19)	-0.85 (0.16)
10 year	17.9	1.41 (0.02)	0.09 (0.63)	-0.005 (0.60)
Healthcare utilization				
Primary care	15.0	0.91 (0.01)	0.92 (0.17)	0.87 (0.16)
Mental health services	12.5	1.58 (0.19)	2.49 (0.27)	2.52 (0.27)
Inpatient	0.45	0.05 (0.001)	0.01 (0.01)	0.003 (0.01)
Emergency room	1.21	0.08 (0.002)	0.03 (0.02)	0.02 (0.02)
Homelessness services	0.87	-0.09 (0.003)	-0.21 (0.07)	-0.23 (0.07)

Notes: This table presents ordinary least squares (OLS) and two-stage least squares (2SLS) estimates of the effect of disability compensation (scaled to $300 per month in 2018 dollars). The outcomes variables are mortality rates (measured at 1, 5, and 10 years after the disability rating process) and healthcare utilization (represented as the number of utilization-specific encounters measured up to 5 years after the rating process). For year 10, mortality is analyzed in the subsample for which there is an adequate time horizon to assess mortality within that period ($N = 646,079$). Outcome variables are detailed in Section 3.1. Column (1) displays mean values. Column (2) displays the OLS estimates with standard errors in parentheses, clustered at the individual level. Column (3) displays the 2SLS estimates derived with both the caseworker and physician leniency instruments, as described in equations (1.1) and (1.2). Column (4) displays the 2SLS estimates derived with both the caseworker and physician leniency instruments, along with their interaction term. Standard errors for the 2SLS estimates in parentheses are clustered at the caseworker and physician levels.

Table 1.3: Effect of Disability Compensation on Morbidity

	(1) Mean	(2) OLS	(3) 2SLS	(4) Interacted 2SLS
Cardiovascular	51.5	1.32 (0.02)	-1.94 (0.59)	-2.17 (0.51)
General hypertension	47.7	1.18 (0.02)	-2.08 (0.57)	-2.31 (0.49)
Complicated hypertension	8.0	0.75 (0.01)	-0.76 (0.19)	-0.85 (0.16)
Cardiac arrhythmia	13.9	0.66 (0.01)	-0.57 (0.22)	-0.64 (0.20)
Congestive heart failure	6.1	0.63 (0.01)	-0.22 (0.13)	-0.30 (0.11)
Diabetes	23.3	0.80 (0.01)	-1.01 (0.38)	-1.22 (0.33)
General diabetes mellitus	22.7	0.76 (0.01)	-0.90 (0.37)	-1.10 (0.33)
Complicated diabetes	11.8	0.68 (0.01)	-0.64 (0.21)	-0.77 (0.19)
Mental health	45.4	2.46 (0.02)	5.58 (0.44)	5.58 (0.44)
Psychosis	5.4	0.63 (0.01)	0.70 (0.12)	0.65 (0.12)
Depression	44.4	2.34 (0.02)	5.49 (0.43)	5.50 (0.43)

Notes: This table presents ordinary least squares (OLS) and two-stage least squares (2SLS) estimates of the effect of disability compensation (scaled to $300 per month in 2018 dollars). The outcome variables are Elixhauser morbidity indicators measured up to 5 years after the disability rating process. Outcome variables are constructed from the outpatient and inpatient diagnostic codes detailed in Section 3.1. The analysis was conducted for a subset of individuals with ≥ 1 encounter within 5 years of the rating process ($N = 733,250$). Column (1) displays mean values. Column (2) displays the OLS estimates with standard errors in parentheses, clustered at the individual level. Column (3) displays the 2SLS estimates derived with both the caseworker and physician leniency instruments, as described in equations (1.1) and (1.2). Column (4) displays the 2SLS estimates derived with both the caseworker and physician leniency instruments, along with their interaction term. Standard errors for the 2SLS estimates in parentheses are clustered at the caseworker and physician levels.

Table 1.4: Effect of Disability Compensation on 5-Year Mortality across Impairment Types

	(1) N	(2) 5-Year Mortality	(3) Empirical Bayes OLS	(4) Empirical Bayes 2SLS
Auditory	184,261	10.39	0.71 (0.03)	-2.82 (0.24)
Musculoskeletal	180,674	6.35	0.49 (0.02)	-0.09 (0.31)
Mental health	79,381	6.97	0.29 (0.03)	-1.08 (0.23)
Neurologic	32,252	8.71	1.12 (0.04)	0.79 (0.32)
Cardiovascular	33,401	13.43	1.47 (0.05)	0.82 (0.58)
Dermatologic	25,096	8.79	0.45 (0.05)	-0.19 (0.57)
Respiratory	23,428	14.29	2.52 (0.07)	-0.81 (0.69)
Other	157,323	3.98	0.42 (0.02)	-0.04 (0.11)

Notes: This table presents empirical Bayes– (EB-) adjusted ordinary least squares (OLS) and two-stage least squares (2SLS) estimates of the effect of disability compensation (scaled to $300 per month in 2018 dollars) on 5-year mortality, stratified by the impairment type associated with the disability claim. Appendix Section 1.5 describes how the EB procedure is implemented. Column (1) displays the number of individuals claiming each impairment type. Column (2) displays the 5-year mortality rate for each impairment type. Column (3) displays the EB-adjusted OLS estimates with standard errors in parentheses, clustered at the individual level. Column (4) displays the EB-adjusted 2SLS estimates derived with both the caseworker and physician leniency instruments, as detailed in equations (1.1) and (1.2). Standard errors for the 2SLS estimates in parentheses are clustered at the caseworker and physician levels.

Table 1.5: Complier Characteristics

	(1) Overall	(2) Caseworker Compliers	(3) Physician Compliers
Elixhauser index	0	-0.097 (0.01)	-0.051 (0.01)
Social deprivation index	0	-0.012 (0.012)	-0.007 (0.006)
Age $\geq$ 50	53.0	-7.76 (0.88)	-3.38 (0.49)
Male	90.7	-1.31 (0.32)	-0.40 (0.49)
White	74.7	-0.56 (0.56)	-0.64 (0.32)
Black	15.7	1.65 (0.37)	1.34 (0.24)
Hispanic	7.5	1.04 (0.18)	0.66 (0.14)
Asian/Pacific Islander	3.9	-0.08 (0.15)	0.17 (0.09)

Notes: This table provides mean characteristics for all veterans in the sample, compliers for the caseworker instrument, and compliers for the physician instrument. The treatment variable has been dichotomized to be either above or below $1,000 in 2018 dollars (Section 7.2). We calculate differences between the overall and complier characteristics with two-stage least squares, utilizing the first stage in equation (1.2) and replacing the outcome variable in equation (1.1) with $X_i D_i$, where X_i is the characteristic for veteran i. Standard errors in parentheses are clustered at either the caseworker or physician level, depending on the complier.

Table 1.6: Medical Acuity by Resistance of Treatment

	(1) Caseworker	(2) Physician
Panel A. Response type	*Elixhauser index*	
Never-takers	0.164	0.158
	(0.005)	(0.005)
Always-takers	-0.060	-0.051
	(0.005)	(0.004)
Panel B. Unobserved resistance	*Relative mortality, Y(0)*	
20^{th} percentile	$+14.5$	$+14.3$
30^{th} percentile	$+6.4$	$+6.4$
40^{th} percentile	$+0.1$	$+0.1$
50^{th} percentile	-5.0	-6.1
60^{th} percentile	-13.2	-12.9

Notes: Panel A (above) provides mean characteristics for always-takers and never-takers for both the caseworker and physician instruments. The treatment variable is dichotomized to be either above or below $1,000 in 2018 dollars (Section 7.2). Always-takers are defined as veterans who receive the dichotomized treatment when they are assigned an instrument below the median. Never-takers are defined as veterans who do not receive the dichotomized treatment when they are assigned an instrument above the median. The Elixhauser index is a measure of acuity based on pre-existing comorbidities, standardized to have a mean of zero and a standard deviation of one. Standard errors in parentheses are clustered at either the caseworker or physician level, depending on the instrument. Panel B (below) presents estimated counterfactual mortality rates $Y(0)$ relative to the mean mortality of the entire sample, as a function of veterans' unobserved resistance to treatment. Refer to Section 7.4 for the estimation of counterfactual mortality rates.

Chapter 2

Inappropriate Prescribing to Older Patients by Nurse Practitioners and Primary Care Physicians

1. Introduction

The United States is facing a major shortage of primary care physicians (Association of American Medical Colleges 2021). There is considerable potential for nurse practitioners (NPs) and other nonphysician providers to ameliorate access problems associated with this shortage (Iglehart 2013; Pohl et al. 2010). However, restrictive scope of practice laws and longstanding concerns about quality and safety of care constrain the roles such providers are permitted to play in the delivery of primary care services (American Medical Association 2018; Xue et al. 2016; National Council of State Boards of Nursing 2018).

In the last decade, driven by the need to improve access to primary care and curb its rising costs, many states have enacted laws liberalizing scope of practice rules to enhance the autonomy of nonphysician providers (National Conference of State Legislatures 2017). A recent wave of reforms conferring prescriptive authority on NPs is particularly noteworthy. A total of 32 states and the District of Columbia have legislated to allow NPs to prescribe medications without physician supervision; half of these states—including the populous states of New York, Illinois, Florida, Massachusetts, and California—made this move in the last decade.

Professional medical organizations have vociferously opposed these reforms (Institute of Medicine 2011; Robeznieks 2022). Opposition rests chiefly on the contention that expanding prescribing authority to nonphysician providers will have adverse effects on quality of care (Institute of Medicine 2011). Does the safety or appropriateness of NP prescribing fall short? Previous studies comparing the performance of NPs and physicians have not demonstrated such inferiority (Laurant et al. 2018; Yang et al. 2021; Yang et al. 2018; McMichael 2021b; Alexander and Schnell 2019; Mundinger et al. 2000; Jiao et al. 2018; McMichael 2021a). But these studies are relatively small in scale and have tended to focus on specific forms of prescribing.

We analyzed the prescribing patterns of NPs and physicians in 29 states where NPs had prescriptive authority during the 2013-2019 study period. Using comprehensive data on medications prescribed to Medicare beneficiaries in those states, we calculated and compared rates of inappropriate prescribing to older patients by NPs and primary care physicians, defining inappropriateness according to the Beers Criteria developed by the American Geriatrics Society (American Geriatrics Society 2012;**?**;American Geriatrics Society 2019). To align with Medicare data and the population to which Beers Criteria apply, we defined older patients as those aged 65 or older. We hypothesized that NPs would not have higher rates of inappropriate prescribing than primary care physicians.

2. Methods

2.1 Data

We formed the study cohort by linking two data sets from Medicare's Provider Utilization and Payment Data collection: the Part B Supplier file and the Part D Prescriber file. The Part B Supplier file consists of annual, clinician-level service data drawn from Medicare Part B claims. Clinicians with a valid National Provider Identifier (NPI) who submitted 11 or more noninstitutional services for reimbursement in a calendar year entered the file for that year. The Part D Prescriber file consists of prescription (original and refill) data drawn from Medicare Part D claims. Medications are categorized by their generic name. Clinicians who prescribed 11 or more medications in a year entered the file for that year. We obtained the Part B and Part D files for years 2013 through 2019.

2.2 Study cohort

After using the NPI to link the Part B and Part D files at the clinician level, we constructed the study cohort at the clinician-year level (Appendix Section 4.4). We restricted the sample to clinicians in the 29 states that had granted NPs independent prescriptive authority by January 1, 2019. The authority granted is independent in the sense that scope of practice rules do not mandate that NPs be supervised by or collaborate with physicians when prescribing. (Additional details of how these states were identified and classified are provided in Appendix Section 2.2).

NPs and primary care physicians were selected according to CMS specialty codes in the Part B file: code 50 for NPs, and codes 01 (general practice), 08 (family practice), 11 (internal medicine), and 38 (geriatric medicine) for physicians. For clinicians who reported multiple specialty codes, we chose a primary specialty by selecting the code associated with the largest number of billed services. Finally, we excluded clinicians who wrote prescriptions for fewer than 100 patients aged 65 years or older ("older patients").

2.3 Study variables

The American Geriatrics Society's Beers Criteria identifies medications with an unfavorable risk-benefit ratio in older adults. This guideline was developed by a panel of experts in geriatric care and pharmacotherapy, and has been used widely in studies of prescribing quality among older patients (Budnitz et al. 2011; Jano and Aparasu 2007; Rankin et al. 2018; Kotwal et al. 2021). Three successive versions of the Beers Criteria were promulgated during our study period (in 2012, 2015, and 2019); for each study year, we applied the most recent guideline. Details of the Beers Criteria are provided in Appendix Section 2.3.

Our primary outcome measure focused on medications classified by the Beers Criteria as "potentially inappropriate in most older adults". Approximately half of all medications listed in the guideline carried this rating and, unlike other Beers Criteria categories, the directive is universal in nature and untethered to particular patient or clinical circumstances. We specified the outcome measure as a rate: the fraction a clinician's total prescriptions written for older patients during years the clinician was under observation that referred to a drug in the category described above. Thus, the measure took the form of a continuous variable ranging from 0 to 100. To better understand patterns of compliance with the guideline, we analyzed three secondary outcomes: percentile of inappropriate prescribing rate; inappropriate prescribing rate stratified by medication class; and inappropriate prescribing rate stratified by geographic area (practice state and hospital referral region).

Additional variables drawn from the Part B Supplier and Part D Prescriber files included clinician gender and patient volume. We classified clinical settings according to evaluation and management codes (99091 – 99499) and practice locations according to the population size of core-based statistical areas in which they were located. The files included patient risk scores, calculated by CMS at the clinician-year level, which are intended to provide an overall measure of the illness severity of a clinician's Medicare patient population (Pope et al. 2004). Practice experience, defined as the number of years since obtaining an NP or medical degree, was derived from the Medicare Provider Enrollment, Chain, and Ownership System. Additional details on the covariates used in our analyses are provided in Appendix Section 2.4.

2.4 Statistical analysis

We conducted logistic regression analysis at the clinician-year level. To account for the repeated observation of the same physicians over multiple years, we clustered standard errors at the clinician level. We adjusted the models for 7 clinician-level variables: gender, patient volume, years of practice experience (10-year bands), patient risk score (quarter-point bins), clinical practice setting (7 categories), practice location (5 categories), state, and year. Estimates from the models represent population-averaged effects (Diggle et al. 2002).

To estimate how NP and physician populations were arrayed across percentiles of inappropriate prescribing, we used hierarchical mixed-effects logistic regression (Normand et al. 2016; Christiansen and Morris 1997), adjusting for the covariates as the main model. Posterior means for NPs and physicians were estimated using the empirical Bayes approach, combining clinicians' own data and the empirical distribution, weighted by their respective uncertainties. Such individual estimates borrow strength from overall estimates and can be better predictors of individual performance (Carlin and Louis 2008; Efron and Morris 1973). Further explanation and details of our approach to the statistical analysis are provided in Appendix Section 2.5.

We also conducted six sensitivity analyses to probe the robustness of our results (details of these analyses are provided in Appendix Section 2.7). First, we re-ran our main analyses after excluding patient risk scores as an adjustor. The goal was to assess potential bias arising from the fact that we did not adjust for patient-level characteristics (Oster 2019). Second, to address possible biases introduced by missing data on practice experience for 24% of NPs and 19% of physicians, we computed worst-case bounds for our estimates (Horowitz and Manski 2000). Third, to assess whether lower prescribing volume among NPs affected our distributional results, we re-calculated our main estimates on a subsample that excluded clinicians who prescribed for fewer than 500 older patients per year. Fourth, recognizing that Beers Criteria may not be appropriate quality measures for patients receiving hospice or palliative care, we re-ran our main analysis after excluding observations from clinicians who predominantly worked in three settings where such care is often delivered: nursing facilities, assisted living facilities, and home visits. Fifth, we probed our regression-based approach by comparing it to propensity score matching and doubly robust methods. Finally, a small proportion of NPs deliver specialty care, and they are not directly identifiable in CMS data. To test whether their prescribing behavior may have skewed our results, we used diagnosis and clinical setting codes to identify NPs who likely were not practicing in a primary care setting, and re-estimated our main results using a sample that excluded them.

3. Results

3.1 Sample characteristics

The study cohort comprised 23,669 NPs and 50,060 primary care physicians (Table 2.1). A total of 89% of NPs were female, compared with 39% of physicians. On average, NPs had fewer years of practice experience than physicians (7.1 vs. 19.8 years), treated fewer older patients each year (199 vs. 272 patients), and wrote fewer than one third the annual number of prescriptions for older adults (1,217 vs. 3,800 prescriptions). The older patient populations NPs treated had slightly higher average risk scores than those physicians treated (1.49 vs. 1.36).

Both NPs (79%) and physicians (75%) treated most of their older patients in outpatient settings. NPs were more likely than physicians to treat their older patients in facility settings (10.8% vs. 5.4%) and physicians were more likely to treat in inpatient settings (16.3% vs. 5.6%). In addition, NPs were more likely than physicians to practice outside large metropolitan areas (59% vs. 48%).

3.2 Crude rates of inappropriate prescribing

On average, crude rates of inappropriate prescribing for NPs and primary care physicians were very similar (Figure 2.1). Among NPs, the crude rate was 1.63 per 100 prescriptions (median, 0.74; standard deviation, 2.36); among physicians, it was 1.69 per 100 prescriptions (median, 1.49; standard deviation, 1.57).

3.3 Adjusted odds and rates of inappropriate prescribing

In the multivariate analysis, the odds of inappropriate prescribing for NPs and physicians were virtually identical (odds ratio, 0.99; 95% CI, 0.97 to 1.01). The similarity was also evident in a comparison of population-averaged rates of inappropriate prescribing among NPs (1.66; 95% CI, 1.63 to 1.69) and physicians (1.68; 1.67 to 1.69) (Figure 2.1).

However, these averages mask substantial underlying differences in inappropriate prescribing patterns. NPs were overrepresented both among clinicians who were consistently compliant with Beers Criteria and among clinicians who were least compliant (Figure 2.2). Specifically, NPs accounted for 32.1% of the total study sample, but 51.8% of clinicians in the lowest decile of inappropriate prescribing and 48.8% of clinicians in the highest decile. Other calculations indicating NPs overrepresentation at the highest and lowest levels of inappropriate prescribing are provided in Appendix Section 2.6.

3.4 Inappropriate prescribing by medication class

Ten medication classes accounted for 99.5% of all inappropriate prescribing detected in our study. Figure 2.3 compares inappropriate prescribing rates for NPs and physicians within each of these medication classes. NPs had higher rates of inappropriate prescribing of musculoskeletal medications, such as muscle relaxants (0.37 vs. 0.30) and antispasmodics (0.15 vs. 0.06), whereas physicians had higher rates of inappropriate prescribing of sedatives, such as antidepressants (0.48 vs. 0.39) and hypnotics (0.44 vs. 0.36). Within other medication classes, there were relatively small differences in the inappropriate prescribing rates of NPs and physicians.

3.5 Inappropriate prescribing by state

NPs had lower adjusted rates of inappropriate prescribing than physicians in 9 of the 29 states that permitted NPs to prescribe medications without physician supervision during our study period (Figure 2.4). In 6 states, NPs exhibited higher adjusted rates of inappropriate prescribing than physicians. In 14 states, there were minimal or no differences in the observed rates of inappropriate prescribing. Overall, these differences were modest; two-thirds of states had ratios of NP-to-physician rates of inappropriate prescribing between 0.9 and 1.1 (mean, 0.98; median, 0.95).

By contrast, differences in rates of inappropriate prescribing *between* states was large. For instance, rates of inappropriate prescribing for both NPs and physicians in Utah and Virginia, which were the highest among the states we examined, were nearly two-fold greater than the corresponding rates in Hawaii and the District of Columbia, which had among the lowest. A comparison of inappropriate prescribing by hospital referral region showed a similar pattern: there was considerably more variation across regions than between NPs and physicians within most regions.

3.6 Sensitivity analyses

The six sensitivity analyses produced estimates that were largely consistent with the main results with respect to both mean inappropriate prescribing rates and the overrepresentation of NPs among clinicians with the highest and lowest rates. First, excluding patient risk scores also produced a non-inferiority result for NPs' quality of prescribing. Second, imputing worst-case bounds associated with missing data on practice experience did not change the adjusted rates of inappropriate prescribing. Third, restricting the analytic sample to high-volume prescribers resulted in the loss of 89% of clinician-years. Analysis of this restricted sample detected a lower rate of inappropriate prescribing among NPs (OR, 0.75; 95% CI, 0.70-0.79), thus reinforcing a non-inferiority result. Fourth, excluding clinicians who delivered care predominately in nursing facilities, assisted living facilities, and home visits produced estimates that were nearly identical to our main results. Fifth, alternative matching and doubly robust methods produced results that were consistent with our regression-based approach. Finally, our main results did not change after excluding from the analysis NPs whose patterns of care suggested they were not practicing in primary care settings. Detailed results for all five sensitivity analyses are provided in Appendix Section 2.7.

4. Discussion

In 29 states that have granted NPs prescriptive authority, this study found that NPs were no more likely than primary care physicians to prescribe inappropriately to older patients. Both types of clinicians averaged approximately 1.7 inappropriate prescriptions for every 100 prescriptions written. However, the distribution of inappropriate prescribing differed substantially between the two groups, with NPs disproportionately likely to appear in the "tails". Specifically, NPs were overrepresented both among clinicians in the highest and the lowest deciles of inappropriate prescribing.

Most previous studies comparing the prescribing performance of physicians and non-physician providers have found little or no differences (Laurant et al. 2018; Yang et al. 2021; Yang et al. 2018; McMichael 2021b; Alexander and Schnell 2019; Mundinger et al. 2000; Jiao et al. 2018; McMichael 2021a). For example, a cohort study that followed veterans with diabetes for five years found no difference in prescribing patterns between NPs and physicians (Yang et al. 2018), though differences were detected within emergency departments (Chan and Chen 2023). These studies exemplify another feature of previous comparisons of NPs' and physicians' prescribing performance: they have tended to focus on specific medications (e.g., diabetes) or settings (e.g., Veterans Health Administration).

Our study was substantially larger in scope than previous ones. It examined 760 million prescriptions written over a 7-year period by more than 50,000 primary care physicians and nearly 25,000 NPs. The states we studied account for 31% of the country's current NP workforce (Kaiser Family Foundation 2022). Our bottom-line finding—that NPs are not inferior in terms of their odds of inappropriate prescribing—aligns with the non-inferiority detected in most other studies.

A frequently overlooked aspect of comparative health services research is that quality and safety measures that have similar mean values may have very different underlying distributions. Our results demonstrate this along two dimensions. The differences in inappropriate prescribing rates *among* NPs and primary care physicians, respectively, were much larger than the mean differences between the two groups. Similarly, the differences in inappropriate prescribing rates across states and hospital referral regions for both NPs and primary care physicians were large.

Our finding that NPs were disproportionately overrepresented among clinicians with the highest rates of inappropriate prescribing raises concern about the "tail" of prescribing performance among NPs. Steps to identify this population and facilitate improvement, including better adherence to guidelines like Beers Criteria, are warranted. However, elimination or broad-based contraction of prescriptive authority for NPs would be a crude response to this issue, given NPs were also substantially more likely than physicians to appear among clinicians with the lowest rates of inappropriate prescribing.

Our study has several limitations. First, it was restricted to prescriptions for older patients and may not be generalizable to younger patient populations. Older adults are nonetheless a key group in which to measure prescribing quality. They account for a high proportion of all prescriptions written; they are also especially vulnerable to adverse drug events from inappropriate prescribing, owing to comorbidities, prevalent polypharmacy, and age-related physiological changes (Hilmer and Gnjidic 2009).

Second, while the American Geriatrics Society's Beers Criteria is the most widely used guideline for prescribing in older patients, it has several weaknesses as a quality measure (Aguiar et al. 2021; Grace et al. 2014; Motter et al. 2018). It does not address the appropriateness of many drugs, including some (e.g., warfarin, insulin) that are frequently associated with adverse drug events (Budnitz et al. 2007); our study narrowed even further, focusing on Beers Criteria drugs that carried a general recommendation against prescribing. In addition, the Beer Criteria does not address underuse of helpful medicines, nor may it be applicable for older adults receiving hospice or palliative care where special risk-benefit calculations are required, although our estimates proved robust to sensitivity analyses that attempted to excluded care delivered in those settings.

