ISBN

9798189270499

Author

Costantini, Sydney

Publication Date

2026-05-01

Peer reviewed|Thesis/dissertation

Essays in Applied Microeconomics

by

Sydney Costantini

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

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David Card, Chair
Professor David Chan
Professor Patrick Kline
Professor Christopher Walters

Spring 2026

Essays in Applied Microeconomics

Abstract

Essays in Applied Microeconomics

by Sydney Costantini

Doctor of Philosophy in Economics
University of California, Berkeley

Professor David Card, Chair

This dissertation presents three essays in applied microeconomics, studying how policy interventions and information affect health and political outcomes.

In Chapter 1, "Housing First or Treatment First? Evidence from the VA's Homelessness Programs," I study the impact of Housing First programs, in which subsidized housing is offered unconditionally and immediately, relative to Treatment First interventions, which combine short-term housing with treatment for issues like substance use. I construct a national sample of nearly 300,000 unhoused, mentally ill Veterans potentially eligible for the VA's Housing First program (HUD-VASH) or a Treatment First program. Using an instrumental variables design, I find that enrolling in Housing First reduces three-year mortality by 4.6 percentage points relative to a no-program counterfactual. In contrast, Treatment First has no long-term effects on health. Housing First is also cost-saving, as it causes individuals to substitute away from lengthy inpatient stays.

Chapter 2, "How do mental health treatment delays impact long term mortality?," studies how clinic congestion affects mortality for veterans experiencing mental health emergencies. I find that longer waiting times make it more likely that patients miss their follow-up mental health visit, increasing the probability that they permanently disengage from care. A one standard deviation increase in wait time between the ED visit and follow-up appointment (11.7 days) increases two-year mortality by about 1.5%.

Chapter 3, "The Effect of Red Scare Propaganda on US Public Opinion: Evidence from the Newspapers of William Randolph Hearst," examines whether anti-redistributive sentiment has long-run roots in Red Scare-era propaganda. In the 1930s and 40s, the media mogul William Randolph Hearst espoused vehement anti-communist rhetoric in his newspapers. Fears of socialism went head to head with FDR's New Deal, the first wide-scale federal spending program.Using variation in newspaper penetration across US counties, I show that a 10 percentage point increase in Hearst newspaper circulation per capita decreased support for FDR by 2.1 percentage points. These effects persist over the long run, decreasing support for public healthcare and welfare programs through the late 20th century.

Contents

Acknowledgments

I have many people to thank who both made this dissertation possible and made my time at UC Berkeley so meaningful.

First and foremost, I want to thank my dissertation advisor, David Card, who is largely responsible for making my time in the Berkeley Economics Department so great. Dave not only taught me how to write an economics paper—working closely with me starting from my first-year summer—but also took the time to get to know who I was. Our meetings often began with long conversations before turning to research (for Dave, even an hour counted as a "short" meeting). I am incredibly grateful not only for his insight and guidance, but for how deeply he cares about his students' success. Dave made me believe I could be an economist, and I can only hope to pay that forward.

I have many other mentors and advisors to thank who were vital in helping me grow as an economist. I thank Dave Chan for introducing me to VA data and demonstrating how to write a good health economics paper; Christopher Walters for imparting his econometrics wisdom, both in and outside of class; and Pat Kline, who really took me under his wing and knew exactly how to make my job market paper the best version it could be. I am also grateful to Guo Xu for encouraging me to keep working on my economic history paper and for his thoughtful suggestions; to Emmanuel Saez for helping get my job market paper started in Public Therapy; and to Ricardo Perez-Truglia and Jonathan Kolstad for fostering the Berkeley economics community through soccer. I also want to thank my labor economics buddies and office mates—Dustin, Marco, Dani, and Isa—who made the Berkeley community welcoming and collaborative (often through shared suffering). Moreover, I want to thank Sarah Baker and Kaveh Danesh, who were my older student mentors in the Department and gave me a lot of great advice (as well as let me complain).

Last but not least, I want to thank my family and friends, who made my time in Berkeley so fun and kept me grounded throughout graduate school. I am especially grateful to my fellow PhD students: Suvy, for hosting Survivor nights; Margaret, for tennis and reality TV; Petra, for always finding cool events/talks/restaurants in San Francisco; and "marginally social people"—Sharath, Karthik, Lucio, and Nathaniel—for many happy hours and wine tastings. I also want to thank my non-economics friends: Booty Booty Rides, for making soccer (and carpooling) so fun, and especially Gabe, for listening to me ramble about every personal life saga. Finally, I want to thank my family: my sister Julia (and Yanni), for following me to California, and of course my parents and Helena, for supporting me from afar and helping me navigate the many ups and downs of graduate school.

Chapter 1

Housing First or Treatment First? Evidence from the VA's homelessness programs

1.1 Introduction

The number of unhoused individuals in the US reached a record high of 770,000 at the end of 2024 (Porter 2024). The visible presence of homeless people – many seemingly dealing with mental health and/or substance abuse issues – is a source of vehement debate. Some advocates argue that homelessness is, in fact, a housing problem, with permanent housing solutions required to help those most in need. Others argue that those who are chronically unhoused have underlying mental health or substance use issues that must be addressed first in order to make progress in reducing homelessness.

In the 1990s, homelessness advocates developed the concept of Housing First, or that for those most in need, highly-subsidized permanent housing should be given unconditionally and immeditely. The philosophy relies on the idea that requiring pre-conditions to housing merely acts as a barrier to seeking care, and that it is hard to address underlying mental health issues without having a stable home first.

Housing First programs are pitted against more traditional programs that emphasize housing readiness. In these models, individuals generally enter transitional housing, or temporary communal housing in which substance use and mental health treatment are required, sometimes along with sobriety. After a stay in these transitional homes, individuals eventually are deemed ready to enter regular, generally unsubsidized housing. In this paper, I call transitional housing programs "Treatment First" for simplicity.

RCTs analyzing Housing First programs have consistently demonstrated increases in housing stability compared to outcomes for individuals not enrolled in the program (e.g.,Tsemberis and Eisenberg 2000, Tsemberis et al. 2004, and Kertesz et al. 2009). But is housing stability the appropriate goal for homelessness policy, or should homelessness programs focus on larger goals of improved health and general well-being? Advocates of Treatment First or Housing Readiness programs argue the latter, and point out that Housing First's effects on mental and physical health remain unclear (NASEM 2018). A related methodological concern is that Housing First studies typically compare a permanent housing subsidy to an ambiguous counterfactual, which may include a combination of entry to Treatment First programs and no housing intervention at all.

The Department of Veterans (VA) affairs is perhaps the ideal setting to shed light on this debate. Veterans have been historically overrepresented in the population of unhoused individuals, and eliminating Veteran homelessness specifically has been an important federal goal (Tsai and Rosenheck 2015). Since the 1990s, the VA, along with the Department of Housing and Urban Development (HUD), have run a joint Housing First program. Known as HUD-VASH (VA Supportive Housing), the program offers housing vouchers to chronically homeless Veterans in the same style as Section 8. Concurrently, the VA has a large transitional housing program – called the Grant and Per Diem (GPD) program – which contracts out third-party agencies, such as the Salvation Army, to run short-term, Treatment First-style housing, where Veterans can stay up to two years.

In this paper, I evaluate the effects of Housing First versus Treatment First using a national sample of nearly 300,000 unhoused Veterans with mental illness. Beginning with their initial intake assessment, I follow these Veterans for three years, tracking program participation and long-term health outcomes. To disentangle the effects of different program types, I explicitly define three treatments: 1) housing voucher (HUD-VASH); 2) transitional housing (GPD); and 3) no housing intervention.

Program enrollment in this setting is the result of a combination of factors: voucher availability, caseworker recommendations, and Veteran choice. Thus, to address the endogeneity of treatment assignment, I use an instrumental variables framework with two treatment options, relative to the counterfactual of no treatment (Kirkeboen et al. 2016, Heinesen et al. 2022, Mountjoy 2022). I consider arrival to the VA homelessness clinic to be the result of a random shock to housing stability. This allows me to exploit two quasi-random sources of variation based on clinic arrival time. First, for a given area and year, permanent housing vouchers are fixed according to HUD allocations. These vouchers are notoriously in short supply, allowing me to exploit the length of the voucher waiting list. Specifically, I define my first instrument as the leave-out percent of individuals who receive a housing voucher in three months or less among those who arrive to the same clinic within a 60-day window of one's intake date. I show that this variable is strongly predictive of the probability a Veteran is assigned to HUD-VASH, while also satisfying a wide battery of orthogonality checks, conditional on time and location controls.

Secondly, when a Veteran arrives to the homelessness clinic, based on both staffing availability and patient needs, they may be assessed by a generalist who triages them into different programs, or they may be seen by a specialist who oversees a specific program they are interested in. In particular, GPD liaisons are in charge of managing referrals to transitional housing. Those immediately seen by a GPD liaison are far more likely to enter transitional housing, likely both due to social worker encouragement and accelerated entry into a particular housing unit. As patients are assigned to caseworkers based on need, I do not use direct social worker assignment; instead, I exploit the availability of a transitional housing liaison at one's time of arrival as an instrument. Specifically, I define my second instrumental variable as an indicator for whether a transitional housing liaison performs any intake assessment in the week of one's arrival at the VA clinic. I show that that this simple indicator is highly predictive of assignment to a transitional housing program, while also satisfying a parallel set of orthogonality checks.

Using these two instrumental variables and a setup with two endogenous treatments, I find that Housing First decreases three-year mortality by 4.6 percentage points (s.e.=1.7) relative to neither program. In contrast, transitional housing has no statistically significant effect on mortality, with point estimates very close to 0. In fact, I can statistically rule out that Treatment First is as effective as Housing First at reducing mortality.

Prior research has shown that 2SLS estimates with multiple endogenous treatments rely on a no "essential heterogeneity" assumption, or treatment effect heterogeneity that is correlated with selection into treatment (Heckman et al. 2006, Kirkeboen et al. 2016, Heinesen et al. 2022). More specifically, to uncover interpretable and well-defined causal effects, complier groups induced into different treatment statuses cannot have heterogeneous average treatment effects for a particular treatment, ruling out models like selection-on-gains and other forms of heterogeneity that are correlated with behavior. I conduct some simple tests to check whether no essential heterogeneity is a good approximation in my setting. First, I show that complier types all have similar mean predicted mortality, implying that compliers who are induced into the voucher program have similar baseline risk as compliers who are induced into transitional housing. Further, I use overidentification tests to rule out treatment effect heterogeneity with respect to other important individual characteristics. Next, I show that heterogeneous treatment effects across locations are uncorrelated with the propensity to be treated, which is inconsistent with a Roy (1951) selection framework. Finally, using worst-case scenario bounds, I show that I estimate coefficients very close to true treatment effects even under small violations to no essential heterogeneity.

Why does Housing First have such a large effect on long term mortality, while Treatment First appears to have no impact? To shed some light on mechanisms, I analyze housing stability over time, engagement with VA homelessness and mental health services, and cause of death. First, as echoed by other research, I find that Housing First increases long term housing stability. In contrast, housing stability for transitional housing recipients starts high but declines over time. Secondly, I find that both programs massively increase engagement with VA social workers. These effects are permanent for the voucher program, and fading over time for transitional housing. Next, I evaluate engagement with mental health care. Following the program philosophy of Treatment First, I find that transitional housing increases engagement with mental health care in the short term; however, effects fade to 0 as Veterans exit the program. In contrast, I find that Housing First has no effect on mental health care utilization, ruling out that mortality effects come from mental health treatment. Lastly, I show that voucher recipients are much less likely to die from extreme heat or cold, violence, and infections caused by risky injection drug use behavior, such as needle sharing. Overall, my results suggest that long term, stable housing improves health, with required mental health treatment not compensating for the effects of permanent housing.

Even if housing improves health, critics have argued that it may not be cost effective. In rent alone, HUD pays an estimated $581 per month for every HUD-VASH voucher (National Homeless Information Project 2018). However, I find that Housing First substantially reduces inpatient length of stays related to homelessness and mental health. Lower inpatient costs exceed program costs, making Housing First cost-saving in the long term.

This paper contributes to both the literature on Housing First, as well as HUD-VASH specifically. Prior randomized control trials and other retrospective studies have found that HUD-VASH leads to more stable housing, but researchers have generally found imprecise results for health outcomes (Rosenheck et al. 2003, Montgomery et al. 2013, Evans et al. 2019). Other Housing First studies for non-Veteran populations have found that offering permanent housing subsidies to the unhoused reduces inpatient stays, ED visits, and incarceration rates, but has little effect on employment (Basu et al. 2012, Stergiopoulos et al. 2015, Aubry et al. 2016, NASEM 2018, Evans et al. 2019, Brounstein and Wieselthier 2024, Cohen 2024, Hanson and Gillespie 2024). Using the largest sample to date to study the causal impact of Housing First, this paper is the first to show that housing reduces long-term mortality.

Additionally, this paper contributes to a scarce literature on transitional housing and other residential programs than emphasize mental health treatment. There is some limited quasi-experimental evidence than transitional housing improves living conditions and reduces inpatient length of stays even after program exit, but no evidence that it improves long term substance use outcomes (Siskind et al. 2013). Expanding beyond unhoused individuals to those with substance use disorders more broadly, residential rehabilitation programs also resemble Treatment First interventions in that they bundle short term housing and behavioral therapy. However, there is also limited empirical evidence that rehab has long-lasting health effects (Reif et al. 2014).

Most notably, this paper contributes to the literature on Housing First versus Treatment First. Prior research typically compares Housing First interventions to "standard of care" treatment methods, which often includes a substantial number of individuals who receive no treatment at all. These studies make it difficult to clearly delineate the impact of a Housing First versus Treatment First intervention. A single RCT, the Family Options study, explicitly compares permanent housing subsidies to transitional housing (Gubits et al. 2018). Unfortunately, the authors can only measure intention to treat effects, and there are low rates of program take up, especially for transitional housing. The authors find that permanent housing again increases housing stability and financial wellbeing, but find no statistically significant impacts on health, which they measure using surveys. This study is the first to explicitly compare average treatment effects of Housing First versus Treatment First.

1.2 Setting

1.2.1 VA Homelessness Services

The Department of Veterans Affairs (VA) operates over 1,000 facilities across the US that provide integrated healthcare and homelessness services to US Veterans. Veterans in need of housing services can drop by any VA medical center or Community Resource and Referral Center – outpatient clinics targeted towards unhoused Veterans – to be evaluated for various homelessness interventions, including housing and case management. The VA fully centralized their intake and referral process in April 2011 via the Homeless Operations Management and Evaluation System (HOMES).

Every Veteran who contacts homelessness services is given an initial intake assessment, from which they are referred to programs meeting their needs. Veterans with lower needs or those who are merely at risk of homelessness may be offered interventions like case management or temporary rental assistance. Chronically unhoused and mentally ill Veterans are typically referred to longer-term housing interventions. The two largest programs serving this latter group of Veterans are HUD-VASH and the Grant & Per Diem program. Veterans are assigned to particular programs based on both social worker recommendation, Veteran choice, and program availability. Additionally, Veterans who are eligible for VA healthcare – which excludes those with a dishonorable discharge – can be referred to VA primary care, behavioral health, or other relevant healthcare services.

1.2.2 Housing First: HUD-VASH

HUD-VASH is a national program that distributes vouchers to provide permanent housing to formerly unhoused Veterans, with around 100,000 vouchers currently in use. The program began in 1992 and was massively expanded by the Obama administration, with a goal of ending Veteran homelessness. The program functions similarly to the Department of Housing and Urban Development (HUD)'s Section 8 program (Chetty et al. 2016). Those admitted to HUD-VASH receive a housing voucher and are able to choose their own apartment, subject to landlord approval. The voucher pays all or the majority of rent; Veterans are only responsible for contributing 30% of their income, which in many cases is de minimus.[1]

The budget and allocation of HUD-VASH vouchers is decided at the federal level each year. Vouchers are distributed across cities to local public housing agencies (PHAs), who partner with the VA. To qualify for the program, Veterans must be currently unhoused, with vouchers distributed to those most in need if they are in short supply. More specifically, the VA guidelines prioritize those who are chronically homeless – defined as those who have a mental or physical disability and have been unhoused for over a year or experienced multiple bouts of homelessness. Due to low voucher availability, wait times for the program are high. The median Veteran referred to HUD-VASH typically waits three months to receive a voucher. While enrolled in the program, Veterans must also agree to case management services, which can vary in frequency. Veterans can keep their vouchers indefinitely, and voucher receipt is not contingent on sobriety or receipt of any mental health services. In some cases, individuals may lose their voucher if they are evicted from the apartment due to "serious lease violations," such as excessive property damage or illegal activity. Nevertheless, VA guidelines recommend that PHAs only consider voucher termination as a last resort (National Housing Law Project 2008).

1.2.3 Treatment First: Grant & Per Diem

The VA's Grant & Per Diem Program (GPD) is the VA's largest transitional housing program, with over 12,000 beds.[2] The program awards grants to community partners, like the Salvation Army, to provide temporary communal housing and supportive services to Veterans, notably mental health and substance use treatment. Transitional housing facilities offer wrap around support services like case management, counseling, and addiction treatment, with a goal of making clients "ready" for permanent housing.

Given that third parties run transitional housing facilities, there is considerable variation across locations – with some housing requiring sobriety in order to remain in the program. While all programs have case workers, only some have licensed therapists on site; other sites may refer individuals to VA clinics for mental health treatment. Veterans who enter GPD can stay up to two years, after which the program helps find them a more permanent housing solution, whether independent or subsidized. GPD targets a similar group of clients as HUD-VASH. Veterans may choose to enroll in transitional housing rather than HUD-VASH due to client preference, social worker recommendation, or low housing voucher availability.

[1] Some Veterans with a service-connected disability, like PTSD, receive disability payments from the VA, which count towards the 30% income payment.

[2] While HUD-VASH vouchers can be permanent, GPD beds have high turnover, which is why bed numbers are much smaller despite high participation in my sample.

1.3 Data

Data on homelessness program participation comes from the VA's Homelessness Registry, which combines data from HOMES, the Homeless Management Information system (HMIS), and the Corporate Data Warehouse (CDW) to comprehensively document participation in the VA's homelessness programs, including entry and discharge dates. Relevant for my analysis, the registry also tracks initial intake assessments, as well as the primary affiliation of the social worker who completed one's assessment.

I link these data to full clinical records in CDW, the VA's detailed health record system, matching on social security number. CDW contains detailed patient demographic data, as well as documentation of all inpatient and outpatient encounters occurring at any VA hospital or clinic. Moreover, the VA has reliable mortality data, combining data from the Veterans Benefits Administration (VBA), Centers for Medicare and Medicaid Services (CMS), and Social Security Administration (SSA). Validation studies suggest the sensitivity of the available indicator of death is 98% (i.e., the fraction of true deaths that are accurately recorded is 98%). Matching on Social Security Numbers, I also link patient data from the VA with cause of death data, coded by ICD-10 code, from the National Death Index.

Finally, via CDW, I am able to access full social worker and clinical notes for every Veteran that meets with a VA-affiliated provider. To ascertain housing status over time, I process these clinical notes using the Natural Language Processing algorithm ReHouSED, which categorizes each note into "stably housed," "unstably housed," and "unknown" (Chapman et al. 2021). While use of Large Language Models is not authorized for the VA's clinical notes system, this rule-based NLP algorithm far outperforms methods like keyword mentions by understanding sentence context. In particular, the algorithm notes when mentions of housing are historical, hypothetical, or negated.[3] I provide more details in Section 1.5.7.

1.4 Empirical Approach

1.4.1 Sample construction

The VA fully centralized their homelessness intake and referral process in April 2011. Veterans who contact homelessness services complete an initial assessment, after which they are referred to various homelessness services, including permanent and transitional housing. Thus, the sample consists of all individuals who had an initial intake assessment with VA homelessness services between April 2011 and February 2017, allowing me to follow all individuals for three years prior to the start of the COVID-19 pandemic. For Veterans with multiple intake assessments during this period, I use the date of their first assessment. To track Veterans' health outcomes in CDW, I restrict the sample to Veterans also enrolled in VA healthcare. Finally, as both programs target unhoused Veterans with mental illness, I restrict the sample to those with a documented behavioral health and/or substance use disorder. More specifically, the Veteran must have had some healthcare encounter in the prior year in which the provider noted that the patient had a mental health condition, as documented via ICD-9 or ICD-10 codes.

[3] Compared to a human reading the notes, ReHouSED has a precision rate ($\frac{\text{True positives}}{\text{True positives+False positives}}$) of 65.3% and a recall rate ($\frac{\text{True positives}}{\text{True positives+False negatives}}$) of 65.8% (Chapman et al. 2021).

The final sample consists of 289,932 unhoused Veterans with mental illness. As shown in Table 1.1, the sample is 90% male, 36% Black, 7% Hispanic, and has an average age of 50. Comparing these demographic characteristics with the general US population of unhoused individuals (HUD 2024), my sample has a much higher percentage of men, a lower percentage of Hispanic individuals, and excludes homeless children (Table A.1).

Many Veterans in my sample struggle with drug or alcohol use, with around 54% suffering from a substance use issue. Moreover, around 62% of the sample suffers from a chronic physical condition, such as diabetes or heart disease. About 12% of the sample receives a HUD-VASH voucher within three months of intake, and about 18% of the sample enrolls in transitional housing. Characteristics of those enrolling in a housing intervention look similar to mean characteristics, with some differences. Notably, transitional housing recipients are more likely to suffer from substance use and more likely to have received mental health inpatient care, suggesting that transitional housing may be targeted towards those particularly interested in addiction treatment.

1.4.2 Instrumental variables design

I analyze the effects of HUD-VASH and GPD through the lens of a potential outcomes model with three mutually exclusive and exhausive treatment categories for individual i: $D_i \in \{v, t, 0\}$. Let $D_i(z_v, z_t)$ indicate potential treatments if the instruments take on values (z_v, z_t), D_i indicate realized treatment, $Y_{k,i}$ describe potential outcomes given treatment k, and Y_i describe realized outcomes. I define v as receiving a HUD-VASH voucher within three months of one's intake assessment, t as entering transitional housing within three months, and 0 as entering neither program. In the small share of cases where a Veteran enrolls in both HUD-VASH and transitional housing within three months, I define their treatment status based on the program they enroll in first. Veterans assigned one treatment status may eventually enter a different program. For instance, those who initially receive no housing intervention could enter a transitional housing program a year later. However, for simplicity, I define treatment here based on the initial program entered following one's intake assessment.[4]

Those assigned 0 may receive a non-housing intervention like case management services. In rare cases – less than 3% of observations in the sample – Veterans may also receive rapid rehousing or temporary rental assistance services via another VA program, Supportive Services for Veteran Families (SSVF). Given that SSVF is given to a small percentage of the sample and is targeted towards those with lower needs, I ignore this alternative treatment in the baseline analysis. However, in robustness checks, I drop individuals who immediately enroll in SSVF. Overall, one can think of v, t, and 0 as representing Housing First, Treatment First, and no housing intervention, respectively.

Initial treatment status is clearly non-random in this setting. Veterans may enroll in a particular program as the result of a combination of factors, including voucher availability, caseworker recommendations, and Veteran choice. Thus, to address the endogeneity of treatment assignment, I use an instrumental variables framework. Importantly, in contrast to settings with a single intervention, this setting requires a second instrument to separately identify treatment effects for Housing First and Treatment First margins.

I consider time of arrival to the VA homelessness clinic to be as good as random. Veterans

[4] I consider switching across programs in Section 1.5.6.

likely seek out homelessness services after a shock to their housing stability, the timing of which may be difficult to predict. This allows me to exploit two quasi-random sources of variation based on clinic arrival time. First, for a given area and year, permanent housing vouchers are fixed according to HUD allocations. These vouchers are coveted and in short supply, often resulting in long wait times to enroll in the program. While one's exact wait time may be correlated with need, voucher availability at the time of clinic arrival may vary for quasi-random reasons (for instance, due to unpredictable timing of other Veterans entering or exiting the program). Thus, as my first instrument, I exploit a proxy for the length of the voucher waiting list. In particular, let $Z_{v,i}$ be the (leave-out) percent of individuals who receive a housing voucher in three months or less among those who arrive to the same clinic within a 60-day window of one's intake date. When the number of available vouchers is small, idiosyncratic timing in program entry and exit can generate meaningful fluctuations in availability at the margin; this type of finite-sample variation is analogous to cohort-to-cohort differences in group composition in small samples (e.g., Hoxby 2000).

Secondly, when a Veteran arrives to the homelessness clinic, based on both staffing availability and patient needs, they may be assessed by a generalist social worker who triages them into different programs, or they may be seen by a specialist who oversees a specific program. In particular, GPD liaisons are in charge of managing referrals to transitional housing. If first seen by a generalist, individuals must be referred to the liaison in order to be admitted to a particular transition housing unit. Those immediately seen by a GPD liaison essentially bypass this step, resulting in accelerated entry into the program. Moreover, liaisons likely have wider knowledge of transitional housing availability and may encourage Veterans to enter a particular program. As patients are assigned to caseworkers based on need, direct caseworker assignment is non-random in this setting. Thus, instead of using assigned social worker, I exploit the availability of a transitional housing liaison at one's time of arrival as my second instrument. Variation in liaison *availability* reflects week-to-week differences in staffing schedules and demands placed on the liaison by Veterans already residing in GPD transitional housing, rather than on the characteristics of newly arriving Veterans. This variation is similar to that of designs that exploit exogenous changes in provider availability, where quasi-random timing of arrival interacts with staffing schedules to determine treatment exposure (e.g., Chan and Chen forthcoming). As a proxy for availability, let $Z_{t,i} = 1$ if a transitional housing liaison performs any intake assessment in the week one arrives to the clinic.

More explicitly, I define the two instruments used in my setting below:

$$Z_{v,i} = \frac{1}{|I_{j(i)}| - 1} \sum_{i' \in I_{j(i)}} \mathbf{1}(i' \neq i) D_{v,i'} \tag{1.1}$$

$$Z_{t,i} = \mathbf{1}(\text{Tr. liasion available}_{s(i)w(i)}) \tag{1.2}$$

where $s(i)$ and $w(i)$ refers to the individual's VA station and week of arrival, respectively, $I_{j(i)}$ is the set of patients who arrive within a 60-day time window of one's intake date, and $D_{v,i} = \mathbf{1}(D_i = v)$. Across all regressions, I control for VA station × year and VA station × month, controlling for any non-random variation in patient arrival, voucher availability, and staffing that varies across clinic-years or due to seasonality. Additionally, in some cases, individuals may be referred to homelessness services from an inpatient service related to mental health and/or homelessness, particularly the VA's long-term residential treatment programs for substance use. In these cases, patient arrival time may less random, as VA social workers may coordinate with inpatient staff to manage the patient's discharge plan. Thus, in the baseline analysis, I control for whether the patient had an inpatient stay in the prior year related to mental health or homelessness. In robustness checks, I also drop these patients.