Third, our outcome measure is overinclusive. The Beers Criteria does not recommend that the drugs we used to identify inappropriate prescribing *never* be prescribed. Clinical guidelines are rarely that definitive. Rather, these drugs are designated as "potentially inappropriate for most older adults". This is important context for interpreting the absolute rates of inappropriate prescribing we report, but it is unlikely to threaten the fundamental validity of the comparative assessment.

Fourth, some prescriptions in our sample that were attributed to physicians may not have been written by them. Medicare's incident-to billing rules allow certain services performed by nonphysician providers to be billed under the NPI of a collaborating physician. (Moreover, there are financial incentives to bill in this way, because Medicare reimburses those services at 85% of the rate set by the physician fee schedule.) Nationally, an estimated 5% of office visit services billed by physicians are performed by nonphysician providers (Staff 2019). However, because incident-to billing requires the physician to have played a supervisory role in the care, it is not the type of autonomous NP prescribing we sought to measure.

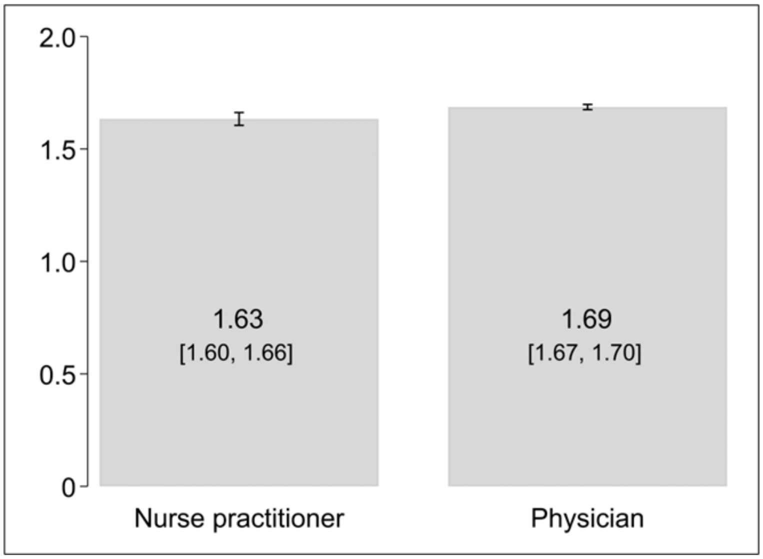
Fifth, our data did not allow us to distinguish between original prescriptions and refills. If NPs were more likely to provide routine or follow-up care, then differences in prescribing patterns between physicians and NPs may be masked by prescribing inertia, possibly inflating estimated rates of inappropriate prescribing for NPs.

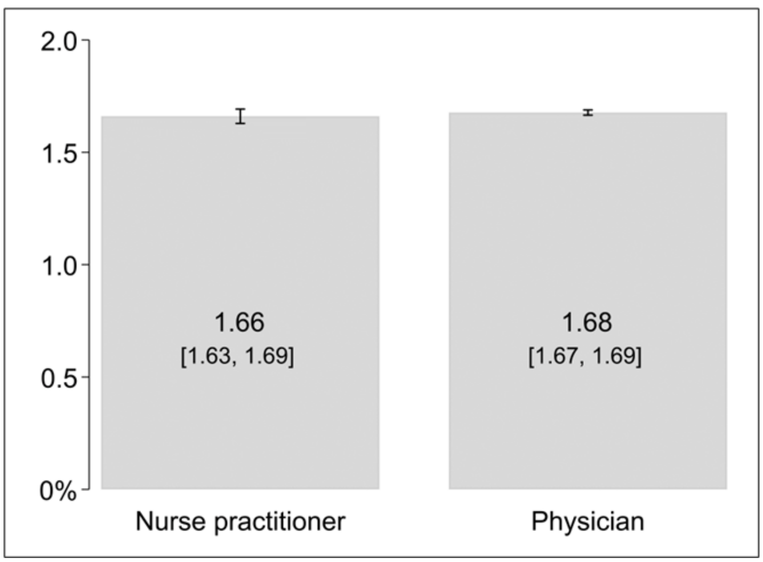
Finally, our analysis did not adjust for patient characteristics. If the patients either NPs or primary care physicians prescribed for were systematically more complicated and sicker, that may have affected the clinician group's propensity to prescribe inappropriately. The risk score variable included in our analysis suggested that severity of illness was similar for the two groups (though slightly higher for NPs), and we adjusted for practice setting, but these are imperfect proxy measures. The extent to which this form of confounding threatens the validity of our results depends on how they are interpreted. As estimates of what would happen to prescribing quality if the patients treated by physicians had been treated by NPs instead, the threat is real. As a reflection of prescribing quality in the real world, where patients naturally sort between NPs and physicians based on a multitude of unobserved factors, such confounding is less consequential, and the type of comparison we have conducted should reasonably reflect the outcome of conferring prescriptive authority on NPs.

This study adds to growing evidence indicating that when prescriptive authority is expanded to include NPs these new prescribers do not perform worse than physicians. Our findings regarding patterns of inappropriate prescribing among NPs and primary care physicians have useful lessons for policymakers, lawmakers, and regulators. If expanding patient access while ensuring quality and safety systemwide is the goal, fixation on the question of whether NPs or other nonphysician providers should be allowed to prescribe may be less impactful than identifying and addressing deficient performance among *all* clinicians who prescribe. Use of clinician-level performance measures, coupled with efforts to improve prescribing at the organizational and individual levels, are one approach for achieving this. Technological interventions, such as prescription drug monitoring, have a role to play, as do initiatives aimed at ensuring better adherence to trusted guidelines like the Beers Criteria.

5. Figures

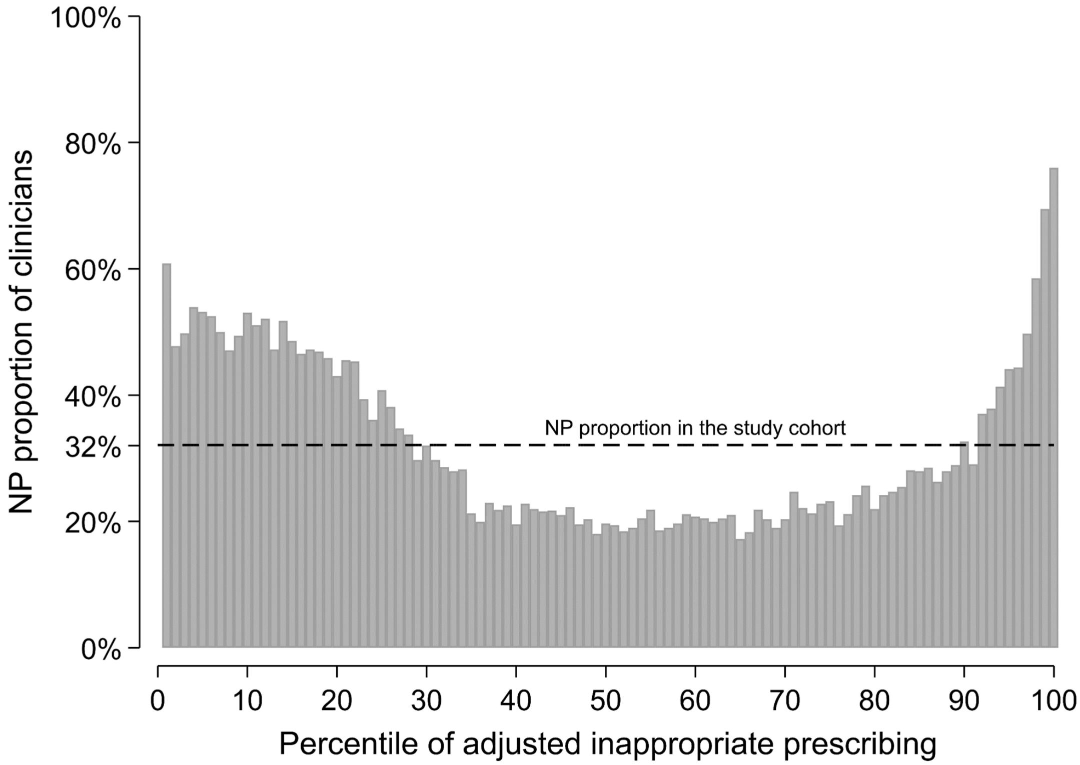
Figure 2.1: Mean Inappropriate Prescribing Rates for Nurse Practitioners and Primary Care Physicians





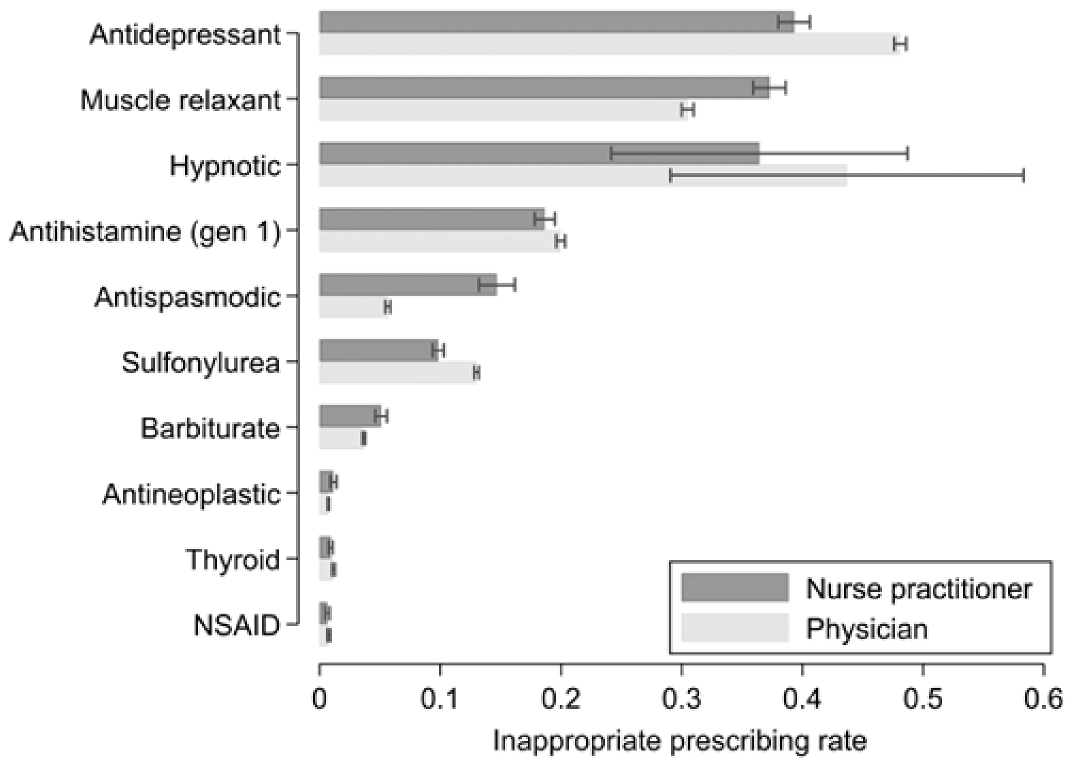
Notes: Panel A (top) refers to the unadjusted rate (per 100 prescriptions) of 2013-19 Medicare Part D medications for older patients for which Beers Criteria generally recommends against prescribing for most adults. Panel B (bottom) refers to the regression-adjusted rate of medications for older patients with the same Beers Criteria designation. Adjusted estimates controlled for clinician gender, years of practice experience, patient volume, patient risk score, clinical setting, practice location, year, and state. Standard errors were clustered at the clinician level, and 95% confidence intervals are shown.

Figure 2.2: Proportion of Clinicians Who Were NPs by Percentile of Adjusted Inappropriate Prescribing Rate



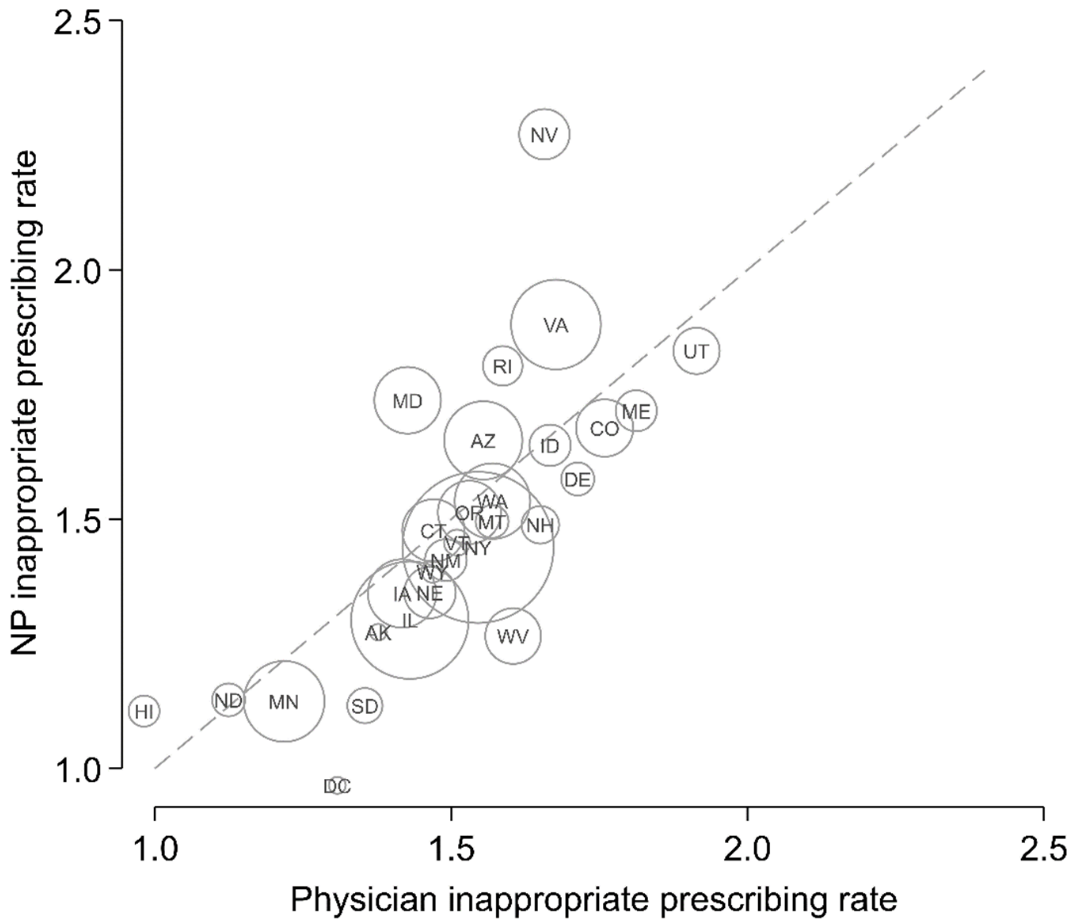
Notes: Clinician-level inappropriate prescribing rates were estimated using hierarchical, mixed-effects logistic regression, which tends to shrink toward the mean to improve the quality of predictions. Percentiles were based on these modeled rates. "NP proportion of clinicians" refers to the share of clinicians in each percentile who were classified as an NP. A total of 32% of clinicians in our study cohort were NPs (dashed line). Adjusted estimates controlled for clinician gender, years of practice experience, patient volume, patient risk score, clinical setting, practice location, year, and state.

Figure 2.3: Inappropriate Prescribing Rates for Nurse Practitioners and Primary Care Physicians, Stratified by Medication Class



Notes: Inappropriate prescribing rates (per 100 prescriptions) were stratified according to medication classes in the American Geriatrics Society's Beers Criteria. The ten most prescribed medication classes shown here accounted for 99.5% of all inappropriate prescribing measured in our study. Estimates were adjusted for clinician gender, years of practice experience, patient volume, patient risk score, clinical setting, practice location, year, and state. Standard errors were clustered at the clinician level, and 95% confidence intervals are shown.

Figure 2.4: Inappropriate Prescribing Rates for Nurse Practitioners and Primary Care Physicians in 29 States, 2013—19



Notes: Each bubble plots the inappropriate prescribing rate (per 100 prescriptions) stratified according to clinicians' practice state. Bubble size represents each state's volume of Medicare Part D claims for older adults in 2019. The dashed 45-degree line indicates equivalent rates of inappropriate prescribing between NPs and primary care physicians. Estimates were adjusted for clinician gender, years of practice experience, patient volume, patient risk score, clinical setting, practice location, and year.

6. Tables

Table 2.1: Demographic, Patient, and Locational Characteristics of Nurse Practitioners and Primary Care Physicians in the Study Sample

	Nurse practitioners N = 23,669	Physicians N = 50,060
Clinician characteristics		
Female, %	88.9%	38.7%
Male, %	11.1%	61.3%
Practice experience, mean	7.1	19.8
Prescriptions per year, mean	1,271	3,800
Physician specialty, %		
Family practice		52.4%
Internal medicine		45.3%
General practice		1.4%
Geriatric medicine		0.9%
Patient characteristics, mean		
Older patients per year	199	272
Patient risk score	1.49	1.36
Clinical setting, %		
Outpatient	78.8%	75.4%
Inpatient	5.6%	16.3%
Emergency	2.3%	1.2%
Critical care	0.1%	0.2%
Nursing facility	8.8%	5.0%
Assisted living facility	2.0%	0.5%
Home visit	1.7%	0.4%
Practice location, %		
Large metropolitan	40.6%	51.8%
Medium metropolitan	23.2%	19.5%
Small metropolitan	17.1%	12.8%
Micropolitan	12.2%	9.5%
Remote	6.9%	6.3%

Notes: Percentages may not total 100 due to rounding. Additional details on the variables shown in this table are provided in the Supplementary Appendix. Practice experience is defined as the number of years since the clinician received their nurse practitioner or medical degree, measured at the year they entered the cohort. Data on practice experience were missing for 5,736 (24%) NPs and 9,622 (19%) physicians.

Chapter 3

Vaccine Incentives Hurt Intrinsic Motivation: Evidence from a Survey Experiment

1. Introduction

Dating back to Pigou (1920), economists have touted monetary incentives as a tool for encouraging behaviors that generate positive externalities, such as conditional cash transfers to promote human capital accumulation (IMF 2019) and payments to improve compliance with health recommendations (Ensor and Cooper 2004). During the COVID-19 pandemic, a number of high-profile economists proposed paying vaccine-hesitant adults to get vaccinated against the virus, arguing that the social benefits of vaccination outweigh the incentive's cost.[1] However, critics of these proposals cautioned against using monetary incentives in this manner, citing instances of when monetary incentives backfired and undermined individuals'

[1]Economists have suggested cash transfer of up to $1,000 for Americans to get vaccinated against COVID-19 (Mankiw 2020; Litan 2020). Private businesses have also began paying their employees to get vaccinated against the virus, and the state of West Virginia offered bonds for vaccination (Krouse 2021).

intrinsic motivation (Loewenstein and Cryder 2020).[2] Indeed, understanding whether vaccine incentives would meaningfully raise vaccination rates requires knowing how many individuals (and for whom) the reward would persuade, deter, or not influence.

It is well-known that monetary incentives induce competing effects, causing their overall impact on the incentivized behavior to be theoretically ambiguous. The price effect makes getting vaccinated more attractive by lowering the relative price of vaccines compared to other goods. Conversely, the psychological effect undermines free-choice intrinsic motivation, pushing individuals' behavior away from the intended objective of the incentives (Ariely et al. 2009; Bowles and Polania-Reyes 2012). Yet, few studies have empirically estimated the magnitude of these psychological effects in the vaccine-hesitant population, and evidence on the behavioral channels underlying these psychological effects is extremely limited.

We conducted a survey experiment to estimate the share of vaccine-hesitant adults whose intrinsic motivation to get vaccinated was harmed by monetary incentives. These individuals needed no incentive to consider getting vaccinated but would have declined the vaccine if primed to consider such a reward. Our experiment employed behavioral priming techniques commonly used in psychology research in order to decouple incentives' psychological effect from their price effect. To give economic structure to the results from our experiment, we developed an econometric model of treatment take-up in the style of Heckman and Pinto (2021), enabling us to characterize individuals whom monetary incentives deterred from getting vaccinated.

In our survey experiment, 513 vaccine-hesitant adults were randomly assigned to one of two treatment arms. Participants in the control arm were asked to consider getting vaccinated against COVID-19 without any mention of monetary incentives, while those in the incentive arm were asked to consider getting vaccinated for varying levels of payment,

[2] For example, a Swedish field experiment found that the offer of payment for blood donation reduced the likelihood that women donated blood (Mellström and Johannesson 2008). In a separate experiment, high school teenagers raised less money for charity when they were paid a commission to do so (Gneezy and Rustichini 2000). See Gneezy et al. (2011) for a review of this literature.

including zero payment. By priming the latter group with the notion of incentives but offering zero payment, our experiment was able to separate the price effect from the psychological effect. Intuitively, if incentives did not harm individuals' attitudes toward vaccination, then their willingness to get vaccinated would have been equal between the control arm and the incentive arm when offered zero payment.

We found that monetary incentives deterred 14% of participants from vaccination; they considered getting the vaccine without any incentive (control arm) but would have rejected it if offered a reward (incentive arm). Deterred individuals differed from other participants in the experiment along several dimensions. They were significantly more likely to be male and be from a racial or ethnic minority group — a pattern that resembles existing disparities in medical mistrust by race and gender. Deterred individuals also had greater baseline levels of skepticism toward the safety of the vaccine, underscoring the limitations of vaccine incentives among those who are the most resistant to vaccination.

What are the channels through which payments discourage people from getting vaccinated? First, individuals may infer from the existence of an incentive that vaccines are unsafe and that the payment is being offered to them in recompense (Cryder et al. 2010; Milgrom and Roberts 1986). Second, the incentive may crowd out intrinsic motivation by introducing doubt about individuals' prosocial motives (Frey and Oberholzer-Gee 1997; Laffont and Martimort 2001; Bénabou and Tirole 2003, 2006, 2011). To test the credibility of these channels, we elicited participants' perceptions of the vaccine's safety and their prosocial attitudes toward vaccination. Participants who were randomized to the incentive arm were 9 pp (21% of the mean) more likely to believe that vaccines are unsafe, and 10 pp (20% of the mean) more likely to believe that people do not have a responsibility to get vaccinated, compared to participants in the control arm. Our mediation analysis implies that one-quarter of vaccine incentives' adverse effects can be explained by the negative shifts in individuals' perceptions of vaccine safety and prosocial attitudes toward vaccination.

Our study contributes to three important strands of literature. First, this study advances our understanding of the determinants of vaccine demand, an important topic for policymakers working to minimize morbidity related to COVID-19 and other infectious diseases. Other studies have attempted to evaluate the efficacy of vaccine payments in the real world, including an array of randomized and natural experiments analyzing demand for COVID-19 vaccines (Campos-Mercade et al. 2021; Chang et al. 2021; Barber and West 2022; Law et al. 2022; Thirumurthy et al. 2022). However, these evaluation studies often come to contradictory conclusions.[3] For example, a modest conditional cash transfer raised COVID-19 vaccination rates in Sweden, but a similarly sized incentive was ineffective in the U.S. When do incentives work and when do they fail? Sensibly, the effect of vaccine payments on vaccine take-up is heterogeneous, and incentives are less effective in populations with a greater share of deterred individuals. We develop a procedure to characterize deterred individuals using data from our survey experiment, which provides policymakers a tool to identify subpopulations that are receptive (or resistant) to vaccine incentives.

Second, this study unifies existing models of intrinsic motivation (Bénabou and Tirole 2003, 2006; Titmuss 1970; Laffont and Martimort 2001; Gneezy et al. 2011; Bowles and Polania-Reyes 2012) with econometric advances in the analysis of treatment-response counterfactuals (Heckman and Pinto 2021; Pinto 2021; Mogstad et al. 2018b). Monetary incentives induce two behavioral effects. The price effect makes the incentivized action more attractive while the psychological effect does the reverse. We formalize a monotonicity condition in order to point identify the share of individuals for whom the psychological effect dominates the price effect. In addition, we develop a mediation model showing that vaccine incentives undermine perceptions of the safety of the vaccine and prosocial attitudes toward vaccination, rationalizing the existence of adverse psychological effects. Contrary to the

[3] Other studies evaluating the efficacy of vaccine incentives include Wigham et al. (2014); Tressler and Bhandari (2019); Mantzari et al. (2015); Banerjee et al. (2010); and Sato and Fintan (2020). One systematic review found incentive-based interventions to be among the least effective at increasing vaccine take-up (Jarrett et al. 2015).

common refrain that cash is always the best incentive, our findings suggest that behaviorally informed strategies could be more effective at increasing vaccination coverage than monetary incentives alone.

Finally, this study highlights an important limitation of using monetary incentives in common experimental research designs. Often, in order to estimate an unconfounded treatment effect, researchers randomly allocate incentives for individuals to accept a treatment, assuming that incentives weakly push individuals toward treatment take-up.[4] This "no defiers" assumption breaks down in our context. Monetary incentives deterred one in seven participants in our experiment from considering getting vaccinated; these participants are disproportionately male and from racial and ethnic minority groups. Two-stage least squares would put negative weight on this subgroup's local average treatment effect (Imbens and Angrist 1994). For settings where treatment effects are heterogeneous and incentives plausibly dissuade individuals from taking up treatment, such as payments to get vaccinated, researchers ought to assume weaker forms of monotonicity (Huber and Mellace 2012; De Chaisemartin 2017) or directly test for the existence of defiers using our procedure and others (Kitagawa 2015; Mourifié and Wan 2017; Kowalski 2020).

2. Research Setting and Data

2.1 Experimental design

Our survey experiment centers around the decision to get vaccinated against COVID-19 during a period of falling demand for vaccines.[5] We recruited participants via Los Angeles

[4]For example, Thornton (2008) randomly allocated cash incentives for rural Malawians to learn their HIV status, using these incentives as an instrument to estimate the impact of knowing one's status.

[5]COVID-19 vaccine take-up rates in the U.S. peaked in April 2021 with over 3 million vaccines administered per day, but plummeted shortly thereafter. Within three months, only one-half million vaccines were administered per day—far below the goals set by the Biden administration. By September 2021 (the end of our study's recruitment period), 72% of adults were partially or fully vaccinated.

County community-based organizations and Qualtrics between May and September 2021. The survey was offered in English, Spanish, Korean, Tagalog, and Vietnamese. After obtaining informed consent, participants responded to multiple questions to determine their eligibility based on place of residence (Los Angeles County), age (18 or older), and vaccination status (not yet vaccinated). In the baseline portion of the survey, participants reported their demographics, self-perceived health risks, attitudes toward health care professionals, and their intention to get vaccinated against COVID-19.

Participants who reported "no" or "not sure" for their intention to get vaccinated entered the experiment. We assigned them to one of two treatment arms using 1:1 simple randomization. Individuals in the control arm were asked: "Some officials have proposed letting any person get vaccinated for the coronavirus at no cost. In this scenario, would you consider getting vaccinated this week at no cost?"[6] Although COVID-19 vaccines had been available for all adults at no cost, we framed the control-arm question this way to preemptively correct misconceptions about having to pay for the vaccine.[7]

Individuals who were randomized to the incentive arm were sequentially asked four questions related to their willingness to get vaccinated:

1. "Some officials have proposed paying people to get vaccinated for the coronavirus. In this scenario, would you consider getting vaccinated this week if you were offered $500?"
2. "Would you consider getting vaccinated this week if you were offered $100?"
3. "Would you consider getting vaccinated this week if you were offered $50?"
4. "Would you consider getting vaccinated this week if you were not offered any payment?"

Payment amounts associated with these hypothetical incentives were based off real-world

[6] When our survey was launched, three COVID-19 vaccines (Pfizer-BioNTech, Moderna, and J&J Janssen) were approved for emergency use by the U.S. Food and Drug Administration. To minimize cognitive load for participants, we did not to ask participants to differentiate between vaccine brands.

[7] One-quarter of survey respondents expressed concern about having to pay for the COVID-19 vaccine.

proposals from U.S. policymakers and employers (Litan 2020; U.S. Chamber of Commerce 2021), and the phrasing of the questions was modeled off of priming experiments commonly used in psychology research (Vohs et al. 2006). The fourth incentive-arm question was designed to prime participants to think about monetary incentives while eliciting their willingness to get vaccinated when offered no payment.

After participants were primed or not primed to think about monetary incentives, we gauged their perception of the safety of the vaccine ("Do you believe the coronavirus vaccine is safe?") and their prosocial attitudes toward vaccination ("Do you believe people have a responsibility to get vaccinated for the coronavirus?"). The purpose of these questions was to assess participants' intrinsic motivation to get vaccinated, as well as the relationship between their intrinsic motivation and willingness to get vaccinated.

Nearly every respondent (>99%) who made it to the randomization portion of the survey completed the survey. Participants received a $10 gift card for completing the survey, and were invited to participate in a half-hour qualitative interview for an additional $20 gift card. We conducted the semi-structured, qualitative interviews via video call or telephone no later than three months after survey completion. Interviews were recorded, transcribed in English, and thematically coded by trained qualitative analysts. Appendix Section 3.1 describes the qualitative interviews and related analysis.

2.2 Summary statistics and balance

Our sample consists of 513 participants who met the eligibility criteria and completed the survey. We exclude nine (1.7%) respondents who completed the survey in fewer than three minutes. Table 3.1 presents baseline summary statistics.[8] Participants in our experiment were on average 34 years old; 38% identified as Latino, 35% Black, 34% White, and 9%

[8] In Appendix Section 3.2, we compare our sample with a representative set of vaccine-hesitant adults from the California Health Interview Survey.

Asian. Half of the participants reported an annual household income below $40,000. One in eight had ever tested positive for COVID-19, and 15% expected to contract a mild or severe case of COVID-19 within three months.

We document substantial misinformation about vaccine eligibility in our sample of unvaccinated individuals, which mirrors previously observed associations between misinformation and vaccine hesitancy in the population (Lin et al. 2020). When our survey was launched, all adults regardless of comorbidity and immigration status were eligible for the vaccine. However, only 69% of participants correctly believed they were eligible to get vaccinated; 22% were unsure and 8% believed they were ineligible. Having erroneous beliefs about vaccine eligibility is uncorrelated with age, gender, and racial/ethnic background but strongly correlates with income. Individuals whose annual household income was below $40,000 are 12 percentage points less likely to report they were eligible to get vaccinated, compared to those whose household income was $40,000 or greater.

Individuals in our study expressed a multitude of concerns about the COVID-19 vaccine, which are measured using a five-point Likert scale. To avoid question-order bias, the order of vaccine concerns presented to respondents was randomized. The most commonly reported concern was the vaccine's safety: 78% were concerned ("agree" or "strongly agree") about the potential side effects of the vaccine. In addition, many participants reported concerns about being forced to get vaccinated (64%), the effectiveness of the vaccine (58%), and being infected with COVID-19 by the vaccine (48%). Some participants also reported concerns related to access, such as not knowing where to get vaccinated (29%) and having to pay for the vaccine (25%).

Demographic variables (household size, age, gender, race and ethnicity) and other baseline variables are well-balanced between the control arm and incentive arm. Columns (2) and (3) of Table 3.1 present baseline variable means for participants in the control arm and incentive arm, respectively. Column (4) presents individual p-values testing for mean differences between treatment arms. Additionally, the p-value associated with an F-test from a linear regression of treatment arm assignment on all baseline variables is 0.33. Mean differences across treatment arms are not statistically significant, suggesting that randomization was implemented correctly.

3. Identifying Incentives' Adverse Effects

3.1 Econometric model

We develop an econometric model to analyze individuals' willingness to get vaccinated over a range of monetary incentives. Building on Pinto (2021), the observed variables are: (1) treatment assignment $Z \in \{0, 1\}$; (2) incentive $D(Z) \in \{500, 100, 50, 0\}$; and (3) choice $Y(Z, D) \in \{0, 1\}$. In this model, the choice to get vaccinated (or not) depends on the incentive, as well as nonpecuniary motivation effects from being exposed to the incentive.[9] The choice equation is:

$$Y = f_Y(Z, D, \epsilon) \tag{3.1}$$

where ϵ is an unobserved error term satisfying $Z \perp \epsilon$.

The causal effect of a monetary incentive is the aggregate of the price effect and psychological effect (Gneezy et al. 2011). The price effect unambiguously makes getting vaccinated more attractive. The psychological effect pushes behavior in the opposite direction and, when it dominates the price effect, crowds out the incentivized behavior.[10] For illustration, consider the potential choices for a participant who is offered \$500 to get vaccinated. The

[9]Note that treatment assignment (Z) is not an instrumental variable in the typical sense because Z influences vaccination choice (Y) beyond its impact on the incentive (D), thus violating the exclusion restriction assumption. It does, however, satisfy the independence and relevance assumptions of instrumental variables.

[10]Bénabou and Tirole (2003; 2006) model an incentive's adverse psychological effect by separating the extrinsic and intrinsic components of utility.

total effect on vaccine take-up combines the price and psychological effects:

$$\underbrace{Y(1,500) - Y(0,0)}_{\text{total effect}} = \underbrace{Y(1,500) - Y(1,0)}_{\text{price effect}} + \underbrace{Y(1,0) - Y(0,0)}_{\text{psychological effect}} \tag{3.2}$$

Other studies have demonstrated the existence of adverse psychological effects in various settings by estimating negative total effects on the incentivized behavior, inferring that the psychological effect dominates the price effect (Mellström and Johannesson 2008; Gneezy and Rustichini 2000; Frey and Jegen 2001). However, this test is overly conservative. As equation (3.2) shows, a non-negative total effect does not imply that the psychological effect is nil. It implies that, on average, the price effect weakly dominates the psychological effect. Our experiment instead aims to separate the price effect from the psychological effect by priming individuals to think about incentives while eliciting their willingness to get vaccinated when they are offered no payment.[11]

To estimate the share of individuals deterred from getting vaccinated ("deterred share"), we partition individuals according to their response type (Kline and Tartari 2016; Heckman and Pinto 2018). A response type specifies an individual's choice for every possible (Z, D) pair. For example, an always-taker would get vaccinated in either treatment arm for any vaccine payment: $Y(z, d) = 1 \; \forall \; z, d$. Although response types are themselves unobservable, statistical moments derived from our experiment can be expressed as a mixture of response types. In addition, many policy-relevant parameters can be expressed as a mixture of response types. Consider the following response matrix, which stacks response types s_i such that each element indicates one of seven choice behaviors t_j:

[11] By priming individuals to think about monetary incentives, the experiment naturally introduces anchoring bias, whereby valuations heavily factor in the initial payment (Tversky and Kahneman 1974; Lieder et al. 2018). Estimates from our experiment should be contextualized with \$500 as the "anchoring" value. We chose not to experimentally vary the anchoring value because our sample size was not powered to jointly test for the priming effect and anchoring effect.

$$\begin{array}{c} \\ Z_0 \\ Z_1 \end{array} \begin{array}{c} \begin{array}{cccccccccc} s_1 & s_2 & s_3 & s_4 & s_5 & s_6 & s_7 & s_8 & s_9 & s_{10} \end{array} \\ \begin{bmatrix} t_1 & t_1 & t_1 & t_1 & t_1 & t_2 & t_2 & t_2 & t_2 & t_2 \\ t_3 & t_4 & t_5 & t_6 & t_7 & t_3 & t_4 & t_5 & t_6 & t_7 \end{bmatrix} \end{array} \tag{3.3}$$

In the control arm (Z_0), an individual with choice behavior t_1 vaccinates and t_2 does not vaccinate. In the incentive arm (Z_1), an individual with choice behavior t_3 vaccinates for any payment including zero, t_4 vaccinates for no less than \$50, t_5 vaccinates for no less than \$100, t_6 vaccinates for no less than \$500, and t_7 does not vaccinate for any payment.[12] In this framework, response type $s_1 \equiv (t_1, t_3)$ denotes the previously described always-takers. Response types $\{s_2, s_3, s_4, s_5\}$ denote deterred individuals: those who consider getting vaccinated without any payment (t_1) but would reject the vaccine if offered an incentive (t_4, t_5, t_6, t_7). Response types $\{s_7, s_8, s_9\}$ denote induced individuals: those who turn down the vaccine without any payment (t_2) but would consider the vaccine for some positive payment (t_4, t_5, t_6). Lastly, response type $s_{10} \equiv (t_2, t_7)$ denotes never-takers.

Finally, we formalize a monotonicity condition to nonparametrically estimate the always-taker share and deterred share. This monotonicity condition eliminates response type $s_6 \equiv (t_2, t_3)$ in the same way that Imbens and Angrist (1994) ruled out defiers to estimate the local average treatment effect (LATE). In other words, our monotonicity condition assumes that no individual gets induced into vaccination by a zero-payment incentive. This monotonicity condition is rooted in two decades' worth of behavioral research showing that, netting out the price effect, extrinsic incentives unambiguously weaken free-choice intrinsic motivation (Deci et al. 1999; Akerlof and Kranton 2010).

How does this monotonicity condition point identify the always-taker share and deterred share? Our identification argument goes as follows. The always-taker plus deterred set

[12]To make this model tractable, we assume that individuals are "money rational". For instance, nobody in the incentive arm rejecting the vaccine for \$500 would subsequently accept it for \$100. This assumption shrinks the number of response types from 64 to 10.

$\{s_1, s_2, s_3, s_4, s_5\}$ is identified by participants in the control arm choosing to get vaccinated. Since treatment assignment Z is random, the always-taker plus deterred share is equivalent between the control arm and incentive arm. By the monotonicity condition, we rule out vaccinations induced by zero-payment incentives $\{s_6\}$, so the always-taker set $\{s_1\}$ is identified by participants in the incentive arm choosing to get vaccinated when offered zero payment. Finally, subtracting the always-taker set from the always-taker plus deterred set yields response types $\{s_2, s_3, s_4, s_5\}$, which constitute deterred individuals.

3.2 Estimated impact on vaccine willingness

We estimate the causal effect of monetary incentives on individuals' willingness to get vaccinated against COVID-19. Figure 3.1 presents our main results. Estimates of differences between the control arm and incentive arm are presented with robust standard errors from a linear regression of vaccination willingness on the treatment indicator.

In the control arm, 25% of participants reported a willingness to get vaccinated without any payment. In the incentive arm, participants' willingness to get vaccinated varied widely by the payment amount: 43% for \$500, 19% for \$100, 13% for \$50, and 11% for zero payment. Section 3.1 outlines the relationship between these statistical moments and the response types characterizing always-takers, deterred individuals, induced individuals, and never-takers. For example, participants in the control arm who were willing to get vaccinated (25%) must be one of five response types $\{s_1, s_2, s_3, s_4, s_5\}$.

We decompose into two groups the response-type mixture representing the quarter of participants willing to get vaccinated in the control arm. Recall that these participants must be either always-takers $\{s_1\}$ or deterred individuals $\{s_2, s_3, s_4, s_5\}$. By our identification argument, one in ten (10.5%) participants are always-takers: they chose to get vaccinated in the incentive arm for zero payment. We thus deduce that one in seven (14.4%) participants are deterred by monetary incentives: they considered getting vaccinated without any payment but would have rejected the vaccine had they been offered a non-zero incentive.

While the always-taker share is point identified, the never-taker share $\{s_{10}\}$ is not identified. Participants who rejected the vaccine for $500 are not necessarily never-takers because the incentive's adverse psychological effect may have dominated the price effect. As a result, the induced share $\{s_7, s_8, s_9\}$ is also not identified. These probabilities are partially identified, however, because response-type probabilities must sum to unity (Heckman and Pinto 2018). Following Honoré and Lleras-Muney (2006), we algorithmically estimate sharp bounds for these parameters by treating identified response-type mixtures as constraints in a linear programming problem. By construction, these sharp bounds include every value that cannot be rejected as the true estimand, exhausting all possible information generated by the experiment and our monotonicity condition. Estimated bounds for the never-taker share are [43%, 57%]; bounds for the induced share are [18%, 32%].

3.3 Characterizing deterred individuals

Who gets deterred by monetary incentives? In theory, how an individual responds to vaccine incentives depends on his or her existing trust in medical institutions, for which there are well-documented disparities by race and gender (Gee and Ford 2011; Thompson et al. 2021). Groups that disproportionately experience abuse from health care providers may be more skeptical of public health interventions, such as vaccine incentives. For example, the Tuskegee experiment, which was unethically conducted without the informed consent of its participants, has been linked to long-term depressed demand for health care services among Black men (Alsan and Wanamaker 2018). To assess whether adverse psychological effects differ across participants in our experiment, we use baseline survey data to characterize participants who were deterred from vaccination. Our approach follows Abadie (2003), who proposed a nonparametric κ-weighting scheme to estimate statistics for the complier population in the LATE model. Likewise, we estimate statistics for the always-taker population and deterred population.

Table 3.2 reports mean baseline characteristics for deterred individuals and always-takers, who comprise 14% and 11% of our sample, respectively. Column (3) reports individual p-values testing for mean differences between both groups. We detect substantial heterogeneity in adverse psychological effects by race and gender. Non-white men, notably, make up one-third of the deterred population but only 4% of always-taker population. We do not observe disparities of this magnitude in any other demographic group, including groups stratified by household size and age. In this way, race and gender seem to be uniquely salient factors affecting how individuals perceive and respond to vaccine incentives.

In addition, baseline concerns about the vaccine's side effects vary substantially between the deterred and always-taker populations. Deterred individuals expressed significantly higher levels of distrust in the safety of the vaccine compared to always-takers: 4.4 versus 3.5 on the Likert (1–5) scale. There are no significant gaps between the two groups for other concerns, including concerns about the vaccine's efficacy and accessibility.

4. Evaluating Behavioral Mechanisms

4.1 Mediation model

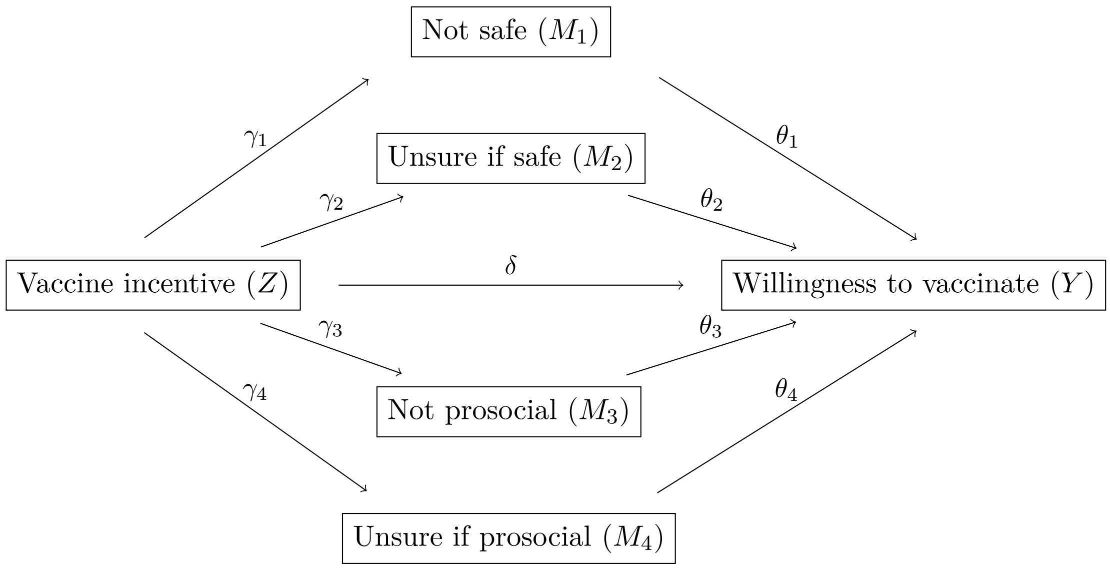
What are the channels through which vaccine incentives discourage individuals from getting vaccinated? We evaluate two mechanisms that could underlie incentives' adverse psychological effects. First, incentives could signal to participants that vaccines are unsafe or cause unwanted side effects ("risk signaling hypothesis"). If high prices convey high quality, then inversely, negative prices may convey elevated risk. For example, clinical studies that offer large participation payments are perceived to be riskier than studies offering no payment (Cryder et al. 2010).

Second, payments could crowd out intrinsic motivation by raising doubt about the prosocial nature of vaccination ("prosocial motivation hypothesis"). Efforts to encourage vaccination often appeal to people's sense of social responsibility because getting vaccinated also prevents the spread of infection to others. However, participants who believe vaccination to be altruistic may no longer hold this belief if they are incentivized to get vaccinated, since pecuniary and prosocial motives cannot coexist in many settings (Frey and Jegen 2001; Akerlof and Kranton 2010).

We develop a mediation model to jointly test the risk signaling and prosocial motivation hypotheses.[13] Because our experiment generates exogenous treatment variation, we can estimate the causal effect of vaccine incentives on participants' beliefs about the COVID-19 vaccine. In the post-randomization portion of the survey, participants reported their perceptions of vaccine safety and their prosocial attitudes toward vaccination: "yes", "no", or "not sure". We specify four mediators based on their responses to these questions. Mediator M_1 captures shifts in beliefs that the vaccine is unsafe; M_2 captures shifts in uncertainty about the vaccine's safety; M_3 captures shifts in beliefs that vaccination is not a responsibility; and M_4 captures shifts in uncertainty about whether vaccination is a responsibility.

For variable M_i to mediate the causal effect of vaccine incentives on an individual's willingness to get vaccinated, the variable must itself be affected by the payment (γ_i) and indirectly influence one's willingness to get vaccinated (θ_i). This causal mediation model is depicted by relationship (3.4), which we operationalize using structural equation modeling (Kline 2015). Using this mediation model, we decompose the adverse psychological effect of vaccine incentives into the direct effect (δ) and indirect effects ($\gamma_i \times \theta_i$). Indirect effects measure the fraction of the psychological effect that can be explained by shifts in vaccine risk perceptions and prosocial attitudes toward vaccination.

[13] Appendix Section 3.1 also provides evidence from qualitative interviews corroborating the risk signaling and prosocial motivation hypotheses.



(3.4)

4.2 Estimated impact on vaccine perceptions

Do incentives harm individuals' perceptions of vaccine safety and/or prosocial attitudes toward vaccination? Figure 3.2 presents the results of our mechanisms analysis. Participants randomized to the incentive arm are 9 percentage points more likely to believe that the vaccine is unsafe compared to those in the control arm (50% vs 41%). We observe a gap of the same magnitude but opposite sign in the proportion of participants expressing being "unsure" about the safety of the vaccine (38% vs 47%), suggesting that vaccine incentives only affected participants with uncertain beliefs about the vaccine's safety. In contrast, participants in the incentive arm are no more or less likely to believe that the vaccine is safe compared to participants in the control arm.

Participants randomized to the incentive arm are 10 percentage points more likely to report that people do *not* have any "responsibility to get vaccinated" compared to those in the control arm (59% vs 49%). Two-third of this effect comes from a drop in the share of participants who believe people have a responsibility to get vaccinated (18% vs 25%). The remainder comes from a drop in the share of participants expressing uncertainty about whether people have a responsibility to get vaccinated (23% vs 26%).

Priming individuals to think about monetary incentives harms their perceptions of the vaccine's safety and their prosocial attitudes toward vaccination. Do these shifts in beliefs rationalize the adverse psychological effects estimated in Section 3.2? We develop a mediation model to quantify the impact of these mechanisms on individuals' willingness to get vaccinated. To estimate the model's parameters, we regress willingness to get vaccinated (Y) on the set of mediators (M_i) and the treatment indicator (Z), and regress each mediator on the treatment indicator, accounting for correlation between error terms across equations. Table 3.3 presents the regression coefficients, corresponding to θ_i, δ, and γ_i in relationship (3.4).

This mediation model allows us to decompose the adverse psychological effects of vaccine incentives into the direct effect (δ) and indirect effects ($\gamma_i \times \theta_i$). The latter represents the portion of adverse effects that can be explained by shifts in participants' perceptions of vaccine safety and prosocial attitudes toward vaccination. These mediators explain approximately one quarter (27%) of vaccine incentives' adverse effects. This finding highlights how beliefs about the COVID-19 vaccine are strongly linked to adverse psychological effects, lending credence to the risk-signaling and prosocial motivation hypotheses.