Finally, in addition to standard IV assumptions, prior research has shown that 2SLS with multiple endogenous treatments rely on a no essential heterogeneity assumption (Kirkeboen et al. 2016, Heinesen et al. 2022). In other words, to uncover well-defined treatment effects, complier groups induced by $Z_{v,i}$ and $Z_{t,i}$ into different treatment statuses must have the same average treatment effects. For example, 2SLS estimands may not be interpretable as causal effects if compliers induced into Housing First have a larger treatment effect from voucher receipt compared to other complier types. In Section 1.5.4, I discuss this assumption in more detail, as well as argue that no essential heterogeneity is likely a reasonable model of treatment effects in my setting.

1.4.3 Understanding complier types

In order to assess the no essential heterogeneity condition, it is useful to define complier groups in this setting. To do so, I assume treatment assignment adheres to a simple choice model, which follows from program context. Individuals choose between a voucher, transitional housing, and neither program based on a combination of offers and preferences. Based on both program availability and eligibility, social workers may give individuals an offer to enroll in a specific program. After an individual is given offers, they can decide among programs according to their preferences.

Higher voucher availability and transition liaison availability increase the probability of receiving a voucher or transitional housing offer, respectively. Let $O_{v,i}(Z_{v,i}) \in \{0,1\}$ be an indicator for receiving a voucher offer, where $O_{v,i}$ is weakly increasing with $Z_{v,i}$. Similarly, let $O_{t,i}(Z_{t,i}) \in \{0,1\}$ be an indicator for receiving a transitional housing offer, where $O_{t,i}(1) \geq O_{t,i}(0)$. An individual has the following choice set:

$$\mathcal{F}_i(Z_{v,i}, Z_{t,i}) = \begin{cases} \{0\} & \text{if } (O_{v,i}(Z_{v,i}), O_{t,i}(Z_{t,i})) = (0,0) \\ \{0, v\} & \text{if } (O_{v,i}(Z_{v,i}), O_{t,i}(Z_{t,i})) = (1,0) \\ \{0, t\} & \text{if } (O_{v,i}(Z_{v,i}), O_{t,i}(Z_{t,i})) = (0,1) \\ \{0, v, t\} & \text{if } (O_{v,i}(Z_{v,i}), O_{t,i}(Z_{t,i})) = (1,1) \end{cases}$$

After receiving offers, an individual can decide among programs according to his or her preferences. I assume an individual derives utility $U_{0,i}$ from enrolling in no program, $U_{v,i}$ from the voucher program, and $U_{t,i}(Z_{t,i})$ from transitional housing, where $U_{t,i}(1) \geq U_{t,i}(0)$. Notably, I assume higher voucher availability only enters treatment choice by increasing the probability of receiving a voucher offer. Alternatively, transitional liaison availability may affect both offers and preferences for transitional housing, as the transitional liaison may encourage Veterans to enter the program. I assume an individual chooses a program based on their set of offers as follows:

$$D_i(Z_{v,i}, Z_{t,i}) = \arg\max{}_{k \in \mathcal{F}_i(Z_{v,i}, Z_{t,i})} U_{k,i}(Z_{t,i})$$

Following this choice framework, $Z_{v,i}$ induces individuals into the voucher program from a counterfactual of either transitional housing or no housing intervention. Similarly, $Z_{t,i}$ induces individuals into transitional housing from a counterfactual of either the voucher program or no housing intervention. Figure 1.1 shows a schematic of how $Z_{v,i}$ and $Z_{t,i}$ induce individuals into different treatment statuses. This specific model of treatment choice is not needed to uncover causal effects (Heinesen et al. 2022). However, it allows me to identify concrete complier types based on behavior and will later allow me to identify complier characteristics, which will be useful in assessing the no essential heterogeneity condition (Mountjoy 2022, Kirkeboen et al. 2016; see Section 1.5.4.).

Individuals may choose a treatment state based on both the value of $Z_{v,i}$ and $Z_{t,i}$ they receive. For example, someone who prefers vouchers over transitional housing may choose a voucher if they are offered both programs, but transitional housing if they are not offered a voucher. To gain intuition, consider binary versions of $(Z_{v,i}, Z_{t,i}) \in \{0,1\}^2$, where $Z_{v,i} = 1$ indicates high voucher availability. Let $\mathbf{D}_i(z_v, z_t) = (D_i(0,0), D_i(1,0), D_i(0,1), D_i(1,1))$ describe potential treatments when $(z_v, z_t) \in \{0,1\}^2$. In Table 1.2, I enumerate potential complier groups consistent with this choice framework and potential behavioral responses to $Z_{v,i}$ and $Z_{t,i}$. In principle there are seven such groups, depending on an individual's potential treatment status when facing each of the four possible values of the combination of $Z_{t,i}$ and discretized $Z_{v,i}$. These include four single-instrument compliers who only respond to one instrument (C_0^v, C_0^t, C_v^t and C_t^v) and three joint-instrument compliers whose decisions depend on the joint realization of both instruments ($C_0^{v+,t-}, C_0^{v-,t+}$, and C_t^{v-}).

Given these complier groups, propensity scores can be used to identify complier share mixtures. For instance, $-(\Pr[D_{t,i} = 1|\ Z_{v,i} = 1, Z_{t,i} = 0] - \Pr[D_{t,i} = 1|Z_{v,i} = 0, Z_{t,i} = 0])$ gives the share of individuals who are induced from t to v by $Z_{v,i}$, fixing $Z_{t,i}$ at 0. This group corresponds to the combined share of C_t^v and C_t^{v-} compliers, according to behavioral responses in Table 1.2. Let $P(c)$ be the population share of c compliers. In Table A.2, I show how propensity scores map to complier groups, and in Figure A.2, I provide more intuition.

Observed propensity score differences generally capture mixtures of complier types, since changes in $Z_{v,i}$ and $Z_{t,i}$ can simultaneously move different complier groups across multiple treatment boundaries. However, if changes in treatment shares are additive with respect to the two instruments, complier shares can be point identified. Specifically, define instrument additivity as:

$$\begin{aligned} &\Pr[D_{k,i} = 1 \mid Z_v = 1, Z_t = 1] - \Pr[D_{k,i} = 1 \mid Z_v = 1, Z_t = 0] \\ &= \Pr[D_{k,i} = 1 \mid Z_v = 0, Z_t = 1] - \Pr[D_{k,i} = 1 \mid Z_v = 0, Z_t = 0] \end{aligned}$$

This condition can be re-written as $\Pr[D_{k,i} = 1|Z_v = 1, Z_t = 1] - \Pr[D_{k,i} = 1|Z_v = 0, Z_t = 0] = (\Pr[D_{k,i} = 1|Z_v = 1, Z_t = 0] - \Pr[D_{k,i} = 1|Z_v = 0, Z_t = 0]) + (\Pr[D_{k,i} = 1|Z_v = 0, Z_t = 1] - \Pr[D_{k,i} = 1|Z_v = 0, Z_t = 0])$, or the change in the share in k from $(z_v, z_t) = (0,0)$ to (1,1) is equal to change from (0,0) to (1,0) plus the change from (0,0) to (0,1). In words, this condition says the change in voucher and transitional housing enrollment when $Z_{v,i}$ and $Z_{t,i}$ increase equals the sum of their separate effects. In Appendix Figure A3, I provide a visualization of this condition in simplex space (described in more detail in Appendix A.1).

Given my monotonic choice structure, where each instrument only increases the probability of enrolling in a particular treatment, this condition implies that the two instruments shift disjoint complier groups. Joint-instrument compliers therefore have zero probability, allowing the remaining complier groups to be point identified. Intuitively, if individuals only respond to a single instrument, the effect of $Z_{v,i}$ on treatment shares will be the same regardless of the value of $Z_{t,i}$, and vice versa, which is precisely the additivity condition. Conversely, if joint compliers existed, the effect of one instrument would depend on the value of the other. Formally, I show that the following holds:

Proposition 1 (Complier shares). *Under the treatment choice model described above, if* $\Pr[D_{k,i} = 1|Z_v = 1, Z_t = 1] - \Pr[D_{k,i} = 1|Z_v = 1, Z_t = 0] = \Pr[D_{k,i} = 1|Z_v = 0, Z_t = 1] - \Pr[D_{k,i} = 1|Z_v = 0, Z_t = 0]$ $\forall k$*, then:*

$$P(C_0^{v+,t-}) = P(C_0^{v-,t+}) = P(C_t^{v-}) = 0$$

The remaining complier shares can be point identified as follows:

$$P(C_0^v) = -(\Pr[D_0 = 1|Z_v = 1, Z_t = 0] - \Pr[D_0 = 1|Z_v = 0, Z_t = 0])$$

$$P(C_0^t) = -(\Pr[D_0 = 1|Z_v = 0, Z_t = 1] - \Pr[D_0 = 1|Z_v = 0, Z_t = 0])$$

$$P(C_t^v) = -(\Pr[D_t = 1|Z_v = 1, Z_t = 0] - \Pr[D_t = 1|Z_v = 0, Z_t = 0])$$

$$P(C_v^t) = -(\Pr[D_v = 1|Z_v = 0, Z_t = 1] - \Pr[D_v = 1|Z_v = 0, Z_t = 0])$$

Proof. See Appendix A.2.1. □

Under this additivity condition, the shares of C_0^v, C_0^t, C_t^v, and C_v^t are fully identified. Conceptually, this is because only a single group is induced from treatment status k by $Z_{j,i}$: C_0^v compliers are the only group induced by $Z_{v,i}$ from 0, C_0^t compliers are the only group induced by $Z_{t,i}$ from 0, C_t^v compliers are the only group induced by $Z_{v,i}$ from t, and C_v^t compliers are the only group induced by $Z_{t,i}$ from v.

This additivity condition corresponds empirically to the absence of an interaction between $Z_{v,i}$ and $Z_{t,i}$ in the first stage.[5] To test this empirically, I estimate the following regressions:

$$D_{k,i} = \alpha_{v,k} Z_{v,i} + \alpha_{t,k} Z_{t,i} + \alpha_{vt,k} Z_{v,i} Z_{t,i} + \zeta_{k,s(i)y(i)} + \phi_{k,s(i)m(i)} + X_i \beta_k + \epsilon_{k,i} \tag{1.3}$$

where $D_{k,i}$ is an indicator for enrolling in treatment k, $\zeta_{k,h(i)y(i)}$ are VA station × year fixed effects, $\phi_{k,h(i)m(i)}$ are VA station × month fixed effects, and X_i are the baseline controls, which includes whether the patient had a prior inpatient stay related to homelessness or mental health. Across all treatments $D_{k,i}$, I find that the interaction coefficient, $\alpha_{vt,k}$, is close to 0 and statistically insignificant.[6] According to the proposition above, this implies that joint-compliers have 0 share, and the shares of C_0^v, C_0^t, C_v^t and C_t^v are fully identified from first stage regressions. For complier shares estimates, see Appendix A.2.2. I will return to these complier types when I describe and provide evidence for the no essential heterogeneity assumption in Section 1.5.4.

[5] See Table A.2 for a full mapping of first stage coefficients to complier shares.

[6] These interaction terms are close to 0 and statistically insignificant whether $Z_{v,i}$ is used in its continuous version or converted into a binary indicator (for high versus low values of the leave-out share assigned a HUD-VASH voucher). Additionally, to assess whether there could be heterogeneity in the interaction between $Z_{v,i}$ and $Z_{t,i}$ across the support of $Z_{v,i}$, I allow the first stage to vary flexibly across deciles of $Z_{v,i}$. Of the 30 decile-by-$Z_{t,i}$ interactions across the first three stages, there is no consistent pattern and only two coefficients are statistically significant, consistent with being due to chance. I thank Evan Rose for this suggestion.

1.5 Results

1.5.1 First stage: how do $Z_{v,i}$ and $Z_{t,i,}$ impact treatment status?

First, I evalute the relevance of the instruments, or whether $Z_{v,i}$ and $Z_{t,i}$ actually shift people across treatments $0, v$, and t. Returning to Eq. 4, in Table 1.3, I report coefficients $\alpha_{v,k}$ and $\alpha_{t,k}$ when $D_{k,i} \in \{D_{v,i}, D_{t,i}, D_{0,i}\}$, or indicators for enrolling in the voucher program, transitional housing, and neither, respectively.[7] I find that $Z_{v,i}$ and $Z_{t,i}$ strongly predict enrollment in the voucher program and transitional housing, respectively (t-stat.= 30, 44). Specifically, a 10 percentage point increase in voucher availability increases the probability of enrolling in the voucher program by about 5.1 percentage points, and transitional housing liaison availability increases the probability of enrolling in transitional housing by about 8.9 percentage points.

Given that treatments are mutually exclusive and exhaustive, it follows that $Z_{v,i}$ and $Z_{t,i}$ decrease the probability of enrolling in other treatments. In particular, $Z_{v,i}$ decreases the probability that $D_i = \{t, 0\}$, and $Z_{t,i}$ decreases the probability that $D_i = \{v, 0\}$. Furthermore, $Z_{v,i}$ and $Z_{t,i}$ both cause much larger decreases in the probability of receiving no intervention than the probability of switching between interventions. In other words, while there is still significant switching between treatment statuses t and v, a much larger share of compliers would have received no intervention under the counterfactual than another housing intervention. For precise estimated complier shares, see Appendix A.2.2. Finally, in column 4, I show reduced form estimates for the effect of $Z_{v,i}$ and $Z_{t,i}$ on three-year mortality. I combine these first stage and reduced form estimates to obtain 2SLS coefficients in Section 1.5.3.

1.5.2 Balance checks: are $Z_{v,i}$ and $Z_{t,i,}$ correlated with individual characteristics?

In order for $Z_{v,i}$ and $Z_{t,i}$ to be valid instruments, they must be as good as randomly assigned, conditional on the baseline controls. Notably, certain VA stations and time periods may systematically have longer wait times. In particular, HUD adjusts voucher allocations to particular areas yearly based on homelessness counts and unmet Veteran need. Therefore, I control for any endogeneity in wait time using detailed VA station $\times$ year and VA station $\times$ month fixed effects. The remaining variation in residualized $Z_{v,i}$ represents random shocks to voucher availability not explained by yearly HUD allocations or seasonality effects.