5. Discussion

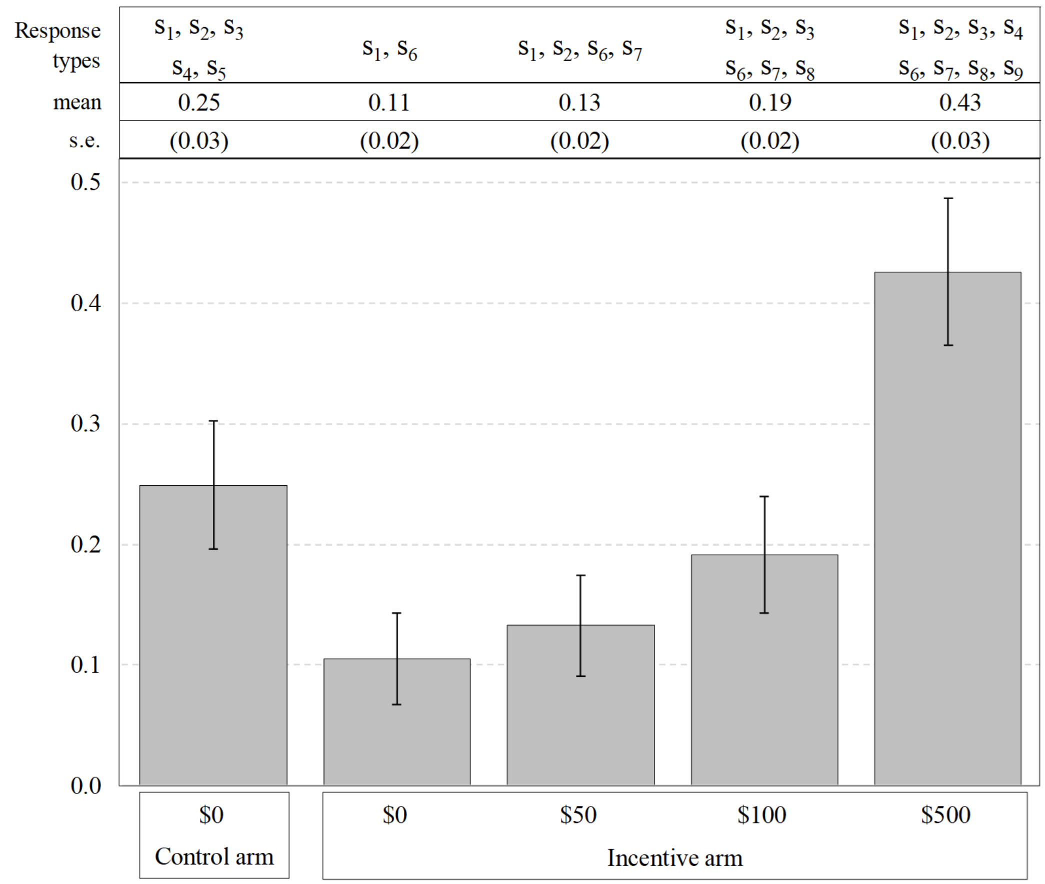
This paper demonstrates that monetary incentives can meaningfully crowd out individuals' motivation to get vaccinated against COVID-19. Using a novel survey experiment, we find that vaccine incentives deterred 14% of participants from considering getting vaccinated. This effect is both statistically and economically significant, as only one in four participants had considered getting vaccinated at the outset. Deterred individuals are more likely to be male, come from racial and ethnic minority groups, and report concerns about the vaccine's side effects. We also develop a mediation model showing how vaccine incentives harm individuals' perceptions of vaccine safety and prosocial attitudes toward vaccination.

Our findings shed new light on an important consequence of racialized medical mistrust. Black men are among the least likely to trust the safety of the COVID-19 vaccine and get vaccinated—a likely rational response to experiences of neglect and abuse from the U.S. health care system (Stoler et al. 2021; Thompson et al. 2021). We also show that this group is, by far, the least receptive to vaccine incentives. Monetary incentives aimed at encouraging vaccination may exacerbate existing racial disparities in COVID-19 vaccination coverage and related morbidity.

In the short run, the efficacy of a monetary incentive depends on how many people it would persuade or dissuade. This logic can explain why vaccine incentives "succeed" in some settings (Campos-Mercade et al. 2021) but "fail" in others (Chang et al. 2021). In the long run, however, vaccine incentives can erode trust in public health institutions and undermine their efforts. We find that exposure to the idea of monetary incentives reduces the proportion of participants believing that vaccination is a social responsibility. Pandemic mitigation strategies—vaccination, mask mandates, and social distancing—are only effective if people adhere to them and assume responsibility in controlling the spread of the disease. Monetary incentives may backfire in this regard.

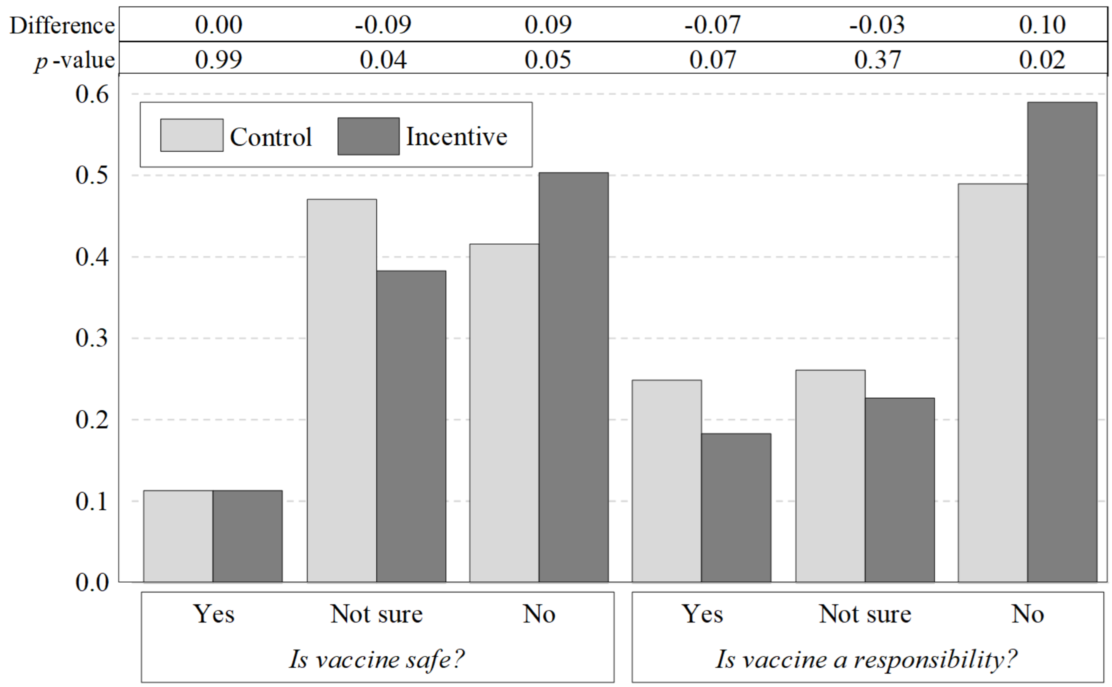
6. Figures

Figure 3.1: Vaccine Willingness by Treatment Arm and Incentive Amount



Notes: This figure presents estimates of individuals' willingness to get vaccinated against COVID-19, stratified by treatment assignment and incentive amount. 95-percent confidence intervals are presented in the bar chart. Mean values and robust standard errors are presented above the figure, as well as response type mixtures associated with each estimate.

Figure 3.2: Beliefs about Vaccination by Treatment Arm



Notes: This figure presents estimates of individuals' perceptions about vaccine safety ("Do you believe the coronavirus vaccine is safe?") and prosocial attitudes toward vaccination ("Do you believe people have a responsibility to get vaccinated for the coronavirus?"), stratified by treatment arm assignment. Percentage-point differences between the incentive arm and control arm are shown alongside two-sided, single-hypothesis p-values associated with equality in estimates between both groups.

7. Tables

Table 3.1: Baseline Summary Statistics by Treatment Arm

	(1)	(2)	(3)	(4)
	Full sample	Control arm	Incentive arm	p-value
Household size	3.4	3.6	3.3	0.13
Age	33.9	33.6	34.2	0.61
White male	0.13	0.14	0.12	0.52
White female	0.21	0.20	0.23	0.51
Non-white male	0.21	0.20	0.23	0.34
Non-white female	0.43	0.44	0.42	0.68
College graduate	0.32	0.30	0.33	0.55
Household income				
Less than $20,000	0.26	0.22	0.30	0.04
$20,000 to $39,999	0.23	0.26	0.21	0.15
$40,000 to $59,999	0.17	0.20	0.15	0.21
$60,000 to $79,999	0.11	0.11	0.10	0.58
$80,000 to $99,999	0.09	0.09	0.09	0.89
Greater than $100,000	0.14	0.12	0.16	0.19
Ever tested positive for COVID	0.12	0.13	0.11	0.50
Best guess for COVID risk				
None	0.85	0.83	0.86	0.28
Mild case	0.12	0.14	0.09	0.10
Serious case	0.04	0.03	0.04	0.48
Vaccine concerns (1–5 scale)				
Side effects	4.1	4.1	4.2	0.70
Not effective	3.7	3.8	3.6	0.24
Causes infection	3.4	3.5	3.3	0.32
Affordability	2.6	2.7	2.5	0.22
Compulsion by law	3.8	3.8	3.8	0.86
Accessibility	2.7	2.8	2.7	0.33
Compulsion by work/school	3.7	3.8	3.7	0.20
Observations	513	257	256	

Notes: This table presents means for baseline variables, which were measured before participants were randomly assigned to their treatment arm. Variable descriptions can be found in the survey text. Column (1) presents variable means for the full sample. Columns (2) and (3) present variable means for participants in the control arm and incentive arm, respectively. Column (4) shows two-sided, single-hypothesis p-values associated with equality in means between the control arm and incentive arm.

Table 3.2: Demographics and Vaccine Concerns by Treatment Response Type

	(1)	(2)	(3)
	Deterred	Always-taker	p-value
A. Participant demographics			
Household size	4.2 (0.5)	3.9 (0.3)	0.66
Age	28.8 (2.8)	33.4 (2.5)	0.34
White male	0.11 (0.09)	0.19 (0.07)	0.61
White female	0.16 (0.11)	0.30 (0.09)	0.45
Non-white male	0.33 (0.10)	0.04 (0.04)	0.01
Non-white female	0.38 (0.13)	0.48 (0.10)	0.60
B. Vaccine concerns (1–5 scale)			
Side effects	4.4 (0.3)	3.5 (0.2)	0.04
Not effective	3.7 (0.3)	3.5 (0.2)	0.60
Causes infection	3.2 (0.3)	3.6 (0.2)	0.36
Affordability	2.7 (0.4)	3.1 (0.2)	0.46
Compulsion by law	3.6 (0.3)	3.3 (0.2)	0.49
Accessibility	3.1 (0.3)	3.2 (0.2)	0.79
Compulsion by work/school	3.6 (0.3)	3.5 (0.2)	0.78
Observations	74	54	

Notes: This table presents baseline variable means stratified by response-type mixtures. Panel A reports participants' demographics. Panel B reports vaccine concerns measured using the Likert (1–5) scale. Column (1) presents means and standard errors for deterred individuals $\{s_2, s_3, s_4, s_5\}$. Column (2) presents means and standard errors for always-takers $\{s_1\}$. Column (3) shows the two-sided, single-hypothesis p-value associated with equality in means between deterred individuals and always-takers.

Table 3.3: Coefficients from the Causal Mediation Model

	Dependent variable				
	Y	M_1	M_2	M_3	M_4
Vaccine incentive (Z)	-0.11	0.09	-0.09	0.10	-0.03
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
Not safe (M_1)	-0.41				
	(0.07)				
Unsure if safe (M_2)	-0.31				
	(0.07)				
Not prosocial (M_3)	-0.29				
	(0.05)				
Unsure if prosocial (M_4)	-0.21				
	(0.06)				
R^2	0.31	0.01	0.01	0.01	0.00
Observations	513	515	513	513	513

Notes: This table presents regression estimates associated with the causal mediation model described by relationship (3.4). Variable Y equals one for participants who considered getting vaccinated against COVID-19 for no payment. Variable Z equals one for participants who were randomized to the incentive arm of the experiment. Mediators $\{M_1, M_2, M_3, M_4\}$ indicate participants' perceptions of vaccine safety and their prosocial attitudes toward vaccination. Robust standard errors are estimated simultaneously, and are shown in parentheses.

Chapter 4

Appendix and Supplementary Material

1. Appendix to "Do Cash Transfers Narrow Health Disparities? Evidence from Veterans with Disabilities"

1.1 Institutional background

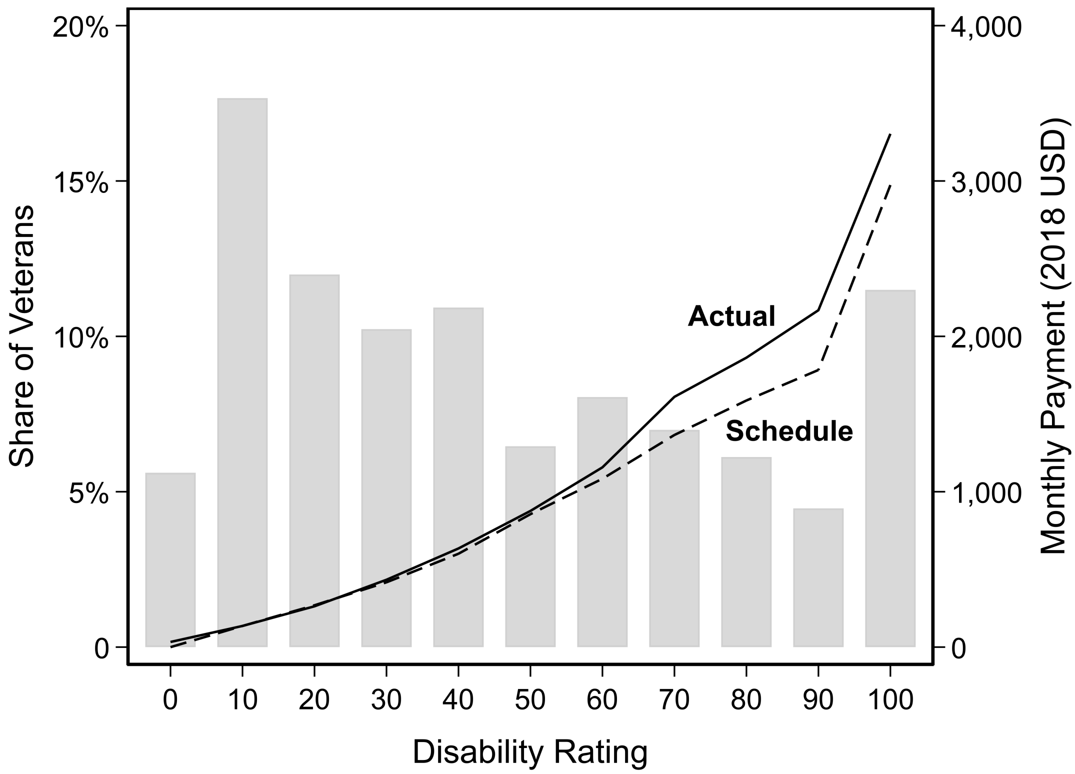
Submitting and adjudicating a disability compensation claim with the VA is a multi-step process that aims to provide benefits to veterans with service-connected disabilities. The first step in this process is for the veteran to submit a claim using VA Form 21-526EZ. This can be done online, by mail, or in person at a VA office. The claim must include details of the impairment in question, and whenever possible, supporting evidence like service treatment records, service personnel records, and private medical records should be submitted alongside the claim.

The disability claim is then assigned to a Veterans Service Representative (VSR), or caseworker, who reviews the claim to determine if additional evidence is needed and requests any missing information from the veteran, service department, or other authoritative sources. During this step, the caseworker orders a VA claim exam, also known as a Compensation & Pension (C&P) examination. This exam helps the VA determine the current severity of a veteran's condition and establish the nexus, or link, between the veteran's current disability and their service. A VA or VA-contracted physician conducts the exam, focusing on the specific areas related to the claimed conditions.

Following the caseworker's and physician's reviews, the disability claim is passed on to a Rating Veterans Service Representative (RVSR), or adjudicator. The adjudicator has the authority to evaluate the claim and make a decision on the level of service-connected disability. They review all the evidence in the veteran's file, including the caseworker's findings and the physician's exam results, and apply the VA's Schedule for Rating Disabilities

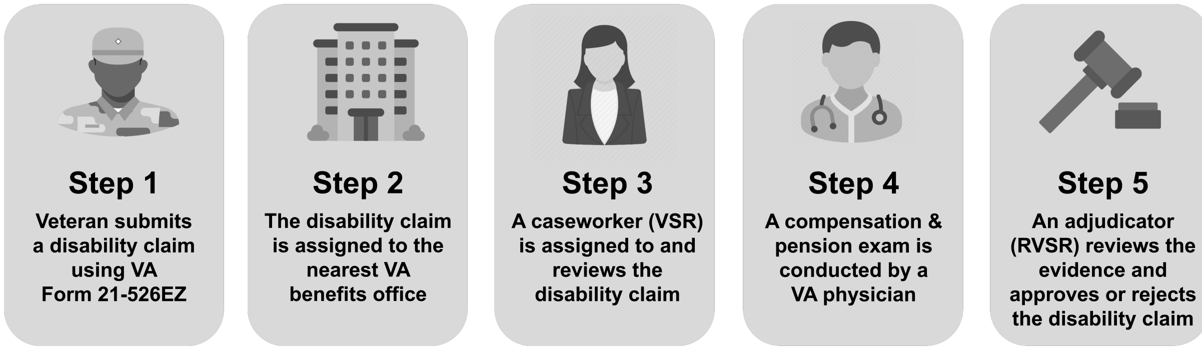
(VASRD) to determine the appropriate disability rating. This rating is crucial because it affects the amount of compensation the veteran will receive. Once the adjudicator makes a decision, the VA sends a written notification to the veteran, outlining the decision, the rating assigned, and the compensation amount, if applicable. If the claim is denied, the notification will include an explanation of why the claim was not approved and information on how to appeal the decision.

Figure 4.1: Histogram of Disability Ratings and 2018 Payment Schedule



Notes: This figure displays the distribution of service-connected disability ratings for veterans in our sample, ranging between 0% to 100%. The solid line represents the VA's 2018 payment schedule for a single veteran with no dependents, as a function of their disability rating. The dashed line represents the actual payment (taking into account additional payments from dependents and VA's Individual Unemployability rules) for veterans in 2018.

Figure 4.2: Multistep Process for Submitting a Disability Compensation Claim



Notes: This figure outlines the multi-step process for submitting and adjudicating a disability compensation claim with the VA, which is described in greater detail below.

1.2 Data and sample

Table 4.1: Most Prevalent Impairments in the VA's Disability Compensation Program in FY 2018

Service-connected disabilities	**Prevalence**
Tinnitus	1,971,201
Hearing loss	1,228,936
Post-traumatic stress disorder	1,039,794
Scars, general	1,036,677
Limitation of flexion, knee	1,021,281
Lumbosacral or cervical strain	989,835
Paralysis of the sciatic nerve	781,178
Limitation of motion of the ankle	636,853
Migraine	548,999
Degenerative Arthritis of the Spine	769,384
Total disabilities	25,127,129
Total veterans with disabilities	4,743,108

Notes: This table presents the most prevalent service-connected disabilities among all VA disability compensation recipients as of fiscal year 2018. The data presented in this table reflect the disabilities that have been formally recognized and approved by the Veterans Benefits Administration as being connected to military service. The total numbers do not include disabilities that are determined not to be related to military service. Source: Veterans Benefits Administration Annual Benefits Report, Fiscal Year 2018.

Table 4.2: VA Stop Codes to Define Outpatient Utilization

Category of Care	**Stop Codes**
Primary care	301,322,323,350
Emergency department	130
Mental health care	502,503,505,506,509,510,516,524,525 532,535,540,550,552,553,554,557,558 559,561,562,567,568,571,572,573,574 575,576,577,580,582,583,588
Homelessness services	504,507,508,511,522,528,529,530,555 556,590

Notes: The VA healthcare system employs a unique coding system known as VA Stop Codes to classify and track outpatient service utilization. These stop codes are specific codes assigned to every type of outpatient service provided within the VA system, ranging from primary care visits to specialized services. We use the following VA stop codes to categorize and analyze the utilization patterns of outpatient services.

1.3 Defining medical acuity

The International Classification of Diseases (ICD), Ninth and Tenth Revisions, are systems of medical classification used for coding diagnoses in conjunction with clinical care. The Elixhauser Comorbidity Index is a method for assessing patient comorbidities, which are medical conditions co-occurring with the primary impairment (Elixhauser et al. 1998). It includes an array of 31 comorbidities, which may or may not be directly related to the primary condition. The Elixhauser Comorbidity Index is widely used in health services research due to its comprehensive approach and effectiveness in predicting hospital and mortality outcomes.

To construct the medical acuity variable, we use logistic regression to predict the probability of a 5-year mortality based on veterans' Elixhauser comorbidities. Each of the 31 comorbidities is used as an independent variable in the regression model. The logistic regression model provides a probability score between 0 and 1 for each veteran, which we subsequently standardize to have a mean of zero and a standard deviation of one.

Table 4.3: Elixhauser Comorbidity Statistics

	(1) Analytic sample	(2) Non-disabled veterans
Congestive heart failure	2.8	1.5
Cardiac arrhythmia	5.7	4.3
Valvular disease	1.6	1.0
Pulmonary circulation disorder	0.5	0.3
Peripheral vascular disease	3.5	2.1
Hypertension	34.7	27.1
Hypertension, complicated	1.3	0.8
Paralysis	0.5	0.2
Neurological disorders	2.4	1.4
Chronic pulmonary disease	9.6	6.6
Diabetes mellitus	16.6	9.1
Diabetes mellitus, complicated	5.0	2.3
Hypothyroidism	3.4	2.6
Renal failure	2.3	1.3
Liver disease	2.2	1.8
Peptic ulcer disease	0.9	0.6
AIDS/HIV	0.2	0.2
Lymphoma	0.4	0.3
Metastatic cancer	0.3	0.2
Solid tumor without metastasis	5.0	3.3
Rheumatoid arthritis	1.3	0.9
Coagulopathy	0.9	0.6
Obesity	10.6	8.4
Weight loss	0.9	0.6
Fluid and electrolyte disorders	2.0	1.2
Blood loss anemia	0.1	0.1
Anemia deficiency	1.2	0.8
Alcohol disorder use	6.8	5.5
Drug disorder use	4.4	3.7
Psychoses	2.7	1.2
Depression	23.1	12.1

Notes: This table provides summary statistics for all 31 Elixhauser comorbidities measured 1 year before the rating process. Columns (1) and (2) display summary statistics (percentages) for veterans in the analytic sample (who submitted a claim for disability compensation) and for non-disabled veterans (those never receiving disability compensation), respectively.

Table 4.4: Characteristics of the Sample, Stratified by Above or Below 50% Disability Rating

	(1) Rating $< 50\%$	(2) Rating $\geq 50\%$
Demographics		
Age	50.8	49.4
Male	91.7	89.4
White	76.1	73.0
Black	14.8	16.9
Hispanic	6.9	8.2
Asian/Pacific Islander	3.7	4.2
Prior-year utilization		
Outpatient	48.6	55.8
Mental health services	7.6	15.4
Inpatient	3.0	5.2
Emergency room	3.2	4.3
Homelessness services	1.2	1.2
Comorbidites		
Hypertension	33.6	36.0
Depression	17.0	30.0
Diabetes mellitus	15.0	18.3
Obesity	10.1	11.1
COPD	8.8	10.7
Alcohol use disorder	6.4	7.4
Mortality rate		
1 year	1.1	1.7
5 year	7.0	9.4
10 year	15.5	21.6
Individuals (N)	458,335	353,946

Notes: This table provides summary statistics for individual demographics, healthcare utilization (any use in the year before the disability rating process), comorbidities, and mortality rates (measured 1, 5, and 10 years after the rating process). For year 10, mortality is analyzed using the subsample for which there is an adequate time horizon to assess mortality within that period ($N = 646,079$). Columns (1) and (2) display summary statistics for veterans in the analytic sample (who submitted a claim for disability compensation, as described in Section 3.2) who subsequently received a disability rating below and above/equal to 50%, respectively. COPD = Chronic obstructive pulmonary disease.