Secondly, when an individual arrives to a VA medical or referral center for homelessness intake, they may be referred to a particular social worker based on their needs and/or preferences. For instance, those particularly interested in transitional housing may immediately meet with a transitional housing liaison upon arrival. In order to adjust for any non-random assignment to particular social worker types, in constructing $Z_{t,i}$, I instead rely on caseworker $availability$. Additionally, as with $Z_{v,i}$, the detailed VA station $\times$ time controls account for any potential correlation between patient demographics at particular stations and staffing.

To directly test whether $Z_{v,i}$ and $Z_{t,i}$ appear to be as good as random, conditional on the baseline controls, I run the following set of regressions:

[7] Given the findings reported in the previous section, I exclude the $Z_{v,i}Z_{t,i}$ interaction term from first stage and 2SLS regressions in the main text. As noted, all the interaction coefficients are close to 0 and insignificant.

$$C_i = \theta_v Z_{v,i} + \theta_t Z_{t,i} + \zeta_{0,s(i)y(i)} + \phi_{0,s(i)m(i)} + X_i\beta_0 + \epsilon_{0,i} \tag{1.4}$$

where C_i are relevant Veteran characteristics. In particular, if a given characteristic C_i is predictive of mortality but uncorrelated with $Z_{v,i}$ and $Z_{t,i}$, this suggests that both high and low risk Veterans are equally likely to be assigned high and low values of $Z_{v,i}$ and $Z_{t,i}$, indicating quasi-random assignment. In Figure 1.2, I show results from Eq. 5 using a wide-range of characteristics, including key demographics such as age, race, and gender; mental health conditions including depression, pychoses, substance use, and alcohol use; and physical health comorbidities. Moreover, in order to create an aggregate measure of patient risk, I construct a predicted mortality for each patient, $\hat{Y}_i$, based on an OLS regression of three year mortality on all patient characteristics (Chan et al. 2023, Costantini 2025).[8] I standardize all characteristics C_i so that coefficients can be interpreted in terms of standard deviations.

I find that coefficients θ_v and θ_t are small and statistically insignificant for 33 of 35 individual C_is for both θ_v and θ_t. For the other two characteristics, coefficients are marginally statistically significant, consistent with this being due to random chance. Perhaps most importantly, for $Z_{v,i}$ and $Z_{t,i}$, the point estimates when $\hat{Y}_i$ is on the left hand side are almost exactly 0, and results are far from statistical significance. To more clearly visualize the magnitude of the coefficients on $\hat{Y}_i$, in Figure A.4, I plot both the relationship between $\hat{Y}_i$ and the instruments (testing balance) compared to Y_i, three year mortality, and the instruments (the reduced form relationship). I find that the reduced form coefficient between $Z_{v,i}$ and Y_i is more than twenty times the magnitude of the balance coefficient. On the whole, these results provide compelling evidence that high risk patients are not systematically assigned to particular values of $Z_{v,i}$ or $Z_{t,i}$.

1.5.3 Mortality effects

To determine how participation in the voucher program and transitional housing impact three year mortality, I run the following 2SLS system:

$$D_{v,i} = \pi_v Z_{v,i} + \pi_t Z_{t,i} + \zeta_{1,s(i)y(i)} + \phi_{1,s(i)m(i)} + X_i\beta_1 + \epsilon_{1,i} \tag{1.5}$$

$$D_{t,i} = \gamma_v Z_{v,i} + \gamma_t Z_{t,i} + \zeta_{2,s(i)y(i)} + \phi_{2,s(i)m(i)} + X_i\beta_2 + \epsilon_{2,i} \tag{1.6}$$

$$Y_i = \beta_v D_{v,i} + \beta_t D_{t,i} + \zeta_{3,s(i)y(i)} + \phi_{3,s(i)m(i)} + X_i\beta_3 + \epsilon_{3,i} \tag{1.7}$$

where Eq. 6 and 7 are first stage regressions for the voucher program and transitional housing, respectively, and Eq. 8 is the structural equation for the outcome Y_i representing 3-year mortality, estimated by replacing $D_{v,i}$ and $D_{t,i}$ with their predicted values from the first stage models.

2SLS results imply that enrollment in the voucher program decreases mortality by 4.6 percentage points (s.e.=1.7), relative to a no-program counterfactual (Table 1.3, column 2). In contrast, the point estimate for the effect of transitional housing on mortality is close to 0. In fact, I can reject that the voucher program and transitional housing have the same effect on mortality at conventional significance levels (F = 4.06; p-value = 0.044). In Figure A.5, I show the mortality effect over time

[8] For the mortality prediction, I include age in 4-year buckets as separate fixed effects.

for both programs. The effect of Housing First rises steadily over time, becoming statistically significant around year two. On the other hand, effects for transitional housing consistently hover near 0. On the whole, the voucher program appears to have a large and statistically significant impact on long term mortality, while transitional housing likely has no or a very small, statistically undetectable impact.

Comparing these estimates to the OLS results (Table 1.3, column 1), 2SLS estimates for β_v and β_t have the same sign but are of larger magnitude. Recall that OLS coefficients are a combination of the average treatment effect plus a selection bias term reflecting non-random enrollment in the voucher and transitional housing programs. Those for whom $D_{v,i} = 1$ are a combination of voucher always takers – or those who always enroll in the voucher program no matter what values of $Z_{v,i}$ and $Z_{t,i}$ they receive – and C_0^v, C_t^v, and C_v^t complier groups. Similarly, $D_{t,i} = 1$ are a combination of transitional housing always takers and C_0^t, C_t^v, and C_v^t complier groups. Finally, the excluded category, $D_{0,i} = 1$, are a combination of never takers – those who never enroll in either program – and C_0^t and C_0^v complier groups. In this setting, voucher and transitional housing always takers are likely higher risk individuals: both programs target chronically unhoused, mental ill Veterans, and higher need individuals in particular are given priority for HUD-VASH. Therefore, OLS results are consistent with the selection bias term attenuating the effect of the voucher program, as always takers likely have higher baseline mortality risk than never takers.[9]

In column 3, I show the IV results when I add all individual X_i control variables (the same variables shown in Figure 1.2). Results are nearly identical. Next, in the baseline analysis, I consider three treatments: voucher, transitional housing, or neither. Conceptually, 'neither' represents individuals who received no housing intervention. However, around 3% of individuals in my sample receive another light touch housing intervention: Supportive Services for Veteran Families (SSVF), which includes short term rental assistance and rapid rehousing programs. Individuals participating in SSVF are unlikely to be compliers in my setting, as this program is targeting towards Veterans with lower needs – in particular, those who are housed but are having trouble paying rent or were recently evicted. However, to ensure that SSVF receipt does not change the results, I drop these individuals in column 4. Finally, around 24.1% of my sample had an inpatient stay related to homelessness or mental health at a VA hospital in the prior year. Social workers often coordinate with VA inpatient staff if Veterans are identified as unhoused, leading to referrals to homelessness intake. In these cases, patient arrival time may less random. In the baseline analysis, I control for whether the patient had a prior inpatient stay. To further ensure that this issue does not bias my results, in column 5, I drop these individuals entirely. Dropping those with prior inpatient stays make the effects of the voucher program look even larger.

1.5.4 Evaluating no essential heterogeneity

Prior research has shown that in order to recover well-defined causal effects, 2SLS with multiple endogenous treatments relies on restrictions on substitution patterns, a no essential heterogeneity assumption, or some combination of both (Kirkeboen et al. 2016, Kline and Walters 2016, Heinesen et al. 2022, Bhuller and Sigstad 2024, Humphries et al. 2025). According to the choice

[9] In principle, OLS estimates could also reflect selection bias coming from different complier groups enrolling in v, t, and 0. However, I argue in Section 1.5.4 that these groups look observably similar to each other in terms of mortality risk.

framework described in Section 1.4.3, I assume that $Z_{v,i}$ only affects treatment choice by increasing the probability that individuals enroll in v(from a counterfactual of either 0 or t), and $Z_{t,i}$ only affects treatment choice by increasing the probability that individuals enroll in t. Under these behavioral restrictions, Heinesen et. al (2022) show that to uncover interpretable and well-defined causal effects, there also must be no essential heterogeneity. In other words, complier groups induced by $Z_{v,i}$ and $Z_{t,i}$ into different treatment statuses also cannot have heterogeneous average treatment effects for a particular treatment. Formally,

$$E[Y_{k,i} - Y_{0,i} | C_i = c] = E[Y_{k,i} - Y_{0,i} | C_i = c'] \tag{1.8}$$

for all complier groups c and for treatment $k \in \{v, t\}$.[10] $Y_{k,i}$ represents potential mortality for individual i receiving treatment k. This model could be violated if those induced into v from 0 benefit more from vouchers than other complier types, for example. Eq. 8 rules out some, but not all forms of selection-on-gains. Among compliers, individuals cannot select into Housing First or Treatment First in way that is correlated with benefits. However, the condition does not impose restrictions on treatment effects for always takers and never takers, who may have different baseline mortality risk than compliers.

Intuitively, if different complier groups look very different from each other, this assumption is more likely to be violated. For example, in a comparison of no college, community college, and four-year college, Mountjoy (2022) finds that compliers on the margin between community college and no college have much lower prior test scores than those on the margin between community college and four year college, casting doubt on whether these different complier groups would experience similar gains from attending a four year college. However, in this setting, treatments are not "ordered;" instead, social workers assert that transitional housing and the voucher program both target the most at risk individuals.

To more formally evaluate the no essential heterogeneity assumption in my setting, I conduct three additional empirical tests. First, using additional instruments and overidentification tests, I evaluate whether one can reject homogenous treatment effects across key observable characteristics X_i. If additional instruments are added to the model, one can conduct a J-test to reveal whether different instruments yield different IV estimates, which may be suggestive of heterogeneous treatment effects. In particular, by interacting $Z_{v,i}$ and $Z_{t,i}$ with characteristics X_i, I can answer the following question: if compliers were only those that had $X_i = x$, would estimated mortality effects be the same? If the J-test fails to reject homogenous effects across key observable characteristics, this would reassure the reader that there is no substantial treatment effect heterogeneity in general, mitigating concern over essential heterogeneity.

In Table 1.4, I show overidentification test results when I interact $Z_{v,i}$ and $Z_{t,i}$ with a given characteristic X_i and add these interactions as additional instruments to the 2SLS model. To ensure instrument validity, in these regressions, I also control for the given characteristic X_i interacted with the baseline station × time fixed effects. I include tests for demographic and mental health characteristics, as well as high predicted mortality. Across all regressions, I fail to reject homogenous treatment effects across observable characteristics.

As an additional test, I examine complier characteristics directly. If different complier groups

[10]If there are additional restrictions on substitution patterns, no essential heterogeneity is not needed. In particular, if all compliers are induced from the same counterfactual treatment state, 2SLS has a causal interpretation even with unrestricted heterogeneity (Heinesen et al. 2022, Bhuller and Sigstad 2024).

have much higher baseline mortality risk $\hat{Y}_i$, on average, than other complier groups, this would cast doubt on the assumption that average treatment effects are equal across groups. In the former test, I failed to reject homogenous treatment effects for compliers that have high versus low predicted mortality (see Table 1.4). Nevertheless, to further reassure the reader that complier groups look similar to each other, I evaluate complier characteristics using a similar approach to that of Abadie (2003). For the case of binary instruments and no control variables, Mountjoy (2022) shows that the following Wald estimator yields mean $\hat{Y}_i$ for compliers who are induced by instrument $Z_{j,i}$ into treatment j from k, holding $Z_{j',i}$ fixed:

$$E[\hat{Y}_i|C_k^j] = \frac{E[\hat{Y}_i D_{k,i}|Z_{j,i}=1, Z_{j',i}] - E[\hat{Y}_i D_{k,i}|Z_{j,i}=0, Z_{j',i}]}{E[D_{k,i}|Z_{j,i}=1, Z_{j',i}] - E[D_{k,i}|Z_{j,i}=0, Z_{j',i}]} \tag{1.9}$$

This result follows from Proposition 1, where the denominator represents the negative of the complier share, $-P(C_k^j)$, and the numerator is $-E[\hat{Y}_i|C_k^j]P(C_k^j)$. Thus, dividing the two yields mean $\hat{Y}_i$ for C_k^j. Eq. 10 can be used to determine mean characteristics for C_0^v, C_0^t, C_t^v, and C_v^t compliers. In practice, I empirically identify $E[\hat{Y}_i|C_k^j]$ by running 2SLS regressions where $\hat{Y}_i D_{k,i}$ is on the left hand side and I instrument for $D_{k,i}$ with $Z_{j,i}$.

In Figure 1.3, I show complier mean predicted mortality across all groups. I cannot reject that all groups have the same mean $\hat{Y}_i$. In other words, all complier groups have similar baseline mortality risk, on average, making it much more likely that they have the same average treatment effects.

Another way of thinking about the no essential heterogeneity assumption is that treatment effect heterogeneity must be unrelated to behavior. In other words, no essential heterogeneity does not rule out all forms of heterogeneity; instead, it says that compliers cannot select into a certain treatment in a way that is correlated with potential benefits. The most obvious model one may have in mind here is that of Roy selection. If those who particularly benefit from the voucher program are voucher compliers, no essential heterogeneity does not hold.