1.4 Empirical strategy

The VA offers several non-compensation benefits to service-disabled veterans, which are managed by the Veterans Benefit Administration. These benefits can be categorized into three major groups: Vocational Rehabilitation and Employment, Housing Grants and Benefits, and Other Grants and Benefits. Vocational Rehabilitation and Employment benefits assist veterans in gaining and maintaining suitable employment, which the VA considers to be an essential aspect of reintegrating into civilian life. Housing Grants and Benefits offer financial support to modify homes, ensuring they are accessible and comfortable for veterans with service-related disabilities. They are made up of two programs: Specially Adapted Housing and Special Housing Adaptation. Other Grants and Benefits encompass a variety of programs designed for specific needs, such as automobile grants, adaptive equipment grants, clothing allowance grants, and life insurance, which cater to various aspects of daily living affected by service-related disabilities.

Table 4.5: Other VA Benefits for Veterans with Disabilities

	FY2019 funding ($ in millions)	FY2019 participants
Disability Compensation	87,647	4,944,275
Vocational Rehabilitation & Employment	1,388	122,249
Housing Grants and Benefits		
Specially Adapted Housing	118	2,055
Special Housing Adaptation	3	215
Other Grants and Benefits		
Automobile grants	40	1,966
Adaptive equipment grants	97	7,039
Clothing allowance grant	131	160,774
Service-disabled life insurance	2,934	279,112

Source: Congressional Research Service (CRS) compilation of data from Department of Veterans Affairs (VA) FY2021 Volume III budget justification.

Table 4.6: First-Stage Estimates for Subgroups Based on Observables

	N	**Caseworker**	**Physician**
18-34 White men	109,556	0.59	0.52
35-49 White men	91,742	0.63	0.53
50-64 White men	172,472	0.35	0.37
65+ White men	121,592	0.27	0.31
18-34 Black men	19,259	0.61	0.50
35-49 Black men	31,081	0.51	0.46
50-64 Black men	33,278	0.34	0.44
65+ Black men	9,949	0.15	0.33
18-34 White women	22,956	0.62	0.48
35-49 White women	12,738	0.54	0.49
50-64 White women	5,311	0.57	0.56
65+ White women	677	0.75	0.37
18-34 Black women	9,277	0.53	0.56
35-49 Black women	7,762	0.53	0.40
50-64 Black women	2,222	0.58	0.44
65+ Black women	78	5.6	-3.4

Notes: This table displays the first-stage coefficients from equation (1.2) estimated for subgroups defined by age, sex, and Black/White race. The first column shows the first-stage coefficients for the caseworker instrument, and the second column shows the first-stage coefficients for the physician instrument. In every specification, we reject the hypothesis that the coefficients are negative at the 5% level.

1.5 Additional results

This section describes the empirical Bayes procedure used to shrink our causal estimates of VA disability compensation. For impairment type or commuting zone k, the effect of \$300 per month on 5-year mortality is β_k, whose distribution is the true underlying distribution of causal effects:

$$\hat{\beta}_k = \beta_k + \nu_k$$

where ν_k is an error term. The aim of our empirical Bayes procedure is to "shrink" individual estimates towards a group average, based on the assumption that all individual estimates are drawn from a common distribution. This shrinkage reduces the sampling error of the estimates, which results in more stable and reliable parameter estimates. We assume:

$$\hat{\beta}_k | (\beta_k, \pi_k^2) \sim N(\beta_k, \pi_k^2)$$

where π_k^2 is the variance of the error term. The prior distribution of the true causal parameter is given by:

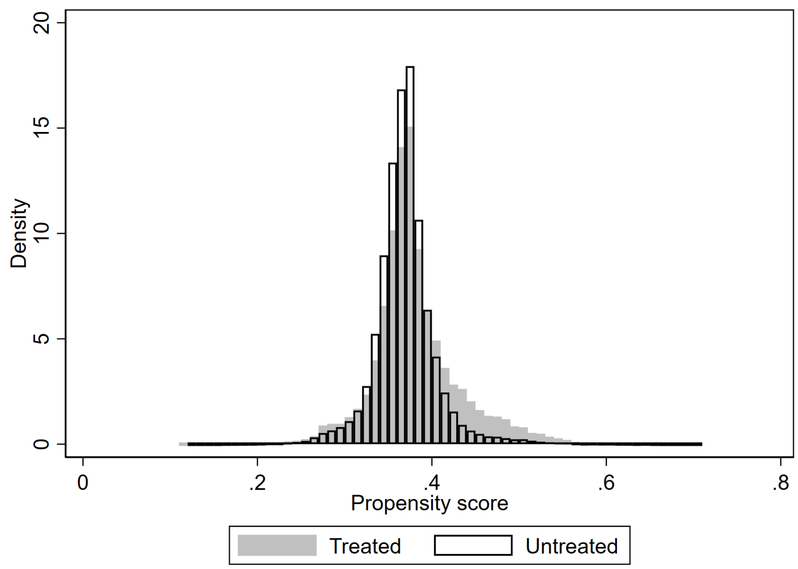
$$\beta_k | (X_k, \lambda, \sigma^2) \sim N(X_k \lambda, \sigma^2)$$

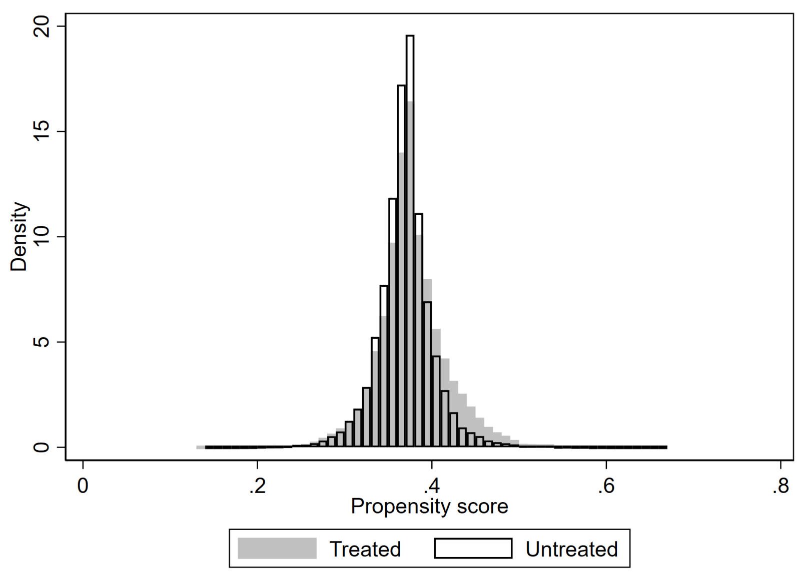
where X_k are observables characteristics and σ^2 is the variance of the causal parameters. Conditioning on observable $\hat{\beta}_k$, the posterior distribution of β_k is given by:

$$\beta_k | (\hat{\beta}_k, X_k, \lambda, \sigma^2, \pi_k^2) \sim N(\beta_k^{EB}, \pi_k^2(1 - B_k))$$

where $\beta_k^{EB} = (1 - B_k)\hat{\beta}_k + B_k X_k \lambda$ and $B_k = \frac{\pi_k^2}{\pi_k^2 + \sigma^2}$. We implement this procedure using the method outlined by Morris (1983) and Chandra et al. (2016).

Figure 4.3: Distribution of Propensity Scores for the Caseworker and Physician Instruments





Notes: Panel A (top) depicts the distribution of propensity scores associated with the caseworker leniency instrument, calculated using the local instrumental variables estimator in equation (1.9). The treatment variable was dichotomized (Section 7.2), to be either above or below $1000 in 2018 dollars. Panel B (bottom) depicts the distribution of propensity scores associated with the physician leniency instrument.

2. Appendix to “Inappropriate Prescribing to Older Adults by Nurse Practitioners and Primary Care Physicians”

2.1 Data linkage and construction of study cohort

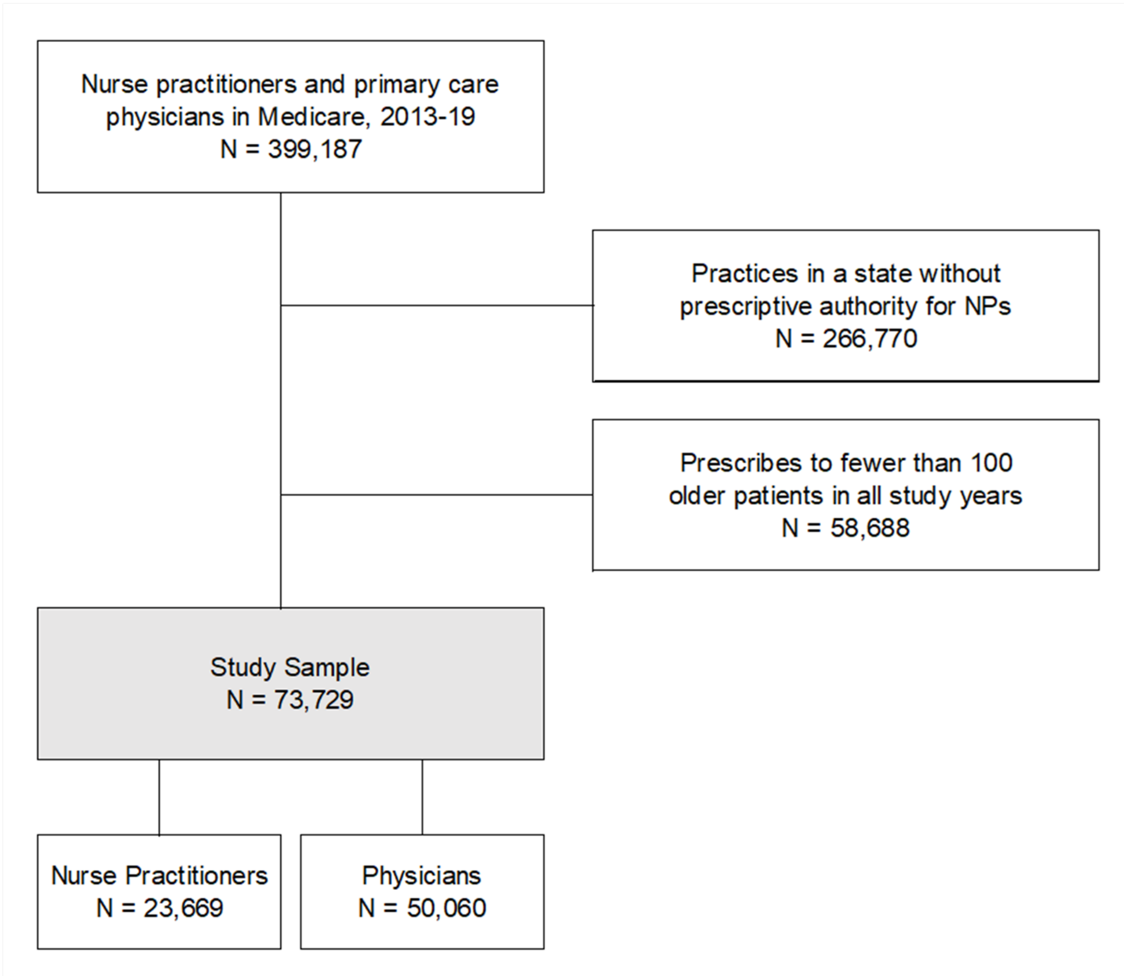
Data source. The Medicare Provider Utilization and Payment Data include publicly available data files that summarize the utilization of and payments for procedures, services, and prescription drugs provided to Medicare fee-for-service beneficiaries by physicians and other clinicians. The datasets are formed using information from reimbursement claims submitted to Centers for Medicare Medicaid Services (CMS) and cover common outpatient services, all physician and other supplier procedures and services, and all Medicare Part D prescriptions.

Data linkage. The primary datasets used in our analysis were formed using Part B and Part D claims in the Medicare Provider Utilization and Payment Data. Clinicians who had a valid National Provider Identifier (NPI) and submitted ≥ 11 noninstitutional services for reimbursement in a calendar year entered the Part B file for that year. Clinicians who had a valid NPI and prescribed ≥ 11 medications in a calendar year entered the Part D file for that year. Observations from the Part B and D files were linked by NPI and year.

Study cohort. Figure S1 depicts the formation of the study cohort. As described in the manuscript, nurse practitioners (NPs) and primary care physicians were identified using CMS specialty codes in the Part B file: code 50 for NPs, and codes 01 (general practice), 08 (family practice), 11 (internal medicine), and 38 (geriatric medicine) for physicians. The study cohort was restricted to clinicians who practiced in one of the 29 states that had granted NPs prescriptive authority by January 1, 2019. Finally, in each year, we excluded clinicians who had prescribed to fewer than 100 “older patients” (defined throughout as patients aged 65 years or older.)

These restrictions left 23,669 NPs and 50,060 primary care physicians in the study cohort. It was an open cohort, with some clinicians entering after 2013 and some exiting after one or more years of observation. On average, clinicians in the study cohort were observed for 3.2 years in the 7-year period 2013 through 2019.

Figure 4.4: Formation of the Study Cohort



Notes: This figure displays the sample selection criteria applied to construct the sample of nurse practitioners and physicians analyzed in this study.

2.2 Laws conferring independent prescriptive authority on nurse practitioners

Previous studies of NP quality of care have analyzed various measures of scope of practice. Some have focused on general practice authority while others have more specifically addressed prescriptive authority. Our study exclusively focused on prescriptive authority.

Following an approach taken previously by McMichael and Markowitz (2023), we defined a state as having granted NPs independent prescriptive authority "if they do not legally mandate any form of supervision by or collaboration with physicians as a condition of NPs practicing and they do not restrict the prescriptive authority of NPs." We began with the 50-state classifications reported in this study. We then reviewed the current law in each state and the District of Columbia to ensure no subsequent changes had occurred, and contacted legislative offices in any jurisdiction in which we had questions or uncertainty about the applicable rules.

As noted above, our analysis focused on states that had granted independent prescriptive authority for NPs prior to and during our study period (2013 – 2019). For states in which the prescriptive authority was enacted during the study period, we included prescription data from clinicians in these states beginning in the year after the law came into effect. For example, NPs were granted prescriptive authority in Delaware in September 2015, so clinicians practicing in Delaware entered the study cohort for years 2016 through 2019.

Table 4.7: State Laws Pertaining to Nurse Practitioner Prescribing Authority, as of February 2022

State	**Effective date**	**Statute or bill**
Alaska	December 1984	Alaska Admin. Code tit. 12, § 44.440.
Arizona	December 1999	Ariz. Rev. Stat. §32-1651.
Arkansas	July 2021	Ark. Code Ann. § 17-87-314
California	January 2023	Cal. Bus & Prof Code § 2836.1.
Colorado	July 2010	Colo. Rev. Stat §12-255-112.
Connecticut	July 2014	Conn. Gen. Stat. §20-87a.
Delaware	September 2015	Del. Code Ann. tit. 24, § 1935.
Florida	July 2020	Fla. Stat. Ann. §464.012.
Hawaii	July 2009	Haw. Rev. Stat. Ann. § 457-8.6.
Idaho	July 2004	Idaho Code § 54-1402(1).
Illinois	January 2018	225 Ill. Comp.Stat. Ann. 65/65-40.
Iowa	1993	Iowa Code §147.107.
Maine	September 1995	Me. Rev. Stat.Ann. tit. 32, §2210.
Maryland	October 2010	Md. Code Ann.Health Occupations § 8-512(a)(2).
Massachusetts	January 2021	1397 Mass. Code Regs. 57(2.10).
Minnesota	January 2015	Minn. Stat. § 148.235.
Montana	Before 1998	Mont. Admin. R.24.159.1461.
Nebraska	March 2015	Neb. Rev. Stat. Ann.§ 38-2315.
Nevada	July 2013	Nev. Rev. Stat.Ann. § 639.1375
New Hampshire	Before 1998	N.H. Rev. Stat.Ann. § 326-B:11.
New Mexico	Before 1998	N.M. Stat. Ann. §61-3-23.2.
New York	January 2015	2020 N.Y. CLS Educ Consol. Laws § 6902(3).
North Dakota	August 2011	N.D. Admin. Code 54-05-03.1-03.
Oregon	Before 1998	Or. Rev. Stat. §678.390.
Rhode Island	June 2013	R.I Gen. Laws §5-34-49
South Dakota	July 2017	S.D. Codified Laws. Ann. § 36-9A-12
Utah	May 2016	Utah Code Ann. § 58-31b-803.
Vermont	June 2011	04-030-170 Vt. Code R. § 8.5.
Virginia	April 2018	Va. Code Ann. §54.1-2957.01.
Washington	July 2005	Wash. Rev. Code Ann. § 18.79.050.
West Virginia	June 2016	W. Va. Code Ann. §30-7-15b.
Wyoming	Before 1998	Wyo. Stat. Ann. § 33-21-120(a)(i)(A).
Washington D.C.	Before 1998	D.C. Code § 3-1206.01.

Notes: This table provides information on independent prescriptive authority of NPs in each of the 50 states and the District of Columbia as of February 2022. For jurisdictions that have conferred this authority, we show the effective date of law enactment and provide a citation to the relevant statute or bill.

2.3 Beers Criteria

The American Geriatric Society's Beers Criteria for Potentially Inappropriate Medication Use in Older Adults (hereafter, "Beers Criteria") is an evidence-based list of medications that are deemed potentially inappropriate for use in adults aged 65 years or older. The guideline is developed and maintained by an interdisciplinary expert panel that reviews evidence using an adapted DELPHI method. The panel consists of 13 clinicians—including physicians, pharmacists, and nurses—with experience in varied practice settings.

Recommendations in the Beers Criteria fall into several categories. Some medications are deemed "potentially inappropriate" for all older adults. For example, the recommendation against meprobamate relates to all adults aged 65 or older. Other recommendations are conditional and call for consideration of drug-disease interactions (i.e., avoid for adults with certain conditions) or drug-drug interactions (i.e., avoid for adults who are taking a contraindicated medication). One example of a conditional recommendation is dronedarone, which prescribers are advised to "avoid in individuals with permanent atrial fibrillation or severe or recently decompensated heart failure". Our analysis focused on medications listed in the category of potentially inappropriate for most older adults, and our main outcome measure (inappropriate prescribing rate) was specified with reference only to medications in this group.

Using generic drug names, we matched the medications of interest in the Beers Criteria to the prescription drug information in the Medicare Provider Utilization and Payment Data. Two investigators independently conducted this matching, and the results were verified for accuracy by a practicing physician and practicing pharmacist.

The Beers Criteria categorizes its recommendations by medication class. As noted in the manuscript, in our analytic dataset the ten most commonly prescribed medication classes accounted for 99.5% of all inappropriate prescribing detected. Those classes are shown in the manuscript, namely: muscle relaxants, antidepressants, hypnotics, first-generation antihistamines, antispasmodics, sulfonylureas, barbiturates, antineoplastics, thyroid, and NSAIDS.

2.4 Model covariates

The variables used as model covariates in our analyses came from the Medicare Provider Utilization and Payment Data (Part B and Part D) and the Medicare Provider Enrollment, Chain, and Ownership System. These variables were clinician gender, patient volume, CMS risk score, clinical setting, practice location, years of experience, year, and practice state. Details of each are provided below.

Patient volume. This variable indicates the number of Medicare beneficiaries aged 65 or older treated by a clinician in a given year.

CMS risk score. CMS developed a risk-adjustment model using hierarchical condition categories to assign risk scores to fee-for-service beneficiaries. Risk scores are based on a beneficiary's age and sex; whether the beneficiary is eligible for Medicaid, first qualified for Medicare on the basis of disability, or lives in an institution; and the beneficiary's diagnoses from the previous year. The data we used included a variable that averaged these patient risk scores at the clinician-year level; they were not available at the patient or encounter level.

Clinical practice setting. We categorized the clinical practice setting information using evaluation and management codes in Medicare Part B claims. The practice settings are outpatient (CPT/HCPCS: 99201, 99202, 99203, 99204, 99205, 99211, 99212, 99213, 99214, 99215, 99354, 99355); inpatient (CPT/HCPCS: 99217, 99218, 99219, 99220, 99221, 99222, 99223, 99224, 99225, 99226, 99231, 99232, 99233, 99234, 99235, 99236, 99238, 99239, 99356, 99357); emergency care (CPT/HCPCS: 99281, 99282, 99283, 99284, 99285), critical care (CPT/HCPCS: 99291, 99292); nursing facility (CPT/HCPCS: 99304, 99305, 99306, 99307, 99308, 99309, 99310, 99315, 99316, 99318); assisted living facility (CPT/HCPCS: 99324, 99325, 99327, 99328, 99334, 99335, 99336, 99337); and home visits (CPT/HCPCS: 99341, 99342, 99343, 99344, 99345, 99347, 99348, 99349, 99350). *Practice location.* This variable indicates the population size of the largest urban core of the core-based statistical area in which a clinician's primary practice was located. We classified practice location into five categories: large metropolitan ($>$ 1 million residents), medium metropolitan (250,000 to 1 million residents), small metropolitan (25,000 to 249,999 residents), micropolitan (10,000 to 24,999 residents), and remote ($<$ 10,000 residents).

Practice experience. We defined this variable as the number of years since the clinician obtained their NP or medical degree; it was calculated using information on graduation years provided in the Medicare Provider Enrollment, Chain, and Ownership System. We categorized the continuous measure of years of experience into four ten-year bins: $<$ 10 years, 10 to 19 years, 20 to 29 years, and $\geq$ 30 years. Practice experience data were missing for 24% of NPs and 19% of primary care physicians in our study cohort. Our main analysis used dummy-variable adjustment to account for the set of clinicians with missing practice experience data. Alternatively, in one of our sensitivity analyses, we evaluated potential bias introduced by missing data on practice experience by computing worst-case bounds for our estimates.

2.5 Statistical approach

Our main statistical analyses were conducted at the clinician-year level. These analyses used longitudinal data on prescriptions written by 73,729 clinicians, many of whom contributed prescription data across multiple years. To analyze variation in prescribing patterns (measured as a fractional rate) across clinician-years, we used logistic regression ("fracreg logit", Stata version 17.0) to estimate the association between the independent variable of interest (NP versus primary care physician) and the outcome (inappropriate prescribing rate, defined according to Beers Criteria). These analyses, also known as marginal or population-averaged analyses, allowed us to investigate between-clinician variation in prescribing patterns holding fixed other covariates in the model.

To ensure our findings were not driven by a particular set of parametric assumptions, we chose a flexible parametric model that estimates the gap in inappropriate prescribing rates between NPs and primary care physicians using a rich set of covariate indicators as control variables, including CMS patient risk score and practice experience.

We cluster-adjusted the standard errors associated with odds ratios derived from the logistic regression models. Clustered standard errors are statistically consistent in settings where observations can be subdivided into smaller groups (i.e., clusters), and where the dependent variable is systematically correlated within each cluster. As we note in the manuscript, clinicians were observed across multiple years in our longitudinal dataset. Therefore, to account for potential within-clinician correlations in prescribing patterns, we clustered the standard errors at the clinician level.

State-specific inappropriate prescribing rates for NPs and primary care physicians were estimated using logistic regression analyses in which the independent variable of interest (NP versus physician) was interacted with state fixed effects. Inappropriate prescribing rates for hospital referral regions (HRR) were estimated by the same procedure but using HRR fixed effects. HRRs were defined according to the clinician's practice zip code using the Dartmouth

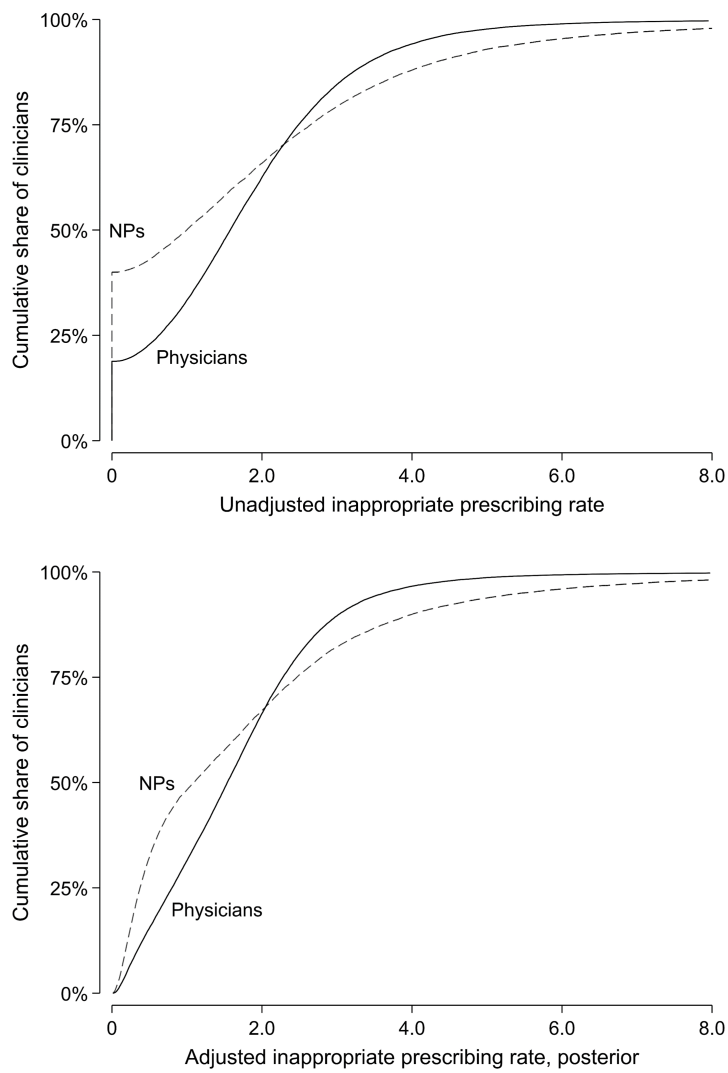
Atlas crosswalk data.

Finally, to estimate and rank clinicians by their adjusted inappropriate prescribing rate, we used a mixed effects logistic regression model ("melogit", Stata version 17.0) with clinician random effects. Random effects model within-clinician correlation in prescribing patterns, controlling for the set of covariates described above. In addition, the empirical Bayes approach we employ pools information across NPs and primary care physicians to estimate the empirical distribution of inappropriate prescribing rates. Posterior means are subsequently shrunken toward a common prior to reduce the variance of the clinician-level rates, especially in cases where individual estimates are heavily influenced by noise and outliers.

2.6 Distribution of inappropriate prescribing rates

On average, unadjusted and adjusted rates of inappropriate prescribing for NPs and physicians were very similar, but their distributions varied substantially. As noted in the manuscript, NPs were overrepresented at the upper and lower ends of the distribution of inappropriate prescribing. The distributions for NPs and primary care physicians differ substantially, with NPs' distribution having thicker tails compared with the physicians' distribution.

Figure 4.5: Cumulative Distribution Function for Rates of Inappropriate Prescribing



Notes: This figure plots the empirical cumulative distribution function (CDF) for unadjusted inappropriate prescribing rates for NPs and primary care physicians. The CDF represents the proportion of NPs and primary care physicians whose rate of inappropriate prescribing was at or below a specified level. For example, the proportion of NPs with an inappropriate prescribing rate of 0 is 43%. The proportion of NPs with an inappropriate prescribing rate of 2 or below per 100 prescriptions is 68%.

2.7 Sensitivity analyses

Sensitivity analysis 1: exclude patient risk score. Our clinician-level data set did not include patient characteristics other than the CMS-generated risk scores. These risk scores are commonly used by researchers as a proxy for the acuity of patient populations. Our primary analysis adjusted for these risk scores, averaging them at the clinician-year level and then categorizing them into six bins for purposes of specifying this measure as a covariate.

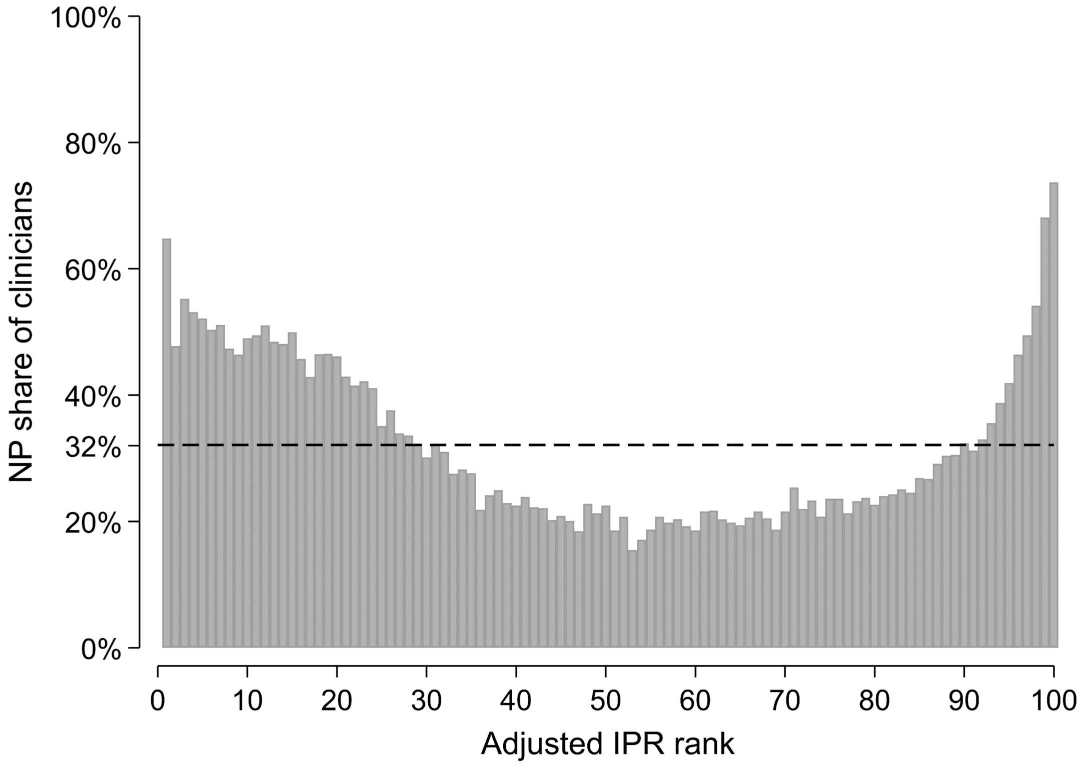
To assess the potential for bias associated with unobserved patient-level characteristics, we excluded the categorical variable indicating patient risk scores and then re-ran the primary analyses. This approach to evaluating the robustness of our model to omitted variable bias involves assessing changes in coefficients after the inclusion/exclusion of relevant control variables. Table S5 shows that the removal of the CMS risk scores did not have a significant impact on our overall results. This provides a degree of assurance that our estimates were not biased by systematic differences in the case mix or acuity of patients treated by NPs and physicians, respectively.

Table 4.8: Adjusted Rates of Inappropriate Prescribing Excluding Patient Risk Score

	Main model		Sensitivity analysis #1	
	Odds ratio	Standard error	Odds ratio	Standard error
Nurse practitioner	0.990	0.011	0.930	0.011

Notes: This table reports odds ratios from a logistic regression model for the inappropriate prescribing rate. Clustered standard errors associated with coefficients are presented.

Figure 4.6: Proportion of Clinicians who were NPs by Percentile of Adjusted Inappropriate Prescribing, Excluding Patient Risk Score



Notes: This figure presents the proportion of clinicians who were NPs within each percentile of the adjusted inappropriate prescribing rate, excluding patient risk score from the set of covariates in the mixed effects logistic regression model used to generate the estimates.

Sensitivity analysis 2: calculate worst-case bounds. Our second sensitivity analysis focused on potential bias arising from missing data on practice experience. As noted above, we defined practice experience as the number of years since a clinician received their NP or medical degree, as measured in the year they entered the cohort. This continuous measure was then sorted into 10-year bins. 5,736 NPs (24%) and 9,622 physicians (19%) in our cohort had missing data on this variable.

To test model robustness to these missing data, we computed worst-case bounds. Specifically, we performed an iterative process assuming the lowest, middle, and highest bound values for the missing practice experience variable and measured the sensitivity of regression coefficients to these changes. In one specification, for example, we assumed that clinicians with missing experience data had fewer than 10 years of experience. In another specification, clinicians with missing experience data were assumed to have between 11 and 19 years of experience. We iterated this process until every possible value of practice experience was exhausted.

The results of this sensitivity analysis were virtually identical to the main result, demonstrating non-inferiority for NPs' quality of prescribing in every iteration. Table S6 shows the adjusted odds ratio of interest, iterating over various imputations of the values for clinician-years that had missing values for the practice experience variable. The coefficient in Lower bound represents the minimum adjusted difference in inappropriate prescribing rates between NPs and physicians. The coefficient in Upper bound represents the maximum adjusted difference in inappropriate prescribing rates between NPs and physicians.

Table 4.9: Bounds for Adjusted Rates of Inappropriate Prescribing

	Lower bound		Upper bound	
	Odds ratio	Standard error	Odds ratio	Standard error
Nurse practitioner	0.990	0.011	0.990	0.011

Notes: This table reports odds ratios from a logistic regression model for the inappropriate prescribing rate. Clustered standard errors associated with coefficients are presented.

Sensitivity analysis 3: analysis of high-volume prescribers. NPs in our study cohort averaged substantially fewer prescriptions per year (1,271) than physicians (3,800). This discrepancy raises the possibility that our main finding—NPs have similar rates of inappropriate prescribing to physicians—may be driven by the fact that NPs are less likely to prescribe any medication to their patients. Our use of prescription thresholds (both types of clinicians had to have prescribed for at least 100 older adults in any year they contributed data) partially guarded against this risk, but not completely.

This sensitivity analysis took the exclusion further, removing from the analytic sample clinician-years (N = 26,688) in which the clinician prescribed medications for fewer than 500 older adults.

This adjustment had moderate effects on our main results, which was not entirely surprising given the very substantial reconstitution of the study sample. Among clinicians who prescribed drugs for 500 or more older adults per year, NPs' inappropriate prescribing rate was 0.50 per 100 prescriptions lower than physicians' inappropriate prescribing rate. The effects on the distributional result were more pronounced, but did not change the direction or qualitative nature of the reported estimates: NPs were still disproportionately overrepresented in the highest and lowest deciles of inappropriate prescribing.

Table 4.10: Adjusted Rates of Inappropriate Prescribing for High-Volume Prescribers

	Main model		Sensitivity analysis #3	
	Odds ratio	Standard error	Odds ratio	Standard error
Nurse practitioner	0.990	0.011	0.747	0.023

Notes: This table reports odds ratios from a logistic regression model for the inappropriate prescribing rate. Clustered standard errors associated with coefficients are presented.

Sensitivity analysis 4: exclude NPs practicing in a facility and home visit settings. The Beers Criteria is not intended for use in adults receiving hospice or palliative care, as discussed in the limitations section of the manuscript. With the Medicare data used in this study, there is no direct way to identify and exclude those patients from the main analysis. However, it was feasible to use information on the clinical settings in which care was delivered to probe the effect on our estimates of including/excluding patients receiving hospice or palliative care. In our analytic sample of NPs and primary care physicians, 14,244 out of 232,794 (6.1%) clinician-years had a majority of evaluation-management claims located in a nursing facility, assisted living facility, or home visit setting—settings in which hospice or palliative care are disproportionately likely to occur (recognizing, of course, that a good deal of other forms of care are also delivered in these settings).

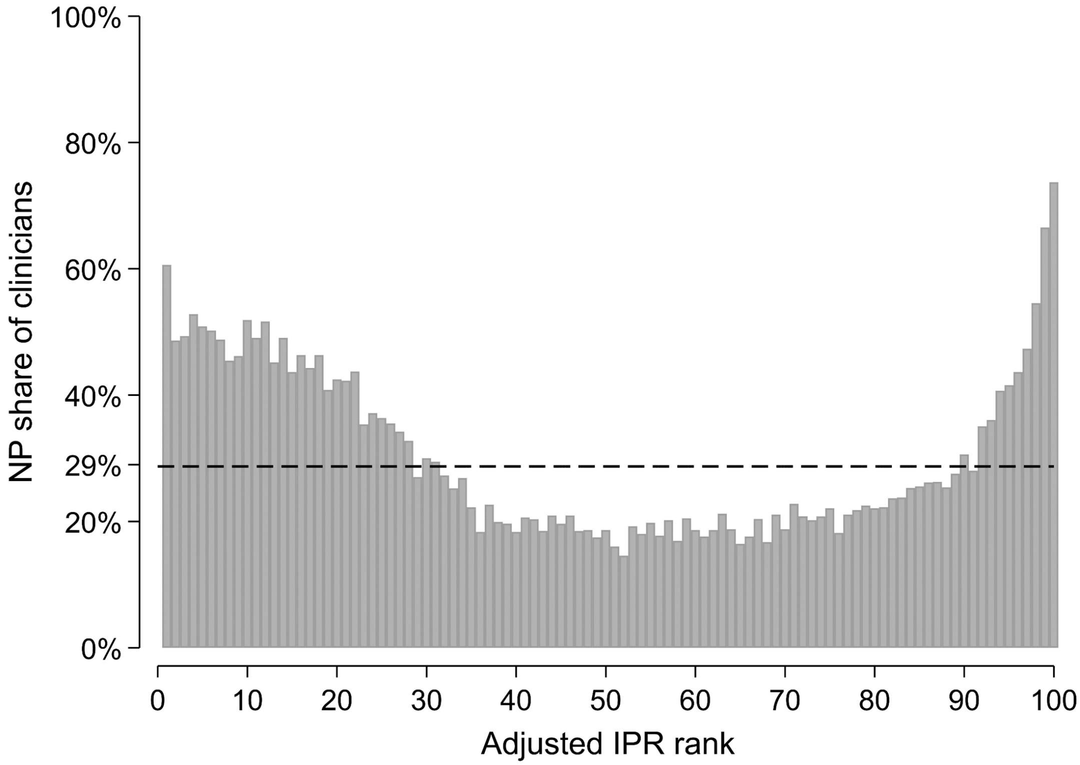
Re-running the logistic regression analysis after excluding these clinician-years produced an estimate that was very consistent with our reported finding: primary care physicians had a fractionally higher rate of inappropriate prescribing than NPs, but the difference was not statistically significant. Table S8 reports the odds ratio from the logistic regression model after exclusion of the clinician-years.

Table 4.11: Adjusted Rates of Inappropriate Prescribing Excluding Facility and Home Visit

	Main model		Sensitivity analysis #4	
	Odds ratio	Standard error	Odds ratio	Standard error
Nurse practitioner	0.990	0.011	0.987	0.012

Notes: This table reports odds ratios from a logistic regression model for the inappropriate prescribing rate. Clustered standard errors associated with coefficients are presented.

Figure 4.7: Proportion of Clinicians who were NPs by Percentile of Adjusted Inappropriate Prescribing, Excluding Facility and Home Visit Settings



Notes: This figure shows the re-estimated distributional results estimate from a sample that excluded clinician-years that had a majority of evaluation-management claims located in a nursing facility, assisted living facility, or home visit setting.

Sensitivity analysis 5: matching and doubly robust methods. We probed our regression-based approach by comparing it to alternative matching-based methods. Specifically, we conducted three forms of matching and re-estimated our results using the comparison groups generated by each of these matching approaches. First, we conducted propensity-score kernel matching between NPs and primary care physicians, using all of the covariates discussed in the Statistical Analysis section to calculate the propensity score. The covariates include an indicator variable for whether practice experience was missing. Bandwidth selection was performed to minimize mean squared error, and standard errors were clustered at the clinician level. In this matching method, each NP was matched with the weighted average of "similar" physicians, where weights were inversely proportional to the distance between NPs' and physicians' propensity scores. Analysis of this matched sample showed that NPs had lower inappropriate prescribing rates than physicians, although the difference was very small and not clinically meaningful.

Second, we conducted propensity-score kernel matching between NPs and physicians using the set of available covariates (as described above) and adding exact matching on state and year. Analysis of this analytic sample produced no significant difference between NPs' and physicians' mean rates of inappropriate prescribing.

Finally, we conducted doubly robust regression analysis combined with inverse propensity score weighting. This approach allowed us to jointly model the outcome variable (using logistic regression, mirroring the main analysis in the revised manuscript) and the provider classification as NP or primary care physician (using logistic regression to estimate the propensity score). Doubly robust estimators are unbiased if either the outcome or propensity score model is correctly specified. The estimates produced by this sample, again, showed a very small and non-meaningful difference in the mean rates of inappropriate prescribing for NPs and physicians.

Table 4.12: Adjusted Rates of Inappropriate Prescribing Using Alternate Statistical Approaches

	Regression	PS match	Exact match	Doubly robust
Outcome model	Logit	Nonparametric	Nonparametric	Logit
Treatment model	Nonparametric	Logit	Logit	Logit
Rate for NPs	1.66	1.58	1.62	1.60
Rate for physicians	1.68	1.69	1.70	1.68
Difference	-0.016	-0.115	-0.085	-0.083
	(-0.052, 0.020)	(-0.203, -0.027)	(-0.193, 0.0002)	(-0.155, -0.012)

Notes: Adjusted means represent population-averaged inappropriate prescribing rates for NPs and primary care physicians, respectively, per 100 prescriptions for older adults. Differences in appropriate prescribing rates should be interpreted as rate differences (per 100 prescriptions) between NPs and primary care physicians. 95% confidence intervals associated with rate differences are shown.

Sensitivity analysis 6: exclude NPs potentially not practicing in a primary care setting. CMS data, like most administrative claims data sets, do not provide variables indicating the clinical subspecialty of NPs, or the precise clinical setting in which NPs are practicing. However, the vast majority of NPs' clinical work occurs in primary care settings. National surveys estimate that approximately 90% of NPs are certified in an area of primary care: 69.7% family medicine; 10.8% adult; 8.8% gerontology; 3.2% pediatrics; and 2.2% women's health.

Despite generally available evidence on the predominance of primary care in NP clinical activities, we probed the possibility that prescriptions by NP subspecialists may be driving the differences we observed at the tails of the distribution of inappropriate prescribing rates for NPs. We did this by employing a data-driven approach to identify NPs who were unlikely to be practicing in a primary care setting, and then re-ran our main analysis after excluding them. The approach took advantage of information at the clinician level on diagnosis codes and evaluation and management codes. Specifically, the approach proceeded through the following steps:

First, we used the CMS data described previously to create a sample of 296,764 physicians. who treated $\geq$ 50 Medicare Part B beneficiaries per year between 2013 and 2019. Information on physicians' specialty allowed us to observe which sample members specialized in primary care (internal medicine, family practice, family practice, general practice, and geriatric medicine) and which had other specialties (emergency medicine, psychiatry, hospitalist, critical care intensivists, geriatric psychiatry, neuropsychiatry, and clinical psychology).

Next, we used logistic regression analysis at the physician level to identify clinical variables associated with primary care specialization. The outcome variable in this analysis distinguished primary care physicians from physicians in other specialties. The independent variables included proportion of patients with a comorbid diagnosis for: atrial fibrillation; Alzheimer's, related disorders, or dementia; asthma; cancer; congestive heart failure; chronic kidney disease; chronic obstructive pulmonary disease; depression; diabetes; hyperlipidemia; hypertension; ischemic heart disease; osteoporosis; rheumatoid arthritis/osteoarthritis; schizophrenia and other psychotic disorders; and stroke. Other independent variables were the proportion of evaluation and management claims in each clinical setting (outpatient, inpatient, emergency department, and critical care). In addition, we added quadratic terms to account for non-linearities. The pseudo R-squared statistic for the estimated logistic regression was 0.59, indicating that these clinical variables, collectively, were quite successful in distinguishing primary care physicians from other types of physicians.

Finally, we used the coefficients estimated in the logistic regression analysis described above to predict the probability each NP in our study sample was working in primary care, as opposed to other specialties. We classified NPs as being likely to be practicing in a primary care setting if their predicted probability was greater than 50%.

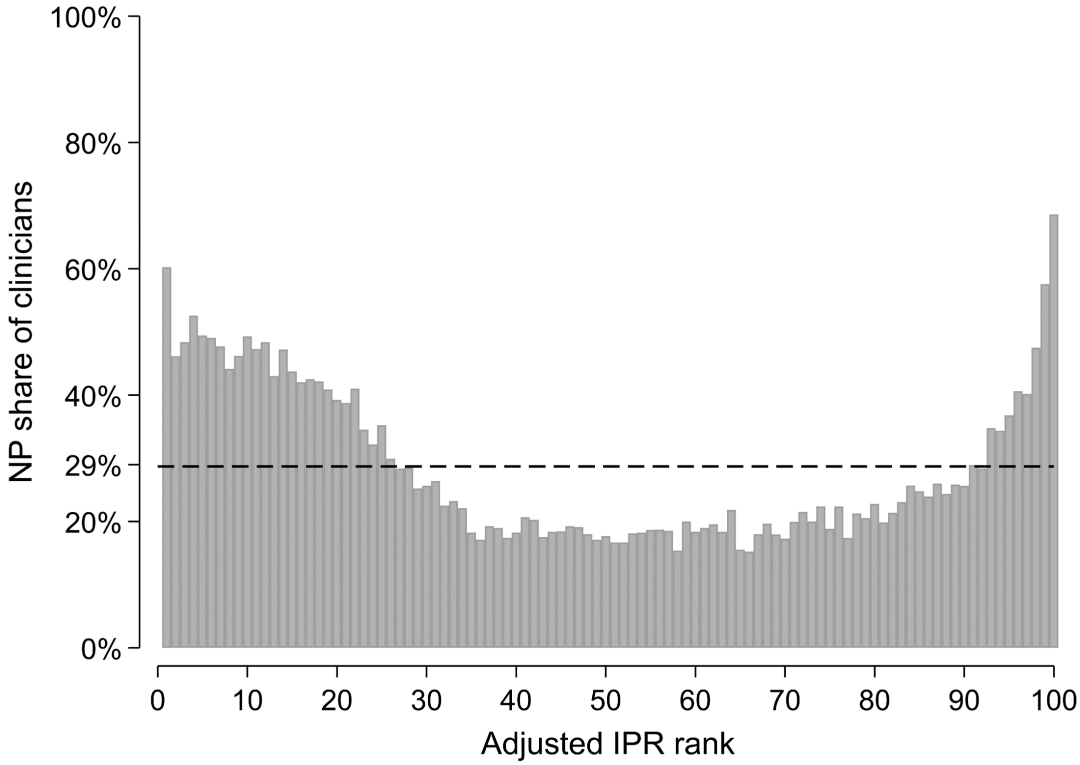
Excluding NPs who probably were not practicing in primary care settings had negligible effects on our main results, estimated on a sample that excluded NPs classified as being likely to be practicing in a primary care setting.

Table 4.13: Adjusted Rates of Inappropriate Prescribing Excluding Potentially Non-Primary Care NPs

	Main model		Sensitivity analysis #6	
	Odds ratio	Standard error	Odds ratio	Standard error
Nurse practitioner	0.990	0.011	0.974	0.012

Notes: This table reports odds ratios from a logistic regression model for the inappropriate prescribing rate. Clustered standard errors associated with coefficients are presented.

Figure 4.8: Proportion of Clinicians who were NPs by Percentile of Adjusted Inappropriate Prescribing, Excluding Non-Primary Care NPs



Notes: This figure shows the re-estimated distributional results estimate from a sample that excluded NPs who were likely to be practicing in a non-primary care setting.

3. Appendix to "Vaccine Incentives Hurt Intrinsic Motivation: Evidence from a Survey Experiment"

3.1 Qualitative interviews and analysis

To identify the determinants of demand for the COVID-19 vaccine, a racially and ethnically diverse subset of survey respondents were recruited for a 30-minute semi-structured interview. The inductive, semi-structured format of the interviews allowed participants to describe and explain what types of interventions, including interventions involving monetary incentives, would increase their confidence in taking the vaccine. Interviews focused on three domains: i) concerns and distrust toward the COVID-19 vaccine, ii) desires for policy interventions and/or government action related to the vaccine, iii) desires for non-pharmacological policy interventions related to the pandemic. Interviews were conducted by the core research team and research assistants trained in reflexive, qualitative interview methods.