To test for evidence of selection-on-gains, I examine the relationship between estimated local average treatment effects and the propensity to be treated (via the first stage). More specifically, starting with the voucher program, I divide the sample into the VA's Veteran Integrated Service Networks (or VISNs), which are based on geographic area. I run first stage stage regressions separately by VISN for the effect of $Z_{v,i}$ on $D_{v,i}$. Then, I estimate the effect of the voucher program on mortality for each VISN (via 2SLS regressions). Finally, I plot the relationship between the first stage and 2SLS estimates. I repeat this process for transitional housing: by VISN, I estimate both the effect of $Z_{t,i}$ on $D_{t,i}$ and the LATE for the effect of transitional housing on mortality. I show results in Appendix Figure A6. Although 2SLS estimates are noisy due to small sample sizes, for both the voucher program and transitional housing, I find that VISN-specific LATEs are uncorrelated with the first stage. In other words, in areas where treatment effects are larger, individuals are not more likely to select into the program – ruling out Roy selection. On the whole, these results suggest that even if there is unobservable treatment effect heterogeneity, this heterogeneity is not correlated with the propensity to be treated.

Finally, what if there are small violations of no essential heterogeneity that are difficult to detect? To evaluate the degree of potential bias, I use a bounding approach. In particular, Heinesen et al. (2022) decompose the multi-instrument 2SLS coefficient into a LATE term and selection terms that involve the difference in average treatment effects across complier groups. The weight of these selection terms depend on the share of compliers who switch between treatments – in my setting, C_t^v and C_v^t compliers. Bhuller and Sigstad (2024) similarly argue that if these complier shares are 0, 2SLS recovers a well-defined LATE. Following this logic, if both the share of C_t^v and C_v^t compliers are small and there are only small violations to the no essential heterogeneity condition, 2SLS estimates will be close to true LATEs. In my setting, C_t^v and C_v^t compliers represent only 16.1% and 15.7% of all $Z_{v,i}$ and $Z_{t,i}$ compliers, respectively (see Table A.3). More rigorously, I can precisely estimate bounds on the true LATE under the assumption that average treatment effects across complier groups differ by a factor of ϵ,or

$$|E[Y_{k,i} - Y_{0,i}|C_i = c] - E[Y_{k,i} - Y_{0,i}|C_i = c']| < \epsilon \tag{1.10}$$

In Appendix A.3, I derive analytical worst-case scenario bounds on $LATE_v = E[Y_{v,i} - Y_{0,i}|C_0^v]$, the LATE of Housing First (compared to no intervention) for C_0^v compliers. The same logic can be applied to derive bounds for $LATE_t = E[Y_{t,i} - Y_{0,i}|C_0^t]$, the LATE of Treatment First for C_0^t compliers. In Table 1.6, I show bounds on both LATEs under different values of ϵ. I find that even with an ϵ as large as one percentage point, or around 20% of the size of the treatment effect, the bounds remain tight: $LATE_v \in (-0.049, -0.043)$ and $LATE_t \in (-0.010, -0.004)$. In fact, $LATE_v$ and $LATE_t$ have non-overlapping bounds as long as $\epsilon < 0.06$ p.p. These results suggest that β_v and β_t are only far from true LATEs if there are large violations to Eq. 9, which is likely implausible given the tests presented in this section.

1.5.5 Evaluating exclusion

Finally, in order for my instruments to be valid, they must influence mortality only by influencing treatment status. In other words, voucher wait time and transitional liaison availability can only have an impact on mortality by shifting individuals between treatments 0, t, and v. Most relevant for this analysis, $Z_{v,i}$ and $Z_{t,i}$ cannot affect mortality by changing how individuals interact with the homelessness system as a whole – for instance, by getting into care faster or increasing engagement in alternative programs, such as case management services.

In Table 1.7, I test various additional concerns related to exclusion. First, one concern is that homelessness wait times are correlated; in other words, if it takes longer to get a housing voucher, this could be because homelessness services are backed up in general. Therefore, in column 2, I add a control for the mean (leave-out) wait time (in days) to one's follow up appointment with a social worker (regardless of which homelessness program the patient enrolled in). IV results are very similar, suggesting that the results are not driven by more efficient homelessness services in general. Second, $Z_{t,i}$ relies on the availability of a transitional housing liaison, a specialty social worker in VA homelessness clinics. This brings up a related concern: if a homelessness referral center has a specialty staff member available, perhaps they are better staffed in general, which may improve the referral process on the whole. To mitigate this concern, in column 3, I control for the number of unique staff available at patient intake. Lastly, a final concern is that transitional housing liaisons may do more than refer patients to GPD. Importantly, the VA also has emergency housing and case management services. If transitional housing liaisons are more (or less) likely than other social workers to refer individuals to these programs, that could bias mortality results. Thus, in columns 4 and 5, I control for entering emergency housing and case management within three months of intake.[11] Finally, as discussed in Section 1.5.3, in Table 1.3, I also drop individuals who enroll in an alternative housing program, SSVF. Overall, the results are highly robust across specifications.

1.5.6 Switching between treatment statuses

Until now, I have considered a static model of treatment status: individuals are in v, t, or 0 based on the first program they enroll in within the first three months following intake. In practice, individuals can switch between treatment statuses: someone initially enrolled in neither program may receive a housing voucher at month six, or an individual in transitional housing might enroll in HUD-VASH once they exit their transitional housing unit.[12]

Descriptively, only 7.8% of those in 0 – who initially enroll in neither program – receive a housing voucher within a year following intake. In contrast, about 20.8% of those enrolled in transitional housing receive a housing voucher within a year. There are several possible explanations for why rates are much higher for those in tr. First, compared to the broader sample population, those who enroll transitional housing are likely more similar to HUD-VASH recipients, making them more likely to qualify for and enroll in the program. Second, individuals originally interested in HUD-VASH may enroll in transitional housing while they wait for their housing voucher to come through. Finally, this could partially be a treatment effect: part of the goal of transitional housing is to help you secure more permanent housing following program exit, which could include transitional housing social workers helping you sign up for HUD-VASH.

To ensure that my static model of treatment status does not impact the results, I run a series of robustness checks related to switching between treatment statuses in Table A.4. First, I redefine $D_{v,i}$ as an indicator for *ever* receiving a housing voucher within a year of intake, even for those who initially enroll in transitional housing. This helps address two concerns: 1) that the voucher availability instrument may effect availability beyond three months, and 2) that some individuals may only enroll in transitional housing for a short period while they await their voucher. In Table A.5 column 1, I show the baseline first stage (with $D_{v,i}$ as the dependent variable) for reference. In column 2, I show the first stage when $D_{v,i}$ is re-defined as receiving a housing voucher within a year. Interestingly, the coefficient on $Z_{t,i}$ becomes insignificant in this model. In other words, for those who would have received Housing First under the counterfactual – or C_v^t compliers – transitional housing merely delays their entry into HUD-VASH.

In Table A.4 column 3, I show the baseline 2SLS mortality results for reference. In column 4, I show 2SLS results when I replace $D_{v,i}$ with receiving a voucher within a year, which are nearly identical. If anything, in this specification, the effect of transitional housing becomes a more precise 0. Finally, in column 5, I simply drop individuals whose initial treatment status is 0 or tr but later enroll in the voucher program. Coefficients are also highly aligned with the baseline results.

[11] A related concern is that case workers may refer individuals to an inpatient service related to homelessness and/or mental health, notably the VA's mental health residential treatment programs and domiciliary care. About 6.5% of the sample enters one of these programs within three months of intake. I also control for referral to an inpatient service, and coefficients are nearly identical to the baseline specification.

[12] Since individuals can keep housing vouchers permanently, this is less of a concern for people who initially enroll in v.

1.5.7 Mechanisms

Why does Housing First have such a large impact on long term mortality, but Treatment First has little effect? In this section, I evaluate potential explanations. First, previous research has found that offering unhoused individuals permanent, subsidized housing substantially reduces the number of days spent homeless compared to the status quo (Rosenheck et al. 2003, Montgomery et al. 2013). Thus, one potential explanation is that housing alone decreases mortality. Unsurprisingly, Meyer et al. (2023) find that the unhoused population experiences 3.5 times the mortality risk of those who are housed, controlling for demographic characteristics. Additionally, if individuals stay enrolled in Housing First, they may become more attached to the homelessness support system, especially compared to those enrolled in programs that are temporary in nature. Without monitoring by social workers and other staff, Veterans who disappear from the system may be particularly likely to experience poor outcomes. Finally, an alternative explanation, often espoused by Housing First advocates, is that housing stability is a necessary pre-condition for individuals to seek the care they need – particularly in addressing mental health and substance use. To shed light on this point, I evaluate Veterans' engagement with mental and physical healthcare over time.

Housing status over time

First, I evaluate Veterans' housing status over the three year period following program enrollment. Housing status is a particularly hard outcome to study in a non-RCT setting, as it is typically not recorded systematically in clinical records. Additionally, there are no records of housing status for those who disengage entirely from VA homelessness services, making full ascertainment difficult.

To partially address these issues, using the social worker and clinical notes available within CDW, I analyze Veteran notes using the Natural Language Processing (NLP) algorithm ReHouSED. Described and validated in detail by Chapman et. al (2021), this rule-based NLP algorithm classifies individual clinical notes into "stably housed," "unstably housed," and "unknown" categories. The algorithm outperforms methods like keyword matching by understanding sentence context – in particular, it identifies cases where housing stability is hypothetical (e.g., "he would like to own his own apartment"), negated (e.g., "he is not homeless"), or historical (e.g., the patient's past medical history lists a homelessness ICD code, which is no longer current). I customize the algorithm for my context so that transitional housing is counted as stable housing, while shelter stays, doubling up, and street homelessness count as unstable housing. Appendix Figure A7 provides examples of how RehouSED interprets and classifies notes into stably housed and unstably housed. If the algorithm sees no clear mention of stable or unstable housing, it classifies the note as unknown.

To employ RehouSED, I collect all clinical or social worker notes for Veterans in my sample, identifying over 50 million clinical notes that include a housing keyword.[13] After running RehouSED on these notes, I generate a month-level stable housing variable, which equals 1 if the Veteran was ever identified as stably housed in that month. Then, I create a quarter level mean, $H_{q,i}$, or the percent of months during that quarter in which the Veteran was identified as stably housed.[14] If the Veteran had no mentions of housing status during that period, I classify them as

[13]Reducing the notes to those that include a housing keyword was necessary to reduce the computational intensity of the algorithm.

[14]I exclude notes classified as "unknown" from this quarter-level mean if the individual had at least one documented housing status during the quarter. In other words, if the patient was identified as stably housed in months 4 and 6 and

having unknown housing status during that quarter with an indicator $U_{q,i}$.

Maintaining the category of unknown housing status eliminates concern over attrition bias, as those participating in no housing intervention are much less likely to have their housing status recorded. Importantly, however, the interpretation of regression results is subtle in the setting: stable housing essentially captures the joint outcome between having stable housing and having housing status recorded – e.g., by meeting with a social worker or in some cases, a primary care doctor.

After classifying housing status for all Veterans during the three year follow up period, I run 2SLS regressions where the outcome variables are either $H_{q,i}$ or $U_{q,i}$. I show the results in Figure 4. In quarter 1 (days 0-90), transitional housing recipients are more likely to be housed than HUD-VASH recipients. These results reflect the fact that Treatment First compliers are able to enter transitional housing units almost immediately. In contrast, Housing First recipients have to find an apartment that accepts their voucher, a much slower process. Nevertheless, this trend reverses by quarter 2. Housing First recipients maintain stable housing at a much higher rate throughout the study period, with an effect size reaching its peak around 34 percentage points in quarter 3 (days 180 - 270). In contrast, transitional housing has an attenuating effect on housing stability over time, with an effect size of only 9 p.p. by quarter 12 (days 990-1080). These results are also in line with institutional details: while Housing First recipients can keep their vouchers forever, GPD participants must find their own housing after a 6 month to 2 year stay in a transitional housing unit.

Similarly, both programs make one less likely to have unknown housing status, but transitional housing also has an attenuating effect over time. These results imply that social workers stay in contact with and can much better track the whereabouts of those actively enrolled in HUD-VASH or GPD. I further explore engagement with homelessness services in the next section. For more details on housing status over time, including complier means, see Appendix A.4.

Engagement with homelessness services

Next, as implied in the prior section, enrollment in Housing First and transitional housing both increase the probability that one engages with VA social workers. In particular, HUD-VASH participants are required to participate in case management. Therefore, one channel for the mortality effect is that VA caseworkers can keep track of where Veterans are and potentially intervene in a crisis. To systematically track engagement with VA homelessness services over time, for each year following intake, I create an indicator, $SW_{i,y}$, for having a visit with a VA social worker during the year.[15] I then run 2SLS regressions with $SW_{i,y}$ for each year as the outcome variable. I show results in Figure 1.5. Overall, Housing First participants are 40 percentage points more likely to stay in contact with VA social workers, a result that is stable across the study period. Transitional housing participants are also 30 percentage points more likely to stay engaged with homelessness services in year one, with effects fading over time as Veterans exit GPD housing. However, the effect is still persistent: by year three, transitional housing recipients are still nearly 20 percentage points more likely to have a visit with a social worker compared to those in the no intervention group. Thus,

had unknown housing status in month 5, I count them as having 100% stable housing during that quarter.

[15] I exclude any visits in days 0-90, as Veterans must mechanically meet with a social worker in order to enroll in HUD-VASH or GPD.

given than transitional housing has no impact on mortality, it is unlikely that engagement with VA social workers is a key mechanism in explaining the mortality effect.[16]

Engagement with mental and physical health services

Lastly, I evaluate the hypothesis that housing stability is a necessary pre-requisite to address other issues, particularly substance use and mental health. This line of reasoning gets at the root of the Housing First versus Treatment First debate: with Housing First, mental health and substance use treatment is optional. However, proponents hypothesize that individuals are better able to participate and comply with mental health treatment once they have a stable roof over their heads. Alternatively, in Treatment First models, mental health treatment, and sometimes sobriety, are typically required to stay enrolled in the program.