The qualitative study (N=27) was nested within the larger survey of unvaccinated adult residents in Los Angeles County. Interviewers contacted respondents who opted in for a 30-minute semi-structured interview. Participants completed interviews via Zoom or telephone (4 in Spanish and 23 in English). Interviews were recorded, translated into English if needed, and transcribed. Using Dedoose software, transcripts were analyzed thematically through iterative codebook development, constant comparison between codes and transcripts, and consensus decision on emergent themes. Inter-rater reliability testing showed a pooled Kappa statistic of 0.78, indicating substantial team agreement in coding transcripts.

The qualitative findings corroborate this paper's risk-signaling and prosocial motivation hypotheses. The presence of financial incentives solicited adverse perceptions of the COVID-19 vaccines among unvaccinated Los Angeles County residents. In particular, financial incentives sparked or elevated suspicion and skepticism towards vaccine effectiveness, viability (i.e., whether it works or not), and safety among vaccine-hesitant residents. Non-financial incentives, such as free food or event tickets, solicited similar responses. Below, we show some exemplar quotes from participants.

1. “I know at the store down the street, they they’ll give you $50 if you get vaccinated and to me I don’t know that’s like to me, that’s on the lines of like selling yourself for people to test [you]. I don’t know. I think… I, I just- I don’t like it.” — 44 year old White woman

2. “The money incentives to get the vaccines, my friend was extremely suspicious of that. Because she says, if it needs to be incentivized that way, it makes her even more unwilling to get the vaccine because it sounds more like a pushing agenda or why do they need to kind of, not trick people, but incentivize them that way? Like, this seemed less scientific, but more of a hidden agenda behind it to offer perks or benefits.” – 25 year old Asian non-binary person

3. “I feel like the way they’re going about trying to get people to get is like just sketchy. I don’t think in history or during my time being here I’ve ever seen them try to. Like, for example, you get a free Krispy Creme doughnut if you have the vaccine. I’ve seen the other day, they were like there were something involving like a lottery ticket and vaccines.” – 19 year old Black woman

3.2 Comparison with population of vaccine-hesitant adults in California

Table 4.14 presents baseline summary statistics for our sample and for a representative sample of adults in California. The latter comes from the California Health Interview Survey (CHIS 2020). CHIS is an online and telephone survey covering a wide range of health topics, including topics related to COVID-19 vaccine hesitancy. It employs a sampling methodology and extensive questionnaires to collect information that reflects California's diverse populations. CHIS is conducted annually by the UCLA Center for Health Policy Research in collaboration with the California Department of Public Health, and the Department of Health Care Services.

We stratified the CHIS sample based on participants' response to the question: "If a vaccine becomes available for COVID-19, would you get it?" 4,183 individuals responded "no" to this question, representing approximately 7 million vaccine-hesitant adults in California. We compared these individuals' demographics with those from our study. Compared to the vaccine-hesitant population in California, participants in our survey experiment were younger and more likely to be female and Black/African-American.

Multiple factors explain the age, gender, and racial gap between vaccine-hesitant respondents of CHIS and our survey. First, while our survey was available online and via telephone, 99% of respondents opted to take the survey online. Other studies have shown that online respondents are younger and more likely to be female and more highly educated. Second, we oversampled non-white respondents in order to adequately analyze behavioral differences between racial and ethnic subgroups.

Table 4.14: Comparison with Vaccine-Hesitant Adults in CHIS

	CHIS		Our study
Would get COVID-19 vaccine?	Yes	No	No / not sure
Age			
18-34	0.30	0.32	0.55
35-49	0.24	0.27	0.33
50-64	0.23	0.25	0.08
65 and over	0.23	0.16	0.04
Man	0.51	0.44	0.34
Woman	0.49	0.56	0.65
Race & ethnicity			
Latino	0.36	0.50	0.38
Asian	0.15	0.08	0.09
Black	0.04	0.10	0.35
White	0.41	0.29	0.34
College graduate	0.46	0.26	0.32
Ever tested positive	0.14	0.13	0.12
Self-reported health			
Excellent	0.20	0.21	0.23
Very good	0.35	0.31	0.35
Good	0.30	0.32	0.28
Fair	0.13	0.13	0.11
Poor	0.02	0.03	0.03
Survey N	17,766	4,183	513
Population N	22,727,828	6,957,054	

Notes: This table compares baseline variable means from our study's sample with those from the California Health Interview Survey (CHIS; 2020 Adult Survey). The first column reports mean values for CHIS respondents who answered "yes" to the question, "If a vaccine becomes available for COVID-19, would you get it?" The second column reports mean values for CHIS respondents who answered "no" to this question. The final column reports mean values for individuals who participated in our survey experiment. CHIS estimates were strata- and cluster-weighted to represent the California population.

Bibliography

Abadie, Alberto (2003) "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics*, 113 (2), 231–263.

Aguiar, João Pedro, José Gustavo Marques, and Filipa Alves da Costa (2021) "Utility and Limitations of a Screening Tool of Older Person's Prescription Among Psychiatric Elder Patients: A Comprehensive Review," *Aging & Health Research*, 1 (3), 100031.

Aizer, Anna, Sungwoo Cho, Shari Eli, and Adriana Lleras-Muney (2023) "The Impact of Cash Transfers to Poor Mothers on Family Structure and Maternal Well-Being," *American Economic Journal: Applied Economics*.

Aizer, Anna and Joseph J Doyle Jr (2015) "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges," *Quarterly Journal of Economics*, 130 (2), 759–803.

Akerlof, George A. and Rachel E. Kranton (2010) *Identity Economics*: Princeton University Press.

Alexander, Dana and Molly Schnell (2019) "Just What the Nurse Practitioner Ordered: Independent Prescriptive Authority and Population Mental Health," *Journal of Health Economics*, 66, 145–162.

Alsan, Marcella and Marianne Wanamaker (2018) "Tuskegee and the Health of Black Men," *Quarterly Journal of Economics*, 133 (1), 407–455.

Altonji, Joseph G, Todd E Elder, and Christopher R Taber (2005) "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 113 (1), 151–184.

American Geriatrics Society (2012) "2012 Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults," *Journal of the American Geriatrics Society*, 60, 616–631.

——— (2019) "2019 Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults," *Journal of the American Geriatrics Society*, 67, 674–694.

American Medical Association (2018) "Physician-Led Health Care Teams," https://ama-assn.org/system/files/2018-09/physician-led-teams-campaign-booklet.pdf.

Angrist, Joshua D and Guido W Imbens (1995) "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of the American Statistical Association*, 90 (430), 431–442.

Apouey, Benedicte and Andrew E. Clark (2015) "Winning Big but Feeling No Better? The Effect of Lottery Prizes on Physical and Mental Health," *Health Economics*, 24 (5), 516–538.

Ariely, Dan, Anat Bracha, and Stephan Meier (2009) "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially," *American Economic Review*, 99 (1), 544–555.

Arnold, David, Will Dobbie, and Peter Hull (2022) "Measuring Racial Discrimination in Bail Decisions," *American Economic Review*, 112 (9), 2992–3038.

Arnold, David, Will Dobbie, and Crystal S Yang (2018) "Racial Bias in Bail Decisions," *Quarterly Journal of Economics*, 133 (4), 1885–1932.

Arteaga, Carolina (2023) "Parental Incarceration and Children's Educational Attainment," *Review of Economics and Statistics*, 1–17.

Arteaga, Carolina and Victoria Barone (2022) "A Manufactured Tragedy: The Origins and Deep Ripples of the Opioid Epidemic,"Technical report, Working paper. http://www. carolina. com/s/Opioids_ArteagaBarone_Jan2022. pdf.

Association of American Medical Colleges (2021) "The Complexities of Physician Supply and Demand: Projections from 2019 to 2034," https://aamc.org/media/54681/download.

Autor, David, David Dorn, and Gordon Hanson (2019) "When Work Disappears: Manufacturing Decline and the Falling Marriage Market Value of Young Men," *American Economic Review: Insights*, 1 (2), 161–178.

Autor, David H, David Dorn, and Gordon H Hanson (2013) "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 103 (6), 2121–2168.

Autor, David H and Mark G Duggan (2003) "The Rise in the Disability Rolls and the Decline in Unemployment," *Quarterly Journal of Economics*, 118 (1), 157–206.

Autor, David H. and Mark G. Duggan (2006) "The Growth in the Social Security Disability Rolls: A Fiscal Crisis Unfolding," *Journal of Economic Perspectives*, 20 (3), 71–96.

Autor, David H, Mark Duggan, Kyle Greenberg, and David S Lyle (2016) "The Impact of Disability Benefits on Labor Supply: Evidence from the VA's Disability Compensation Program," *American Economic Journal: Applied Economics*, 8 (3), 31–68.

Bakx, Pieter, Bram Wouterse, Eddy Van Doorslaer, and Albert Wong (2020) "Better Off at Home? Effects of Nursing Home Eligibility on Costs, Hospitalizations and Survival," *Journal of Health Economics*, 73, 102354.

Bald, Anthony, Eric Chyn, Justine Hastings, and Margarita Machelett (2022) "The Causal Impact of Removing Children from Abusive and Neglectful Homes," *Journal of Political Economy*, 130 (7), 1919–1962.

Banerjee, Abhijit Vinayak, Esther Duflo, Rachel Glennerster, and Dhruva Kothari (2010) "Improving Immunisation Coverage in Rural India: Clustered Randomised Controlled Evaluation of Immunisation Campaigns With and Without Incentives," *BMJ*, 340.

Barber, Andrew and Jeremy West (2022) "Conditional Cash Lotteries Increase COVID-19 Vaccination Rates," *Journal of Health Economics*, 81.

Baum, Aaron, Juan Wisnivesky, Sanjay Basu, Albert L. Siu, and Mark D. Schwartz (2020) "Association of Geographic Differences in Prevalence of Uncontrolled Chronic Conditions with Changes in Individuals' Likelihood of Uncontrolled Chronic Conditions," *JAMA*, 324 (14), 1429–1438.

Bekker, Paul A (1994) "Alternative approximations to the distributions of instrumental variable estimators," *Econometrica: Journal of the Econometric Society*, 657–681.

Bénabou, Roland and Jean Tirole (2003) "Intrinsic and Extrinsic Motivation," *Review of Economic Studies*, 70 (3), 489–520.

——— (2006) "Incentives and Prosocial Behavior," *American Economic Review*, 96 (5), 1652–1678.

——— (2011) "Identity, Morals, and Taboos: Beliefs as Assets," *Quarterly Journal of Economics*, 126 (2), 805–855.

Benjaminsen, Lars and Stefan Bastholm Andrade (2015) "Testing a Typology of Homelessness Across Welfare Regimes: Shelter Use in Denmark and the USA," *Housing Studies*, 30 (6), 858–876.

Bhuller, Manudeep, Gordon B Dahl, Katrine V Løken, and Magne Mogstad (2020) "Incarceration, recidivism, and employment," *Journal of Political Economy*, 128 (4), 1269–1324.

Björklund, Anders and Robert Moffitt (1987) "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models," *Review of Economics and Statistics*, 42–49.

Black, Dan, Kermit Daniel, and Seth Sanders (2002) "The Impact of Economic Conditions on Participation in Disability Programs: Evidence from the Coal Boom and Bust," *American Economic Review*, 92 (1), 27–50.

Blæhr, Emely Ek and Rikke Søgaard (2021) "Instrumental Variable-Based Assessment of the Effect of Psychotherapy on Suicide Attempts, Health, and Economic Outcomes in Schizophrenia," *Health Economics*, 30 (4), 903–914.

Blandhol, Christine, John Bonney, Magne Mogstad, and Alexander Torgovitsky (2022) "When Is TSLS Actually Late?"Technical report, National Bureau of Economic Research.

Blazer, Dan G and Debara L Tucci (2019) "Hearing Loss and Psychiatric Disorders: A Review," *Psychological Medicine*, 49 (6), 891–897.

Bound, John, David A Jaeger, and Regina M Baker (1995) "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak," *Journal of the American Statistical Association*, 90 (430), 443–450.

Bowles, Samuel and Sandra Polania-Reyes (2012) "Economic Incentives and Social Preferences: Substitutes or Complements?" *Journal of Economic Literature*, 50 (2), 368–425.

Braveman, Paula and Laura Gottlieb (2014) "The Social Determinants of Health: It's Time to Consider the Causes of the Causes," *Public Health Reports*, 129, 19–31.

Brinch, Christian N, Magne Mogstad, and Matthew Wiswall (2017) "Beyond LATE with a Discrete Instrument," *Journal of Political Economy*, 125 (4), 985–1039.

Bruhn, Jesse M, Kyle Greenberg, Matthew Gudgeon, Evan K Rose, and Yotam Shem-Tov (2022) "The Effects of Combat Deployments on Veterans' Outcomes,"Technical report, National Bureau of Economic Research.

Budnitz, Daniel S, Maribeth C Lovegrove, Nadine Shehab et al. (2011) "Emergency Hospitalizations for Adverse Drug Events in Older Americans," *New England Journal of Medicine*, 365, 2002–2012.

Budnitz, Daniel S, Nadine Shehab, Scott R Kegler et al. (2007) "Medication Use Leading to Emergency Department Visits for Adverse Drug Events in Older Adults," *Annals of Internal Medicine*, 147, 755–765.

Butler, Danielle C, Stephen Petterson, Robert L Phillips, and Andrew W Bazemore (2013) "Measures of Social Deprivation that Predict Health Care Access and Need within a Rational Area of Primary Care Service Delivery," *Health Services Research*, 48 (2pt1), 539–559.

California Health Interview Survey (2020) "CHIS 2020 Adult Public Use Files. Version: May 2022," UCLA Center for Health Policy Research, Los Angeles, CA. Accessed on September 22, 2022.

Campos-Mercade, Pol, Armando N. Meier, Florian H. Schneider, Stephan Meier, Devin Pope, and Erik Wengström (2021) "Monetary Incentives Increase COVID-19 Vaccinations," *Science*, 374 (6569), 879–882.

Carlin, Bradley P and Thomas A Louis (2008) *Bayesian Methods for Data Analysis*: CRC Press.

Carneiro, Pedro, James J Heckman, and Edward J Vytlacil (2011) "Estimating Marginal Returns to Education," *American Economic Review*, 101 (6), 2754–2781.

Case, Anne (2004) *Does Money Protect Health Status? Evidence from South African Pensions*, Chap. 7, 287–312: University of Chicago Press.

Case, Anne and Angus Deaton (2015) “Rising Morbidity and Mortality in Midlife Among White Non-Hispanic Americans in the 21st Century,” *Proceedings of the National Academy of Sciences*, 112 (49), 15078–15083.

——— (2017) “Mortality and Morbidity in the 21st Century,” *Brookings Papers on Economic Activity*, 2017, 397.

Cesarini, David, Erik Lindqvist, Robert Östling, and Björn Wallace (2016) “Wealth, Health, And Child Development: Evidence From Administrative Data On Swedish Lottery Players,” *Quarterly Journal of Economics*, 131 (2), 687–738.

Chan, Darren C and Yu Chen (2023) “The Productivity of Professions: Evidence from the Emergency Department,” *National Bureau of Economic Research*, w30608.

Chan, David C, Matthew Gentzkow, and Chuan Yu (2022) “Selection with Variation in Diagnostic Skill: Evidence from Radiologists,” *Quarterly Journal of Economics*, 137 (2), 729–783.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson (2016) “Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector,” *American Economic Review*, 106 (8), 2110–2144.

Chang, Tom, Mireille Jacobson, Manisha Shah, Rajiv Pramanik, and Samir B. Shah (2021) “Financial Incentives and Other Nudges Do Not Increase COVID-19 Vaccinations among the Vaccine Hesitant,” Working Paper.

Cheng, Lingguo, Hong Liu, Ye Zhang, and Zhong Zhao (2018) “The Health Implications of Social Pensions: Evidence from China’s New Rural Pension Scheme,” *Journal of Comparative Economics*, 46 (1), 53–77.

Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas

Turner, Augustin Bergeron, and David Cutler (2016) "The Association between Income and Life Expectancy in the United States, 2001-2014," *JAMA*, 315 (16), 1750–1766.

Christiansen, Cindy L and Carl N Morris (1997) "Improving the Statistical Approach to Health Care Provider Profiling," *Annals of Internal Medicine*, 127, 764–768.

Chyn, Eric (2018) "Moved to Opportunity: The Long-Run Effects of Public Housing Demolition on Children," *American Economic Review*, 108 (10), 3028–3056.

Coile, Courtney, Mark Duggan, and Audrey Guo (2021) "To Work for Yourself, for Others, or Not at All? How Disability Benefits Affect the Employment Decisions of Older Veterans," *Journal of Policy Analysis and Management*, 40 (3), 686–714.

Cornelissen, Thomas, Christian Dustmann, Anna Raute, and Uta Schönberg (2018) "Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance," *Journal of Political Economy*, 126 (6), 2356–2409.

Cryder, Cynthia E., Alex John London, Kevin G. Volpp, and George Loewenstein (2010) "Informative Inducement: Study Payment as a Signal of Risk," *Social Science and Medicine*, 70 (3), 455–464.

Currie, Janet (2009) "Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development," *Journal of Economic Literature*, 47 (1), 87–122.

Cutler, David M, Wei Huang, and Adriana Lleras-Muney (2016) "Economic Conditions and Mortality: Evidence from 200 Years of Data,"Technical report, National Bureau of Economic Research.

Cutler, David M., Adriana Lleras-Muney, and Tom Vogl (2012) *Socioeconomic Status and Health: Dimensions and Mechanisms*: Oxford University Press.

De Chaisemartin, Clement (2017) "Tolerating Defiance? Local Average Treatment Effects Without Monotonicity," *Quantitative Economics*, 8 (2), 367–396.

Deaton, Angus (2002) "Policy Implications of the Gradient of Health and Wealth," *Health Affairs*, 21 (2), 13–30.

Deci, Edward L., Richard Koestner, and Richard M. Ryan (1999) "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation," *Psychological Bulletin*, 125 (6).

Depalo, Domenico (2020) "Explaining the Causal Effect of Adherence to Medication on Cholesterol Through the Marginal Patient," *Health Economics*, 29, 110–126.

Deryugina, Tatyana and David Molitor (2020) "Does When You Die Depend on Where You Live? Evidence from Hurricane Katrina," *American Economic Review*, 110 (11), 3602–3633.

Deshpande, Manasi (2016) "Does Welfare Inhibit Success? The Long-Term Effects of Removing Low-Income Youth from the Disability Rolls," *American Economic Review*, 106 (11), 3300–3330.

Deshpande, Manasi and Lee M. Lockwood (2022) "Beyond Health: Nonhealth Risk and the Value of Disability Insurance," *Econometrica*, 90 (4), 1781–1810.

Deshpande, Manasi and Michael Mueller-Smith (2022) "Does Welfare Prevent Crime? The Criminal Justice Outcomes of Youth Removed from SSI," *Quarterly Journal of Economics*, 137 (4), 2263–2307.

Diggle, Peter J, Patrick Heagerty, Kung-Yee Liang et al. (2002) *Analysis of Longitudinal Data*, Oxford, UK: Oxford University Press.

Doyle Jr, Joseph J (2007) "Child Protection and Child Outcomes: Measuring the Effects of Foster Care," *American Economic Review*, 97 (5), 1583–1610.

——— (2008) "Child Protection and Adult Crime: Using Investigator Assignment to Estimate Causal Effects of Foster Care," *Journal of Political Economy*, 116 (4), 746–770.

Duggan, Mark, Robert Rosenheck, and Perry Singleton (2010) "Federal Policy and the Rise in Disability Enrollment: Evidence for the Veterans Affairs' Disability Compensation Program," *Journal of Law and Economics*, 53 (2), 379–398.

Dwyer, Ryan J and Elizabeth W Dunn (2022) "Wealth Redistribution Promotes Happiness," *Proceedings of the National Academy of Sciences*, 119 (46), e2211123119.

Efron, Bradley and Carl Morris (1973) "Stein's Estimation Rule and Its Competitors — An Empirical Bayes Approach," *Journal of the American Statistical Association*, 68 (34), 117–130.

Einav, Liran, Amy Finkelstein, Yunan Ji, and Neale Mahoney (2022) "Voluntary Regulation: Evidence from Medicare Payment Reform," *Quarterly Journal of Economics*, 137 (1), 565–618.

Eli, Shari J, Trevon D Logan, and Boriana Miloucheva (2023) "The Enduring Effects of Racial Discrimination on Income and Health," *Journal of Economic Literature*, 61 (3), 924–940.

Elixhauser, Anne, Claudia Steiner, D Robert Harris, and Rosanna M Coffey (1998) "Comorbidity Measures for Use with Administrative Data," *Medical Care*, 8–27.

Ellis, Randall P, Bruno Martins, and Wenjia Zhu (2017) "Health Care Demand Elasticities by Type of Service," *Journal of Health Economics*, 55, 232–243.

Ensor, Tim and Stephanie Cooper (2004) "Overcoming Barriers to Health Service Access: Influencing the Demand Side," *Health Policy and Planning*, 19 (4), 69–79.

Eren, Ozkan and Naci Mocan (2021) "Juvenile Punishment, High School Graduation, and Adult Crime: Evidence from Idiosyncratic Judge Harshness," *Review of Economics and Statistics*, 103 (1), 34–47.

Evans, William N, Sarah Kroeger, Caroline Palmer, and Emily Pohl (2019) "Housing and Urban Development–Veterans Affairs Supportive Housing Vouchers and Veterans' Homelessness, 2007–2017," *American Journal of Public Health*, 109 (10), 1440–1445.

Feeney, Kevin (2018) "Dying for More: Cash Transfers and Health,"Technical report, Mimeo.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams (2021) "Place-based Drivers of Mortality: Evidence from Migration," *American Economic Review*, 111 (8), 2697–2735.

Finkelstein, Amy, Matthew J Notowidigdo Frank Schilbach, and Jonathan Zhang (2023) "The Impact of the Great Recession on Health and Welfare."

Forman-Hoffman, Valerie L., Kimberly L. Ault, Wayne L. Anderson, Joshua M. Weiner, Alissa Stevens, Vincent A. Campbell, and Brian S. Armour (2015) "Disability Status, Mortality, and Leading Causes of Death in the United States Community Population," *Medical care*, 53 (4), 346.

Frandsen, Brigham, Lars Lefgren, and Emily Leslie (2023) "Judging Judge Fixed Effects," *American Economic Review*, 113 (1), 253–277.

French, Eric and Jae Song (2014) "The Effect of Disability Insurance Receipt on Labor Supply," *American Economic Journal: Economic Policy*, 6 (2), 291–337.

Frey, Bruno S. and Reto Jegen (2001) "Motivation Crowding Theory," *Journal of Economic Surveys*, 15 (5), 589–611.

Frey, Bruno S. and Felix Oberholzer-Gee (1997) "The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out," *American Economic Review*, 87 (4), 745–755.

García-Gómez, Pilar and Anne C. Gielen (2018) "Mortality Effects of Containing Moral Hazard: Evidence from Disability Insurance Reform," *Health Economics*, 27 (3), 606–621, 10.1002/hec.3635.

Garin, Andrew, Dmitri Koustas, Carl McPherson, Samuel Norris, Matthew Pecenco, Evan K Rose, Yotam Shem-Tov, and Jeffrey Weaver (2023) "The Impact of Incarceration on Employment, Earnings, and Tax Filing," *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2023-108).

Gee, Gilbert C. and Chandra L. Ford (2011) "Structural Racism and Health Inequities: Old Issues, New Directions," *Du Bois Review: Social Science Research on Race*, 8 (1), 115–132.

Gelber, Alexander, Timothy J Moore, and Alexander Strand (2017) "The Effect of Disability Insurance Payments on Beneficiaries' Earnings," *American Economic Journal: Economic Policy*, 9 (3), 229–261.