As all individuals in my sample were assigned a prior mental health diagnosis code, the majority previously attended at least one mental health visit. Thus, I construct two *intensive* margin measures of engagement with mental health care over time. First, I construct a simple variable of the count of mental health visits attended in year y. Secondly, as no shows are a big issue in mental health care, I construct a no show rate for each year y, or the percentage of visits one did not attend over the total number scheduled. As everyone in my sample has mental health or substance use condition, I let the no show rate equal 100% if one did not schedule or attend any visits during year y.

I show 2SLS regression results on use of mental health care in Figure 6. I find that transitional housing both increases one's use of mental health care in the first year and decreases the no show rate, consistent with transitional housing programs often requiring mental health treatment.[17] However, these effects are impermanent, fading to 0 by year 2 as Veterans exit GPD housing. In contrast, Housing First appears to have no impact on engagement with mental health care.[18] This evidence suggests that the mortality effect is not driven by mental health treatment. However, I cannot rule out that housing alone improves one's mental health, even though Veterans are not seeking more care once they are housed.[19]

Finally, in Figure A.9, I also show results for engagement with VA primary care over time. I define engagement with primary care as attending any visit in year y, following general guidelines for checkup frequency. I find that neither program has an impact on overall engagement with primary care.

[16] One hypothesis is that social workers could help Veterans by increasing their disability compensation (Silver and Zhang 2026). I find no evidence that enrolling in HUD-VASH or GPD increases the amount one receives in disability benefits.

[17] Effects of transitional housing on use of mental health care are likely undermeasured here, as I can only observe visits at the VA. While in transitional housing units, some Veterans may participate in in-house therapy delivered via third party organizations.

[18] I find no effect of Housing First on mental health care utilization even for those with no prior specialty mental health care engagement.

[19] My results on mental healthcare engagement contrast with the results of Hanson and Gillespie (2024), who find that enrolling individuals in Housing First increases psychiatric office visits. Their setting differs from mine in that everyone in my sample has a diagnosed mental health condition and has a prior attachment to VA healthcare services. In contrast, in Hanson and Gillespie (2024), Housing First may facilitate access to outpatient mental health care through the program's embedded care coordination.

1.5.8 Cause of death

Does housing alone explain the mortality effect, or do Veterans who recieve a housing voucher also benefit from improved mental health and reduced substance use? To further illuminate mechanisms, I link my sample to National Death Index data on cause of death, dividing deaths into those caused primarily by housing alone, those related to mental health or drug use, and those related to other causes. I show 2SLS results on cause of death in Table 1.8.

First, I investigate potential causes of death primarily explained by housing alone. Individuals who lack housing – particularly those who live on the streets – are more exposed to the elements, making them particularly vulnerable to extreme heat or cold. In Table 1.8 column 1, I find that voucher recipients much less likely to die from exposure to the elements. Secondly, without a safe place to call home, unhoused individuals are much more likely to subject to violence or assault. In column 2, I also find evidence that voucher recipients are much less likely to die from assault. These two outcomes demonstrate clearly that shelter alone lowers mortality risk.

Next, I investigate causes of death primarily related to mental health or substance use. Most intuitively, if mental health is improving and substance use is declining among Housing First recipients, one may expect suicide or overdose rates to decline. In column 3, I show some suggestive, though underpowered evidence that voucher recipients have lower rates of suicide and overdose. Lastly, I investigate a cause of death closely linked to both substance use and homelessness: HIV/AIDS, Hepatitis B, and Hepatitis C infections. Transmitted via blood, these infections are often spread among people who inject drugs via sharing needles or other injection drug use equipment. According to NASEM (2018), unhoused injection drug users are particularly likely to contract these viruses, as they often lack places to cleanly and safely inject drugs. In column 4, I also find that voucher recipients are at much lower risk of dying from HIV, Hepatitis B (HBV), or Hepatitis C (HCV). Decomposing this coefficient further, I find that HCV deaths explain around 76% of the effect.

The interpretation of this result is more subtle. While individuals often contract these viruses via unsafe injection practices, one can often survive many years with these conditions even if left untreated. However, HIV weakens the immune system, making one much more vulnerable to new infections. Similarly, viral hepatitis weakens the liver over time, leading to increased mortality risk from new infections or alcohol binges. Thus, cause of death here is multifaceted: first, it could suggest decreased drug or alcohol use, as well as safer injection practices. Prior descriptive research on people who inject drugs find a strong correlation between homelessness and risky injection behaviors, such as syringe and equipment sharing (Hotton et al. 2021). This mechanism is closely connected to "harm reduction" philosophy of addiction treatment, which espouses that even if drug use continues, safe injection practices – such as providing individuals sterile needles via needle exchanges – can help reduce drug use's harmful effects. Second, among those with pre-existing infections, it could suggest decreased risk of contracting non-drug related infections. For example, those who lack housing often have poorer hygiene, leading to skin or soft tissue infections. Finally, mortality reductions could be caused by improved adherence to antiviral medications. I find some weak evidence of this phenomenon: Housing First makes one 2.5 p.p. (s.e.=1.5) more likely to fill an antiviral medication for HIV, HCV, or HBV during the study period. Overall, causes of death ostensibly related to homelessness, mental health, or both explain about half of the mortality effect.

1.5.9 Comparison to other studies

Two US-based studies have linked prior RCT evidence on Housing First with administrative data on mortality (Raven et al. 2020, Hanson and Gillespie 2024). These studies both find imprecise results for the effect of Housing First on mortality, with confidence intervals that include both meaningful reductions and increases.

Both Raven et al. (2020) and Hanson and Gillespie (2024) look at mortality effects at shorter time horizons and report intention-to-treat effects. To compare my results more directly with these studies, in Table A.5, I report two-year mortality treatment-on-the treated estimates for my study compared to these benchmarks, scaling Raven et al. and Hanson and Gillespie by their reported first stages: 50% and 82%, respectively. This latter statistic is likely an overestimate of the true first stage, as Hanson and Gillespie do not report what percent of the control group received some alternative housing assistance, which was nearly 36% in Raven et al. (2020). Nevertheless, I marginally fail to reject any difference between my estimate and that of these alternative studies when I use the same time horizon (p-values=0.16, 0.051). Secondly, as Hanson and Gillespie targeted individuals with a high number of prior jail stays, I construct an alternative sample defined as individuals who completed a homelessness intake assessment where they indicidated they had spent over a month in jail in their lifetime (not restricting to those with a prior mental health condition as in the primary sample). Here, I find that estimates are even more aligned with Hanson and Gillespie (p-value=0.31).[20] Taken together, these comparisons suggest that differences between my results and prior US studies are likely explained by differences in estimands, follow-up horizons, and sample composition.

1.5.10 Policy implications and cost effectiveness

Are any types of transitional housing effective?

Transitional housing at the VA is administered by third party agencies, who differ in both services provided and program requirements (such as requiring weekly therapy or even sobriety to stay enrolled). In fact, the VA explicitly defines different transitional housing types, which may be more or less service-intensive. This raises the question: are any types of transitional housing effective, and if so, what lessons can be learned from them? In particular, critics argue that transitional housing with strict requirements may merely increase barriers to obtaining a permanent home. In the extreme case, an individual may be kicked out their transitional housing unit and lose both immediate housing and social worker support.

Thus, one potential hypothesis is that transitional housing may be more effective if it: 1) does not have strict program requirements, and 2) primarily emphasizes placing individuals in long term housing. Two types of VA transitional housing fall more closely under this model: "low demand" and "bridge housing." In the "low demand" model, mental health treatment is offered, but explicitly not required, and services are focused on finding permanent housing. Similarly, the "bridge" model emphasizes very short term stays in transitional housing units, with the focus being on securing permanent housing. While I cannot observe the program type that individuals enroll in, I can

[20] This provides some evidence of treatment effect heterogeneity in this alternative sample at a 2-year time horizon. Nevertheless, when I interact prior jail status with $Z_{v,i}$ and $Z_{t,i}$ and conduct a J-test using my primary sample, I cannot reject homogenous treatment effects at a three-year time horizon.

observe the percent of transitional housing beds in each VISN that are classified as low demand and/or bridge. To examine program heterogeneity, I divide VISNs into those with above and below median shares of low demand/bridge beds. Then, I rerun the mortality IV regressions separately for the above and below median samples. I show results in Table A.6. Though treatment effects are noisy, I find suggestive evidence that low demand and bridge housing may be more effective than other transitional housing types.

Is Housing First worth funding?

Even if Housing First programs improve health, are they worth paying for? One way to think about cost-effectiveness in this setting is to consider housing to be a health intervention. In public health research, medical interventions in the US – such as a new cancer treatment – are generally considered cost-effective if they cost less than \$100,000 - \$150,000 per year of life saved. In deciding what programs are worth funding, researchers construct incremental cost-effectiveness ratios (or ICERs), comparing a given intervention against this benchmark. To assess cost-effectiveness, I consider the ICER of HUD-VASH (compared to no intervention) over the three year study period as:

$$\text{ICER} = \frac{\Delta\text{cost}}{\Delta\text{life years}} \tag{1.11}$$

On the cost side, I first document if/when an individual in my sample received housing through HUD-VASH. I conservatively assume that once an individual obtains housing, they continue to use their HUD-VASH voucher throughout the remainder of the study period. Then, I run 2SLS regressions where the outcome variable is the number of months the individual used a housing voucher. I find that HUD-VASH enrollment causes one to use a housing voucher for 23.0 (s.e.=0.7) additional months over the study period. Excluding the cost of case management, HUD pays an estimated \$581 per month for every HUD-VASH voucher (National Homeless Information Project 2018). Thus, estimated housing costs per treated individual are approximately \$13,380.

Next, I estimate the additional costs of social worker visits compared to those who receive no housing intervention. I approximate the cost of social worker visits using the VA's Health Economics Resource Center (HERC), which estimates costs to the VA using Medicare reimbursement rates. Using the total cost of social worker visits as the left hand side variable in regressions, I estimate an additional cost for those who enroll in HUD-VASH of \$9,667 (s.e.=885).

I also assess changes in healthcare and homelessness services utilization that could lead to cost-savings among HUD-VASH recipients. First, I find that enrolling in the voucher program leads to an average of 7.9 (s.e.=2.98) less days of VA emergency shelter use over the study period. I estimate the average daily cost of a shelter stay to be \$31 (HUD 2010),[21] leading to a cost savings of \$246. Finally, and most significantly, the VA operates various inpatient treatment programs related to homelessness and mental health, including residential rehabilitative treatment programs and domiciliary care. I find that enrolling in Housing First substantially decreases the length of stay in one of these programs by 12.57 days (s.e.=4.77) over the study period.[22] According to HERC

[21] This estimate averages the daily cost across various shelter types.

[22] In contrast, I find that transitional housing has no effect on inpatient length of stay (coeff=0.91; s.e.=4.00).

estimates, this decline in length of stay translates to a cost-saving of $41,734 (s.e.=$16,429).[23,24]

This result is consistent with descriptive studies from the medical literature, which find both very large healthcare spending among homeless individuals, as well as substantial cost differences once these individuals are housed. For chronically homeless individuals with alcohol use – a similar population to that of this paper – a study from Seattle finds that average total healthcare spending prior to housing was $15,072 per year. This number declined to $8,592 in the year following housing, translating to $19,440 in healthcare cost savings over three years (Larimer et al. 2009). A study from Massachussets finds that healthcare costs were around $32,331 per year for unsheltered homeless individuals, compared to only $9,648 for those with shelter, with the primary difference coming from inpatient costs (Koh et al. 2022). This translates to $68,049 in healthcare cost-savings over three years. Relatedly, other descriptive studies have noted that homeless individuals have longer hospital stays than non-homeless individuals. A study from New York City finds that homeless patients stay around 4.1 extra days in the hospital per discharge, with the majority of these stays being related to substance use or mental illness (Salit et al. 1998). Physicians reported delaying patients' discharge due to lack of housing and worry about compliance with follow up care. Considering these massive cost savings from shorter hospitalizations, one could think of Housing First as a diversion program – keeping individuals out of costly psychiatric care.

On the whole, due to a substantial reduction in inpatient costs, I find that Housing First is actually *cost-saving* over the three-year study period: compared to those who enrolled in no housing intervention, HUD-VASH saved $18,931 per participant. Thus, even if the intervention incurred no health benefit, it would be worth funding purely from a cost perspective. I summarize these findings in Table 1.9. This cost analysis should be interpreted as the costs incurred by either the VA or HUD, which is what I can observe. However, I likely underestimate broader cost-savings of HUD-VASH due to reduced use of other public services. For example, Housing First may reduce incarceration rates (Cohen 2024). In addition, I cannot observe non-VA emergency shelter use or Medicaid or Medicare spending.

Given that HUD-VASH incurs lower costs and has higher benefits than no housing intervention, from the ICER perspective, Housing First is a dominant intervention. Thinking about costs and benefits from an ICER perspective is also closely connected to the public finance literature on the Marginal Value of Public Funds (see Hendren and Sprung-Keyser 2020) in cases where fiscal externalities are close to 0. Although I am unable to observe employment for Housing First participants, Brounstein and Wieselthier (2024) estimate minimal labor supply effects from enrollment in permanent housing, likely due to large rates of unemployment among this population. Thus, if participants have any positive willingness to pay for improved quality of life and increased longevity due to permanent housing, the marginal value of public funds in this setting is infinite.

[23] Importantly, I replicate the RCT result from Rosenheck et al. (2003), which finds no effect of HUD-VASH on the *number* of inpatient stays. However, cost estimates change substantially once length of stay and visit-specific HERC cost estimates are calculated.

[24] I do not find any other significant effects on healthcare utilization, including outpatient mental health, primary care, ED visits, and hospitalizations related to physical health.

1.6 Discussion

This study is not only the first to show that Housing First unequivocably reduces mortality, but it is the first to explicitly show that Housing First far outperforms Treatment First type interventions in terms of both long term housing stability and health outcomes. My results add to a growing body of evidence that the primary solution to homelessness is housing, and that psychosocial issues do not need to be addressed prior to addressing issues of housing instability (e.g., Rosenheck et al. 2003, Gubits et al. 2018, Montgomery et al. 2013). Moreover, echoing descriptive results from the medical literature, I find that Housing First is cost-saving due to large reductions in psychiatric hospitalization length of stay.

However, my results cast doubt on the narrative that Housing First models provide the stability needed for individuals with mental health issues to begin to address them. Despite high rates of mental health treatment non-attendance in this population, Housing First does not lead to an increase in mental health treatment nor improve the rate of appointment attendance. In contrast, Treatment First models do improve adherence to mental health treatment in the short term, but effects are short-lasting and do not seem to translate into large health gains. Thus, for the large population of unhoused, mentally ill individuals, other interventions may be necessary to improve long term engagement with mental health care. Despite having little impact on healthcare utilization, Housing First has a large effect on overall health and wellbeing, as indicated by lower mortality rates. In particular, permanent housing reduces deaths from extreme weather, violence, and infections caused by unsafe drug injection. This latter result provides some suggestive evidence that "harm reduction" in addiction treatment – or providing resources to allow individuals to safely inject drugs – could reduce mortality.

Comparing outcomes from Housing First versus Treatment First interventions is of critical policy relevance. In an executive order from July 24, 2025, entitled "Ending Crime and Disorder on America's Streets," the Trump administration officially ended federal support for Housing First policies. Instead, the administration hopes to shift funding towards programs focused on mental health and substance use treatment. Emphasizing high rates of behavioral health disorders and addiction among the homeless, the Trump adminstration stated that Housing First policies "deprioritize accountability and fail to promote treatment, recovery, and self-sufficiency." Despite the adminstration's promise to curb street homelessness, this study suggests that scaling back Housing First initiatives will likely have the opposite of the desired effect.

Table 1.1: Veteran characteristics

	Full sample	Voucher recipients	Transitional housing enrollees
Male	0.90	0.84	0.92
Black	0.36	0.37	0.39
Hispanic	0.07	0.07	0.06
Age	50.0	49.0	50.7
Depression	0.66	0.67	0.63
Psychoses	0.13	0.11	0.13
Alcohol use	0.41	0.37	0.49
Drug use	0.40	0.37	0.49
Chronic physical disease	0.62	0.62	0.61
Prior inpatient - homelessness	0.12	0.10	0.15
Prior inpatient - mental health	0.23	0.18	0.31
Voucher receipt	0.12	1.0	0.0
Transitional housing	0.18	0.0	1.0
N	289,933	34,285	52,901

Note: This table shows mean characteristics for the full sample, voucher recipients, and transitional housing recipients. Voucher and transitional housing refers to enrolling in that program within the first 3 months of intake.

Table 1.2: Complier groups

Complier group	$\mathbf{D}_i(z_v, z_t) =$ $(D_i(0,0), D_i(1,0), D_i(0,1), D_i(1,1))$	Description
C_0^v	$(0, v, 0, v)$	Compliers who only respond to $Z_{v,i}$
C_0^t	$(0, 0, t, t)$	Compliers who only respond to $Z_{t,i}$
$C_0^{v+,t-}$	$(0, v, t, v)$	Compliers who respond to both, but $Z_{v,i}$ is more compelling
$C_0^{v-,t+}$	$(0, v, t, t)$	Compliers who respond to both, but $Z_{t,i}$ is more compelling
C_t^v	(t, v, t, v)	Compliers who always respond to $Z_{v,i}$, with a counterfactual of t
C_t^{v-}	(t, v, t, t)	Compliers who respond to $Z_{v,i}$ only when $Z_{t,i} = 0$, with a counterfactual of t
C_v^t	(v, v, t, t)	Compliers who always respond to $Z_{t,i}$, with a counterfactual of v

Note: This table defines allowed complier groups according to my treatment choice model. I convert both instruments to binary variables for simplicity, but one can think of $z_v = 1$ as high voucher availability. Column 2 shows potential treatments for each complier group when z_v and z_t are both 0, when $z_v = 1$ and $z_t = 0$, when $z_v = 0$ and $z_t = 1$, and when both instruments equal 1. Column 3 provides a conceptual description. Note that under this model, $\mathbf{D}_i(z_v, z_t) = (v, v, t, v)$ is not a possible behavioral response pattern. For these individuals, $O_{v,i}(0) = 1$ (given that they choose v when $z_v = 0$). When $(z_v, z_t) = (0, 1)$, they choose t, which implies that $\arg\max_{k \in \{0,v,t\}} U_{k,i} = t$. Thus, when $(z_v, z_t) = (1, 1)$, they must also choose t, as their choice set and preferences remain unchanged.

Table 1.3: First stage and reduced form

	(1) Voucher	(2) Transitional	(3) Neither	(4) 3-year mortality
Perc. receiving voucher	0.510	-0.050	-0.460	-0.02316
	(0.017)	(0.016)	(0.020)	(0.00861)
Tr. liaison available	-0.013	0.089	-0.076	-0.00004
	(0.002)	(0.002)	(0.002)	(0.00001)
Dep. var. mean	0.118	0.182	0.699	0.066
Observations	289,932	289,932	289,932	289,932

Note: This figure shows first stage (columns 1-3) and reduced form (column 4) regression coefficients for $Z_{v,i}$ (perc. receiving voucher) and $Z_{t,i}$ (tr. liaison available). In columns 1-3, indicators for enrolling in the voucher program, transitional housing, or neither are the left-hand side variables. In column 4, 3-year mortality is the left-hand side variable. Baseline controls include VA station × year, VA station × month, and indicators for having a prior inpatient stay related to homelessness or mental health. Standard errors for all regressions are clustered by VA station × year × month.

Table 1.4: Effect of voucher and transitional housing on mortality

	(1) OLS	(2) IV	(3) All controls	(4) Drop SSVF	(4) Drop prior inpatient
Voucher	-0.013	-0.046	-0.046	-0.045	-0.058
	(0.001)	(0.017)	(0.016)	(0.018)	(0.019)
Trans. housing	-0.004	-0.007	-0.012	-0.012	-0.003
	(0.001)	(0.015)	(0.014)	(0.015)	(0.019)
Dep. var. mean	0.066	0.066	0.066	0.066	0.063
Observations	289,932	289,932	289,932	281,474	219,937

Note: This table shows the main results for the effect of the voucher program and transitional housing on three-year mortality. Column 1 shows OLS results, and column 2 shows 2SLS IV results. Column 3 shows 2SLS regressions when I add all patient characteristics as control variables. Column 4 shows the estimates when I drop patients who enroll in SSVF, or temporary rental assistance services. Finally, column 5 shows results when I drop patients with a prior inpatient mental health or homelessness stay. Baseline controls include VA station $\times$ year, VA station $\times$ month, and indicators for having a prior inpatient stay related to homelessness or mental health. Standard errors for all regressions are clustered by VA station $\times$ year $\times$ month.

Table 1.5: Overidentification tests

Additional instruments		Voucher	Transitional housing
Male $\times Z_{v,i}$	2SLS	-0.044	-0.006
Male $\times Z_{t,i}$		(0.017)	(0.015)
	First stage F	234	460
	Overid. p-value	0.304	
Black $\times Z_{v,i}$	2SLS	-0.039	-0.007
Black $\times Z_{t,i}$		(0.016)	(0.015)
	First stage F	248	465
	Overid. p-value	0.279	
Hispanic $\times Z_{v,i}$	2SLS	-0.047	-0.007
Hispanic $\times Z_{t,i}$		(0.017)	(0.015)
	First stage F	228	466
	Overid. p-value	0.818	
Depression $\times Z_{v,i}$	2SLS	-0.044	-0.005
Depression $\times Z_{t,i}$		(0.017)	(0.015)
	First stage F	227	462
	Overid. p-value	0.585	
Psychoses $\times Z_{v,i}$	2SLS	-0.049	-0.007
Psychoses $\times Z_{t,i}$		(0.017)	(0.015)
	First stage F	227	462
	Overid. p-value	0.867	
Alcohol use $\times Z_{v,i}$	2SLS	-0.044	-0.011
Alcohol use $\times Z_{t,i}$		(0.017)	(0.015)
	First stage F	227	475
	Overid. p-value	0.463	
Drug use $\times Z_{v,i}$	2SLS	-0.048	-0.011
Drug use $\times Z_{t,i}$		(0.018)	(0.015)
	First stage F	229	461
	Overid. p-value	0.184	
High $\hat{Y}_i \times Z_{v,i}$	2SLS	-0.047	-0.011
High $\hat{Y}_i \times Z_{t,i}$		(0.016)	(0.015)
	First stage F	230	465
	Overid. p-value	0.153	
All interactions	2SLS	-0.039	-0.006
(16 additional		(0.017)	(0.016)
instruments)	First stage F	54	97
	Overid. p-value	0.592	

Note: This table shows 2SLS and overidentification test (J-test) results when I add additional instruments to the model. For a given characteristic X_i, I construct additional instruments as $Z_{v,i}X_i$ and $Z_{t,i}X_i$ and add these to the baseline 2SLS regressions. To maintain instrument validity in these specifications, I also interact X_i with the baseline time $\times$ station controls. High $\hat{Y}_i$ refers to above median predicted mortality.

Table 1.6: 2SLS bounds under small violations to no essential heterogeneity

	(1)	(2)
Epsilon	Voucher	Transitional housing
0.010	(-0.049, -0.043)	(-0.010, -0.004)
0.005	(-0.047, -0.045)	(-0.008, -0.006)
0.001	(-0.046, -0.046)	(-0.007, -0.007)

Note: This table shows estimated upper and lower bounds on true LATEs if there are small violations to no essential heterogeneity, or under the condition that $|E[Y_{k,i} - Y_{0,i}|C_i = c] - E[Y_{k,i} - Y_{0,i}|C_i = c']| < \epsilon$ for various values of epsilon. In column 1, I show bounds on $LATE_v = E[Y_{v,i} - Y_{0,i}|C_0^v]$, the LATE of Housing First (compared to no intervention) for C_0^v compliers. In column 2, I show bounds on $LATE_t = E[Y_{t,i} - Y_{0,i}|C_0^t]$, the LATE of Treatment First for C_0^t compliers. I derive these bounds in Appendix A.3.

Table 1.7: Exclusion restriction tests

	(1) Baseline	(2) Number of staff	(3) Homelessness services speed	(4) Emergency housing use	(5) Case management
Voucher	-0.046	-0.050	-0.049	-0.046	-0.046
	(0.017)	(0.019)	(0.018)	(0.017)	(0.017)
Trans. housing	-0.007	-0.011	-0.008	-0.007	-0.006
	(0.015)	(0.016)	(0.015)	(0.015)	(0.015)
Dep. var. mean	0.066	0.066	0.066	0.066	0.066
Observations	289,932	289,932	289,923	289,932	289,932

Note: This table shows the effect of the voucher program and transitional housing on three-year mortality when I add additional controls related to exclusion. Column 1 shows the baseline result, column 2 controls for the number of staff working at intake, and column 3 controls for the (leave out) average wait time to a follow up appointment with a social worker. Next, column 4 controls for use of emergency housing, and column 5 controls for use of case management. Baseline controls include VA station × year, VA station × month, and indicators for having a prior inpatient stay related to homelessness or mental health. Standard errors for all regressions are clustered by VA station × year × month.

Table 1.8: Cause of death

	(1) Exposure to the elements	(2) Violence/assault	(3) Suicide or overdose	(4) HIV or Hepatitis B/C	(5) Other causes of death
Voucher	-0.0016	-0.0051	-0.0088	-0.0059	-0.0251
	(0.0006)	(0.0018)	(0.0069)	(0.0026)	(0.0159)
Trans. housing	-0.0006	-0.0005	0.0029	-0.0002	-0.0087
	(0.0007)	(0.0013)	(0.0059)	(0.0020)	(0.0136)
Dep. var. mean	0.0001	0.0006	0.0098	0.0014	0.0537
Observations	289,932	289,932	289,932	289,932	289,932

Note: This table shows the mortality effect broken up by cause of death. HIV, Hepatitis B, and Hepatitis C are often caused by injection drug users sharing needles. Exposure to the elements refers to extreme heat or cold. Baseline controls include VA station × year, VA station × month, and indicators for having a prior inpatient stay related to homelessness or mental health. Standard errors for all regressions are clustered by VA station × year × month.