Gelber, Alexander, Timothy Moore, Zhuan Pei, and Alexander Strand (2023) "Disability Insurance Income Saves Lives," *Journal of Political Economy*.

Geocze, Luciana, Samantha Mucci, Denise Caluta Abranches, Mario Alfredo de Marco, and Norma de Oliveira Penido (2013) "Systematic Review on the Evidences of an Association Between Tinnitus and Depression," *Brazilian Journal of Otorhinolaryngology*, 79 (1), 106–111.

Gneezy, Uri, Stephan Meier, and Pedro Rey-Biel (2011) "When and Why Incentives (Don't) Work to Modify Behavior," *Journal of Economic Perspectives*, 25 (4), 191–210.

Gneezy, Uri and Aldo Rustichini (2000) "Pay Enough or Don't Pay at All," *Quarterly Journal of Economics*, 115 (3), 791–810.

Government Accountability Office (2022) "VA Disability Benefits: Compensation Program Could Be Strengthened by Consistently Following Leading Reform Practices,"Technical

Report GAO-22-104488, United States Government Accountability Office, Washington, DC.

Grace, Anthony R, Rosemary Briggs, Ruth E Kieran et al. (2014) "A Comparison of Beers and STOPP Criteria in Assessing Potentially Inappropriate Medications in Nursing Home Residents Attending the Emergency Department," *Journal of the American Medical Directors Association*, 15, 830–834.

Gross, Max and E Jason Baron (2022) "Temporary Stays and Persistent Gains: The Causal Effects of Foster Care," *American Economic Journal: Applied Economics*, 14 (2), 170–199.

Grossman, Michael (1972) "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, 80 (2), 223–255.

Gruber, Jonathan and Jeffrey D Kubik (1997) "Disability Insurance Rejection Rates and the Labor Supply of Older Workers," *Journal of Public Economics*, 64 (1), 1–23.

Guldi, Melanie, Amelia Hawkins, Jeffrey Hemmeter, and Lucie Schmidt (2022) "Supplemental Security Income for Children, Maternal Labor Supply, and Family Well-Being: Evidence from Birth Weight Eligibility Cutoffs," *Journal of Human Resources*.

Hansen, Lars Peter (1982) "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica: Journal of the Econometric Society*, 1029–1054.

Heckman, James (1997) "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *Journal of Human Resources*, 441–462.

Heckman, James J, John Eric Humphries, and Gregory Veramendi (2018) "Returns to Education: The Causal Effects of Education on Earnings, Health, and Smoking," *Journal of Political Economy*, 126 (S1), S197–S246.

Heckman, James J. and Rodrigo Pinto (2018) "Unordered Monotonicity," *Econometrica*, 86 (1), 1–35.

——— (2021) "The Econometric Model for Causal Policy Analysis," *Annual Review of Economics*.

Heckman, James J, Sergio Urzua, and Edward Vytlacil (2006) "Understanding Instrumental Variables in Models with Essential Heterogeneity," *Review of Economics and Statistics*, 88 (3), 389–432.

Heckman, James J and Edward Vytlacil (2005) "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73 (3), 669–738.

Heckman, James J and Edward J Vytlacil (2007) "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation," *Handbook of Econometrics*, 6, 4779–4874.

Hilmer, Sarah N and Danijela Gnjidic (2009) "The Effects of Polypharmacy in Older Adults," *Clinical Pharmacology & Therapeutics*, 85, 86–88.

Honoré, Bo E. and Adriana Lleras-Muney (2006) "Bounds in Competing Risks Models and the War on Cancer," *Econometrica*, 74 (6), 1675–1698.

Horowitz, Joel L and Charles F Manski (2000) "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data," *Journal of the American Statistical Association*, 95 (449), 77–84.

Huber, Martin and Giovanni Mellace (2012) "Relaxing Monotonicity in the Identification of Local Average Treatment Effects," Working Paper.

Iglehart, John K (2013) "Expanding the Role of Advanced Nurse Practitioners—Risks and Rewards," *New England Journal of Medicine*, 368, 1935–1941.

Imbens, Guido W and Joshua D Angrist (1994) "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62 (2), 467–475.

IMF (2019) "A Strategy for IMF Engagement on Social Spending," *International Monetary Fund Policy Paper*, 19 (016).

Institute of Medicine (2011) *The Future of Nursing: Leading Change, Advancing Health. Chapter 3: Transforming Practice*, Washington, D.C.: National Academies Press.

Ito, Koichiro, Takanori Ida, and Makoto Tanaka (2023) "Selection on Welfare Gains: Experimental Evidence from Electricity Plan Choice," *American Economic Review*, 113 (11), 2937–2973.

Jano, Elizabeth and Rajender R Aparasu (2007) "Healthcare Outcomes Associated with Beers' Criteria: A Systematic Review," *Annals of Pharmacotherapy*, 41, 438–447.

Jarrett, Caitlin, Rose Wilson, Maureen O'Leary, Elisabeth Eckersberger, and Heidi J. Larson (2015) "Strategies for Addressing Vaccine Hesitancy – A Systematic Review," *Vaccine*, 33 (34), 4180–4190.

Jensen, Robert T and Kaspar Richter (2004) "The Health Implications of Social Security Failure: Evidence from the Russian Pension Crisis," *Journal of Public Economics*, 88 (1-2), 209–236.

Jiao, Sai, Ian B Murimi, Randall S Stafford et al. (2018) "Quality of Prescribing by Physicians, Nurse Practitioners, and Physician Assistants in the United States," *Pharmacotherapy*, 38, 417–427.

Kaiser Family Foundation (2022) "Total Number of Nurse Practitioners," https://kff.org/other/state-indicator/total-number-of-nurse-practitioners.

Kangas, Olli, S Flour, Miska Simanainen, and M Ylikanno (2020) "Evaluation of the Finnish Basic Income Experiment," *Reports and Memoranda from the Ministry of Social Affairs and Health*.

Kim, Seonghoon and Kanghyock Koh (2021) “The Effects of Income on Health: Evidence from Lottery Wins in Singapore,” *Journal of Health Economics*, 76, 102414.

Kitagawa, Toru (2015) “A Test for Instrument Validity,” *Econometrica*, 83 (5), 2043–2063.

Kline, Patrick and Melissa Tartari (2016) “Bounding the Labor Supply Responses to a Randomized Welfare Experiment: A Revealed Preference Approach,” *American Economic Review*, 106 (4), 972–1014.

Kline, Patrick and Christopher R Walters (2016) “Evaluating Public Programs with Close Substitutes: The Case of Head Start,” *Quarterly Journal of Economics*, 131 (4), 1795–1848.

Kline, Rex B. (2015) *Principles and Practice of Structural Equation Modeling*: Guilford.

Kling, Jeffrey R. (2006) “Incarceration Length, Employment, and Earnings,” *American Economic Review*, 96 (3), 863–876.

Kotwal, Ashwin A, Michael A Steinman, Irena Cenzer et al. (2021) “Use of High-Risk Medications Among Lonely Older Adults: Results from a Nationally Representative Sample,” *JAMA Internal Medicine*, 181, 1528–1530.

Kowalski, Amanda (2020) “Counting Defiers: Examples from Health Care,” Working Paper.

Kowalski, Amanda E (2023) “Behaviour within a Clinical Trial and Implications for Mammography Guidelines,” *Review of Economic Studies*, 90 (1), 432–462.

Krahn, Gloria L., Deborah Klein Walker, and Rosaly Correa-De-Araujo (2015) “Persons with Disabilities as an Unrecognized Health Disparity Population,” *American Journal of Public Health*, 105 (S2), S198–S206.

Krouse, Sarah (2021) “Dollar General Will Pay Workers to Get a COVID-19 Vaccine,” *The Wall Street Journal*.

Kuper, Hannah, Phyllis Heydt, Céleste Danos, Michelle Burbach, Sara Rotenberg, and Frederic Seghers (2022) "Reimagining Health Systems that Expect, Accept and Connect 1 Billion People with Disabilities," *MBI and CHAI*.

Laffont, Jean-Jacques and David Martimort (2001) *The Theory of Incentives: The Principal-Agent Model*: Princeton University Press.

Lang, Matthew, T. Clay McManus, and Georg Schaur (2019) "The Effects of Import Competition on Health in the Local Economy," *Health Economics*, 28 (1), 44–56.

Laurant, Miranda, Mieke van der Biezen, Nancy Wijers et al. (2018) "Nurses as Substitutes for Doctors in Primary Care," *Cochrane Database of Systematic Reviews*, 7 (7), CD001271.

Law, Anica C., Daniel Peterson, Allan J. Walkey, and Nicholas A. Bosch (2022) "Lottery-Based Incentives and COVID-19 Vaccination Rates in the US," *JAMA Internal Medicine*, 182 (2), 235–237.

Lee, David S, Justin McCrary, Marcelo J Moreira, and Jack Porter (2022) "Valid T-Ratio Inference for IV," *American Economic Review*, 112 (10), 3260–3290.

Lee, Sokbae and Bernard Salanié (2018) "Identifying Effects of Multivalued Treatments," *Econometrica*, 86 (6), 1939–1963.

Lewis, John E, SDG Stephens, and L McKenna (1994) "Tinnitus and Suicide," *Clinical Otolaryngology & Allied Sciences*, 19 (1), 50–54.

Liebman, Jeffrey B. (2015) "Understanding the Increase in Disability Insurance Benefit Receipt in the United States," *Journal of Economic Perspectives*, 29 (2), 123–150.

Lieder, Falk, Thomas L. Griffiths, Quentin J.M. Huys, and Noah D. Goodman (2018) "The Anchoring Bias Reflects Rational Use of Cognitive Resources," *Psychonomic Bulletin and Review*, 25 (1), 322–349.

Lin, Cheryl, Pikuei Tu, and Leslie M. Beitsch (2020) “Confidence and Receptivity for COVID-19 Vaccines: A Rapid Systematic Review,” *Vaccines*, 9 (1), 16.

Lindahl, Mikael (2005) “Estimating the Effect of Income on Health and Mortality Using Lottery Prizes as an Exogenous Source of Variation in Income,” *Journal of Human Resources*, 40 (1), 144–168.

Lindqvist, Erik, Robert Östling, and David Cesarini (2020) “Long-Run Effects of Lottery Wealth on Psychological Well-being,” *Review of Economic Studies*, 87 (6), 2703–2726.

Litan, Robert E. (2020) “Want Herd Immunity? Pay People to Take the Vaccine,” *Brookings Institution*.

Lleras-Muney, Adriana and Flavien Moreau (2022) “A Unified Model of Cohort Mortality,” *Demography*, 59 (6), 2109–2134.

Lleras-Muney, Adriana (2022) “Education and income gradients in longevity: The role of policy,” *Canadian Journal of Economics/Revue canadienne d’économique*, 55 (1), 5–37.

Loewenstein, George and Cynthia Cryder (2020) “Why Paying People to Be Vaccinated Could Backfire,” *The New York Times*.

Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand (2013) “Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt,” *American Economic Review*, 103 (5), 1797–1829.

Maestas, Nicole, Kathleen J Mullen, and Alexander Strand (2021) “The Effect of Economic Conditions on the Disability Insurance Program: Evidence from the Great Recession,” *Journal of Public Economics*, 199, 104410.

Malavasi, Chiara and Han Ye (2023) “The Impact of Pension Income on Mortality: Evidence from Germany,” Manuscript.

Mankiw, Gregory N. (2020) "Pay People to Get Vaccinated," *The New York Times*.

Mantzari, Eleni, Florian Vogt, , and Theresa Marteau (2015) "Financial Incentives for Increasing Uptake of HPV Vaccinations: A Randomized Controlled Trial," *Health Psychology*, 34 (2).

Marmot, Michael G., Geoffrey Rose, Martin Shipley, and Peter J. Hamilton (1978) "Employment Grade and Coronary Heart Disease in British Civil Servants," *Journal of Epidemiology Community Health*, 32 (4), 244–249.

Maynard, Charles, Ranak Trivedi, Karin Nelson, and Stephan D Fihn (2018) "Disability rating, age at death, and cause of death in US veterans with service-connected conditions," *Military Medicine*, 183 (11-12), e371–e376.

McGuire, Joel, Caspar Kaiser, and Anders M Bach-Mortensen (2022) "A Systematic Review and Meta-analysis of the Impact of Cash Transfers on Subjective Well-being and Mental Health in Low-and Middle-income Countries," *Nature Human Behaviour*, 6 (3), 359–370.

McMichael, Benjamin J (2021a) "The Access-to-Care Epidemic," *Wake Forest Law Review*, 56, 547.

——— (2021b) "Nurse Practitioner Scope-of-Practice Laws and Opioid Prescribing," *Milbank Quarterly*, 99, 721–745.

McMichael, Benjamin J and Sara Markowitz (2023) "Toward a Uniform Classification of Nurse Practitioner Scope of Practice Laws," *Medical Care Research and Review*, 80 (4), 444–454.

Mellström, Carl and Magnus Johannesson (2008) "Crowding Out in Blood Donation: Was Titmuss Right?" *Journal of the European Economic Association*, 6 (4), 845–863.

Meseguer, Javier (2021) "How Does Mortality Among Disability-Program Beneficiaries Compare with That of the General Population? A Summary of Actuarial Estimates," *Soc. Sec. Bull.*, 81, 19.

Meyer, Bruce D, Angela Wyse, Alexa Grunwaldt, Carla Medalia, and Derek Wu (2021) "Learning About Homelessness Using Linked Survey and Administrative Data,"Technical report, National Bureau of Economic Research.

Miglino, Enrico, Nicolás Navarrete H, Gonzalo Navarrete H, and Pablo Navarrete H (2023) "Health Effects of Increasing Income for the Elderly: Evidence from a Chilean Pension Program," *American Economic Journal: Economic Policy*, 15 (1), 370–393.

Milgrom, Paul and John Roberts (1986) "Price and Advertising Signals of Product Quality," *Journal of Political Economy*, 94 (4), 796–821.

Mogstad, Magne, Andres Santos, and Alexander Torgovitsky (2018a) "Using instrumental variables for inference about policy relevant treatment parameters," *Econometrica*, 86 (5), 1589–1619.

——— (2018b) "Using Instrumental Variables for Inference about Policy Relevant Treatment Parameters," *Econometrica*, 86 (5), 1589–1619.

Mogstad, Magne, Alexander Torgovitsky, and Christopher R Walters (2020) "Policy Evaluation with Multiple Instrumental Variables,"Technical report, National Bureau of Economic Research.

Mogstad, Magne, Alexander Torgovitsky, and Christopher R. Walters (2021) "The Causal Interpretation of Two-Stage Least Squares With Multiple Instrumental Variables," *American Economic Review*, 111 (11), 3663–3698.

Morris, Carl N (1983) "Parametric Empirical Bayes Inference: Theory and Applications," *Journal of the American Statistical Association*, 78 (381), 47–55.

Motter, Fabrício Renato, Juliana S Fritzen, Sarah N Hilmer et al. (2018) "Potentially Inappropriate Medication in the Elderly: A Systematic Review of Validated Explicit Criteria," *European Journal of Clinical Pharmacology*, 74, 679–700.

Mourifié, Ismael and Yuanyuan Wan (2017) "Testing Local Average Treatment Effect Assumptions," *Review of Economics and Statistics*, 99 (2), 305–313.

Mundinger, Mary O, Rosalie L Kane, Elizabeth R Lenz et al. (2000) "Primary Care Outcomes in Patients Treated by Nurse Practitioners or Physicians: A Randomized Trial," *Journal of the American Medical Association*, 283, 59–68.

National Conference of State Legislatures (2017) "Scope of Practice Archive Database," https://www.ncsl.org/research/health/scope-of-practice-overview.aspx#Database.

National Council of State Boards of Nursing (2018) "The Nursing Regulatory Environment in 2018: Issues and Challenges," *Journal of Nursing Regulation*, 9, 52–67.

Normand, Sharon-Lise T, Arlene S Ash, Stephen E Fienberg et al. (2016) "League Tables for Hospital Comparisons," *Annual Review of Statistics and Its Application*, 3, 21–50.

Norris, Samuel, Matthew Pecenco, and Jeffrey Weaver (2021) "The Effects of Parental and Sibling Incarceration: Evidence from Ohio," *American Economic Review*, 111 (9), 2926–2963.

——— (2022) "The Effect of Incarceration on Mortality," *Review of Economics and Statistics*, 1–45.

Nybom, Martin (2017) "The Distribution of Lifetime Earnings Returns to College," *Journal of Labor Economics*, 35 (4), 903–952.

Oster, Emily (2019) "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business & Economic Statistics*, 37, 187–204.

Pierce, Justin R. and Peter K. Schott (2020) "Trade Liberalization and Mortality: Evidence from US Counties," *American Economic Review: Insights*, 2 (1), 47–63.

Pigou, Arthur Cecil (1920) *The Economics of Welfare*: London: Macmillan.

Pinto, Rodrigo (2021) "Beyond Intention to Treat: Using the Incentives in Moving to Opportunity to Identify Neighborhood Effects," Working Paper.

Pohl, Joanne M, Corey Hanson, Jason A Newland et al. (2010) "Unleashing Nurse Practitioners' Potential to Deliver Primary Care and Lead Teams," *Health Affairs*, 29, 900–905.

Pope, Gregory C, John Kautter, Randall P Ellis et al. (2004) "Risk Adjustment of Medicare Capitation Payments Using the CMS-HCC Model," *Health Care Financing Review*, 25, 119–141.

Rankin, Audrey, Cathal A Cadogan, Susan M Patterson et al. (2018) "Interventions to Improve the Appropriate Use of Polypharmacy for Older People," *Cochrane Database of Systematic Reviews*, 9, CD008165.

Reichard, Amanda, Hayley Stolzle, and Michael H. Fox (2011) "Health Disparities Among Adults with Physical Disabilities or Cognitive Limitations Compared to Individuals with No Disabilities in the United States," *Disability and Health Journal*, 4 (2), 59–67.

Robeznieks, Andis (2022) "Inside the AMA's Wide-Ranging Fight Against Scope Creep," https://www.ama-assn.org/practice-management/scope-practice/inside-ama-s-wide-ranging-fight-against-scope-creep.

Ruhm, Christopher J (2000) "Are Recessions Good for Your Health?" *Quarterly Journal of Economics*, 115 (2), 617–650.

Salm, Martin (2011) "The Effect of Pensions on Longevity: Evidence from Union Army Veterans," *The Economic Journal*, 121 (552), 595–619.

Sargan, John D (1958) "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica: Journal of the Econometric Society*, 393–415.

Sato, Ryoko and Benjamin Fintan (2020) "Effect of Cash Incentives on Tetanus Toxoid Vaccination Among Rural Nigerian Women: A Randomized Controlled Trial," *Human Vaccines and Immunotherapeutics*, 16 (5), 1181–1188.

Schwandt, Hannes and Till M von Wachter (2020) "Socioeconomic decline and death: Midlife impacts of graduating in a recession,"Technical report, National Bureau of Economic Research.

Silver, David and Jonathan Zhang (2023) "Invisible Wounds: Health and Well-Being Impacts of Mental Disorder Disability Compensation on Veterans," *National Bureau of Economic Research* (w29877).

Smythe, Tracey and Hannah Kuper (2023) "The Association Between Disability and All-Cause Mortality in Low and Middle-Income Countries: A Systematic Review and Meta-Analysis," *medRxiv*, https://doi.org/10.1101/2023.03.21.23287520.

Snyder, Stephen E. and William N. Evans (2006) "The Effect of Income on Mortality: Evidence from the Social Security Notch," *Review of Economics and Statistics*, 88 (3), 482–495.

Sohn, Min-Woong, Noreen Arnold, Charles Maynard, and Denise M Hynes (2006) "Accuracy and Completeness of Mortality Data in the Department of Veterans Affairs," *Population Health Metrics*, 4 (1), 1–8.

Staff, MedPAC (2019) "Improving Medicare's payment policies for advanced practice registered nurses and physician assistants," *MedPAC Blog*.

Stoler, Justin, Adam M. Enders, Casey A. Klofstad, and Joseph E. Uscinski (2021) "The

Limits of Medical Trust in Mitigating COVID-19 Vaccine Hesitancy among Black Americans," *Journal of General Internal Medicine*, 36 (11), 3629–3631.

Thirumurthy, Harsha, Katherine L. Milkman, Kevin G. Volpp, Alison M. Buttenheim, and Devin G. Pope (2022) "Association between Statewide Financial Incentive Programs and COVID-19 Vaccination Rates," *PloS One*, 17 (3).

Thompson, Hayley, Mark Manning, Jamie Mitchell, Seongho Kim, Felicity Harper, Sheena Cresswell, Kristopher Johns, and et al. (2021) "Factors Associated With Racial/Ethnic Group–Based Medical Mistrust and Perspectives on COVID-19 Vaccine Trial Participation and Vaccine Uptake in the US," *JAMA Network Open*, 4 (5), E2111629–E2111629.

Thornton, Rebecca L. (2008) "The Demand For, and Impact of, Learning HIV Status," *American Economic Review*, 98 (5), 1829–1863.

Titmuss, Richard (1970) *The Gift Relationship: From Human Blood to Social Policy*: George Allen and Unwin.

Tressler, Stacy and Ruchi Bhandari (2019) "Interventions to Increase Completion of Hepatitis B Vaccination in People Who Inject Drugs: A Systematic Review and Meta-Analysis," *Open Forum Infectious Diseases*, 6 (12).

Tulchinsky, Theodore H., Elena A. Varavikova, and Matan J. Cohen (2022) *New Public Health*: Elsevier Academic Press.

Tversky, Amos and Daniel Kahneman (1974) "Judgment Under Uncertainty: Heuristics and Biases," *Science*, 185 (4157), 1124–1131.

United Nations (2018) *Persons With Disabilities: Breaking Down Barriers*, 63–76: United Nations.

U.S. Chamber of Commerce (2021) "Vaccination Incentives," *U.S. Chamber of Commerce: Emerging Issues*.

Vohs, Kathleen D., Nicole L. Mead, and Miranda R. Goode (2006) "The Psychological Consequences of Money," *Science*, 314 (5802), 1154–1156.

Von Wachter, Till von, Jae Song, and Joyce Manchester (2011) "Trends in Employment and Earnings of Allowed and Rejected Applicants to the Social Security Disability Insurance Program," *American Economic Review*, 101 (7), 3308–3329.

Walters, Christopher R (2018) "The Demand for Effective Charter Schools," *Journal of Political Economy*, 126 (6), 2179–2223.

Wigham, Sarah, Laura Ternent, Andrew Bryant, Shannon Robalino, Falko Sniehotta, and Jean Adams (2014) "Parental Financial Incentives for Increasing Preschool Vaccination Uptake: Systematic Review," *Pediatrics*, 134 (4), 1117–1128.

World Health Organization (2022) *Global Report on Health Equity for Persons with Disabilities*, Geneva, Switzerland: World Health Organization.

Xue, Ying, Zhiqiang Ye, Christine Brewer et al. (2016) "Impact of State Nurse Practitioner Scope-of-Practice Regulation on Health Care Delivery: Systematic Review," *Nursing Outlook*, 64, 71–85.

Yang, Bicheng K, Meg Johantgen, Alison M Trinkoff et al. (2021) "State Nurse Practitioner Practice Regulations and US Health Care Delivery Outcomes: A Systematic Review," *Medical Care Research and Review*, 78, 183–196.

Yang, Yu, Qi Long, Sandra L Jackson et al. (2018) "Nurse Practitioners, Physician Assistants, and Physicians Are Comparable in Managing the First Five Years of Diabetes," *American Journal of Medicine*, 131, 276–283.

Yao, Xiaoxi, Allard E Dembe, Thomas Wickizer, and Bo Lu (2015) "Does Time Pressure Create Barriers for People to Receive Preventive Health Services?" *Preventive Medicine*, 74, 55–58.

Zayatz, Tim A (2015) *Social security Disability Insurance Program Worker Experience* (123): Social Security Administration, Office of the Chief Actuary.

Zimmerman, Annie, Emily Garman, Mauricio Avendano-Pabon et al. (2021) “The Impact of Cash Transfers on Mental Health in Children and Young People in Low-income and Middle-income Countries: A Systematic Review and Meta-analysis,” *BMJ Global Health*, 6 (4), e004661.