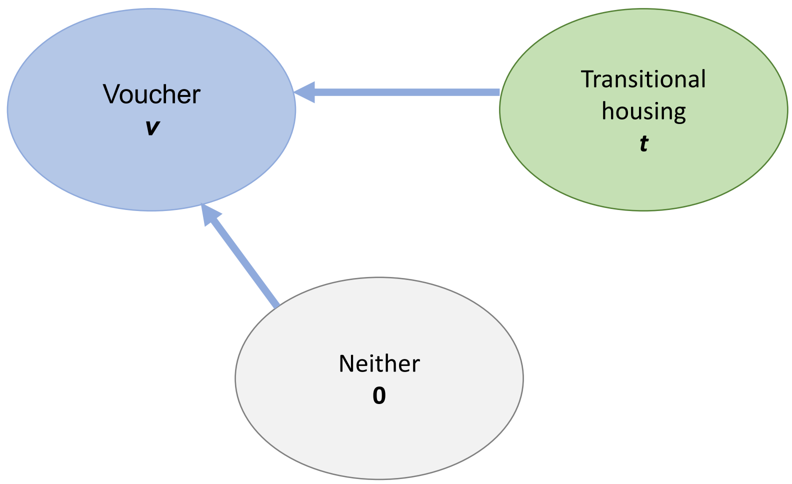
Table 1.9: Housing First cost-benefit analysis

Program costs		Program savings		Total
Housing costs	Social worker visits	Shelter stays	Inpatient stays	
$13,380	$9,667	-$246	-$41,734	-$18,931
($407)	($885)	($92)	($16,429)	

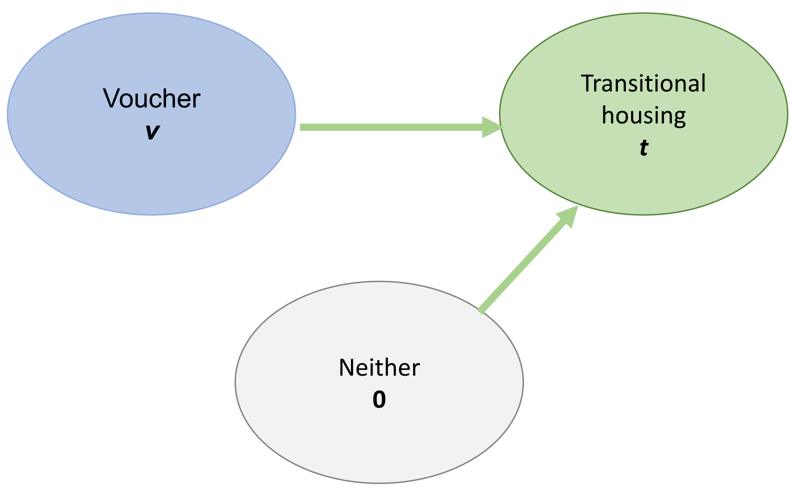
Note: This table shows the costs of funding HUD-VASH over the three year study period, including cost-savings from reduced use of other programs. I include costs incurred by the VA and HUD, which is what I can observe. Housing costs are based on the number of months HUD-VASH participants were actively using a housing voucher, based on social worker documentation (compared to the no-intervention group). Social worker visit costs and the cost of inpatient stays come from estimates from the VA's Health Economics Resource Center (HERC) and are based on Medicare reimbursement rates.

Figure 1.1: Treatment flows induced by $Z_{v,i}$ and $Z_{t,i}$

A: Flows induced by $Z_{v,i}$

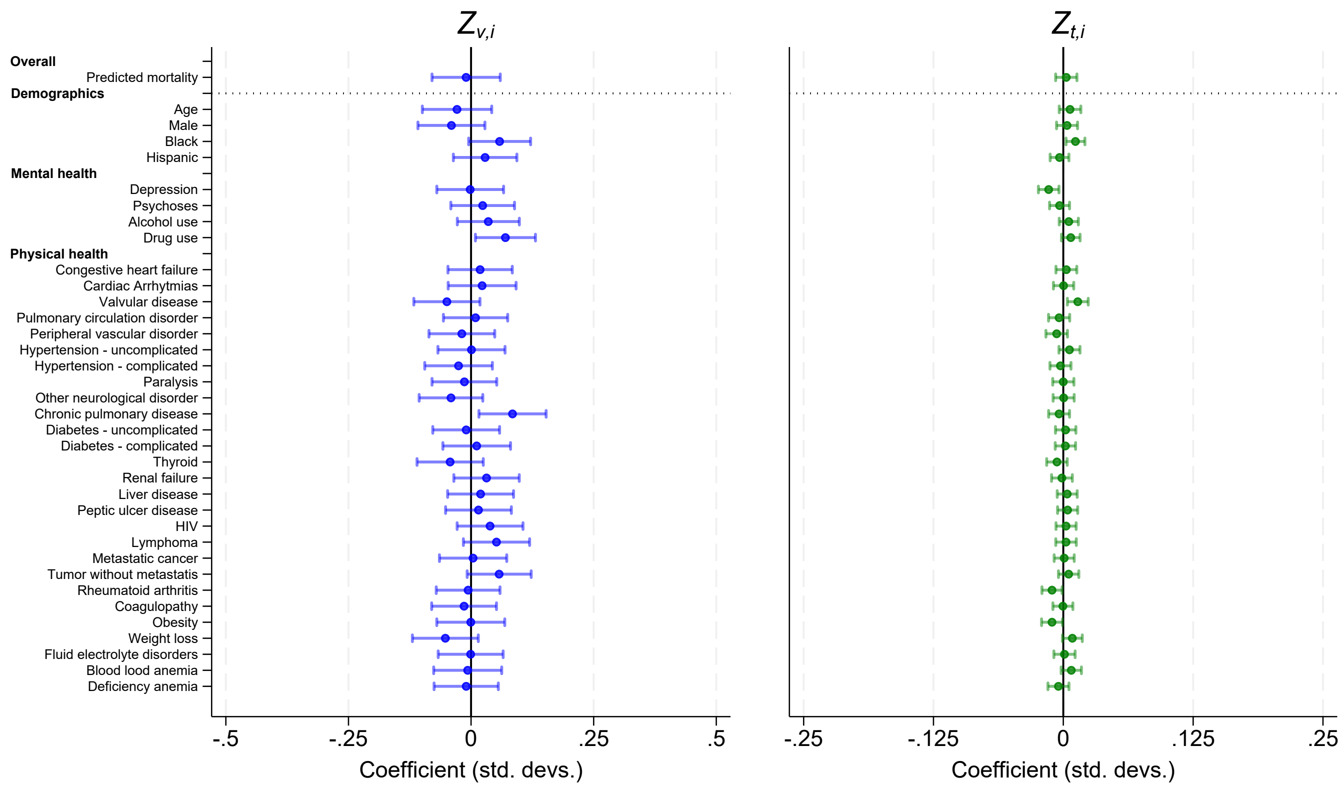


B: Flows induced by $Z_{t,i}$



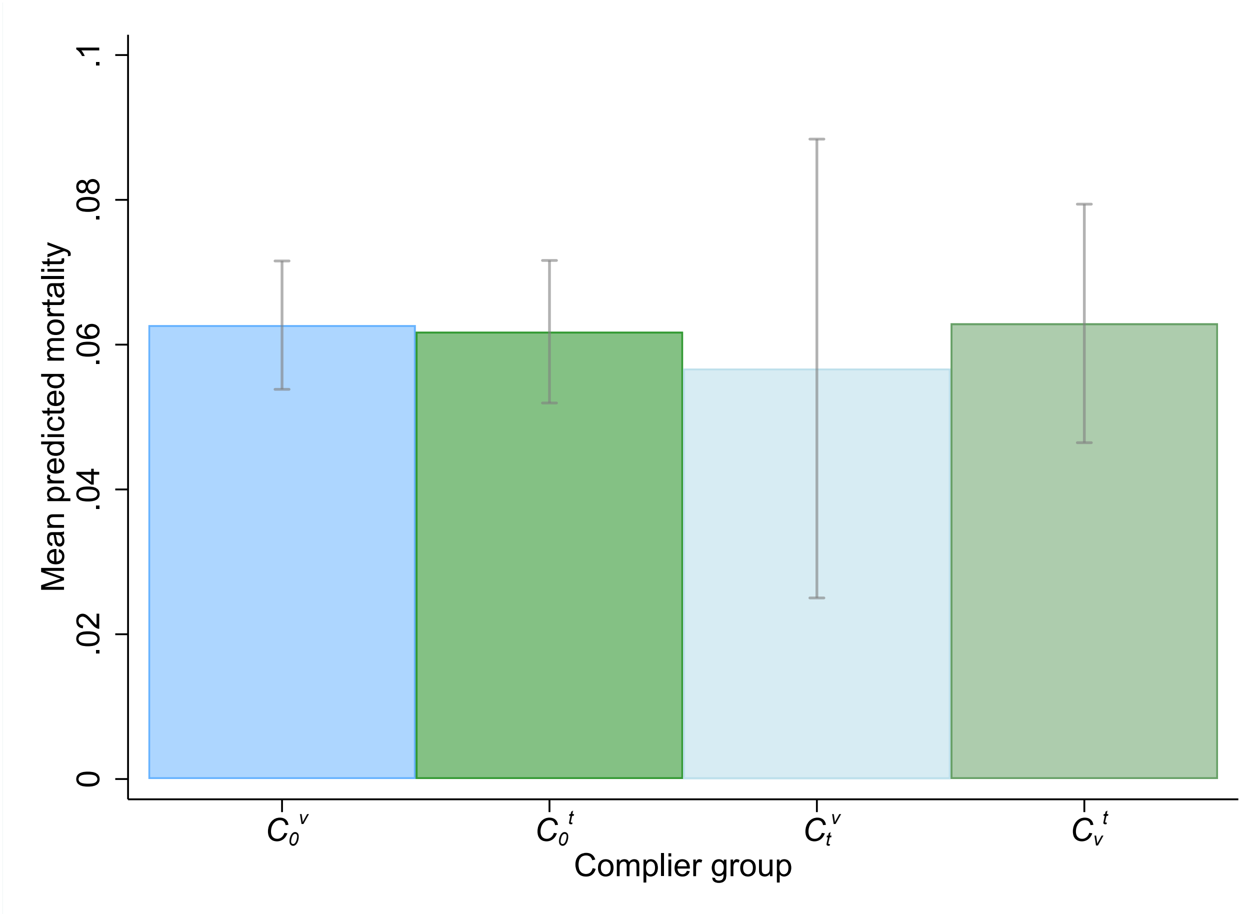
Note: This figure shows allowed flows between treatment states induced by $Z_{v,i}$ and $Z_{t,i}$, according to my treatment choice model.

Figure 1.2: Balance tests



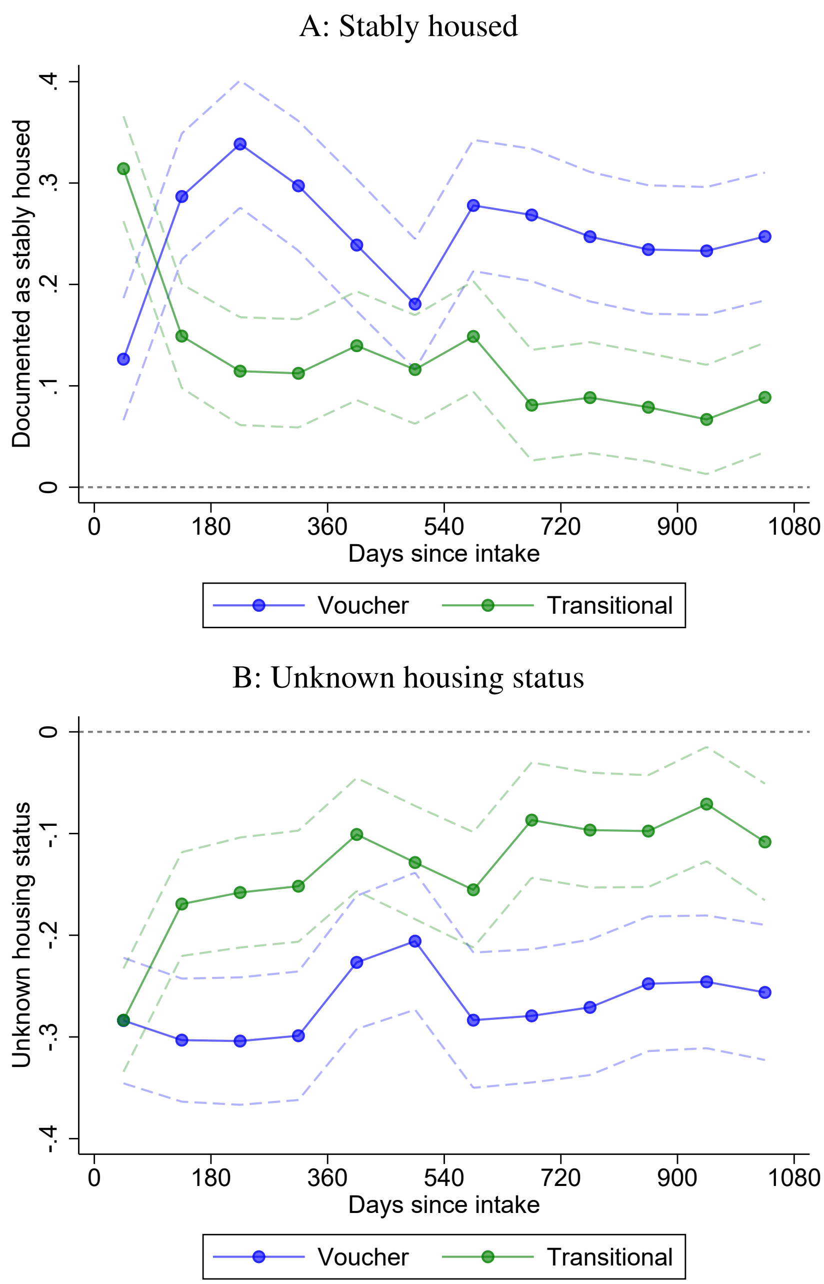
Note: This figure shows balance tests when a given characteristic X_i is the left hand side variable. Dots are point estimates, and bars show standard errors. Results for $Z_{v,i}$ are shown in blue, and results for $Z_{t,i}$ are shown in green. Predicted mortality comes from an OLS regression of 3-year mortality on all characteristics X_i. Baseline controls include VA station $\times$ year, VA station $\times$ month, and indicators for having a prior inpatient stay related to homelessness or mental health. Standard errors for all regressions are clustered by VA station $\times$ year $\times$ month.

Figure 1.3: Mean predicted mortality across complier groups



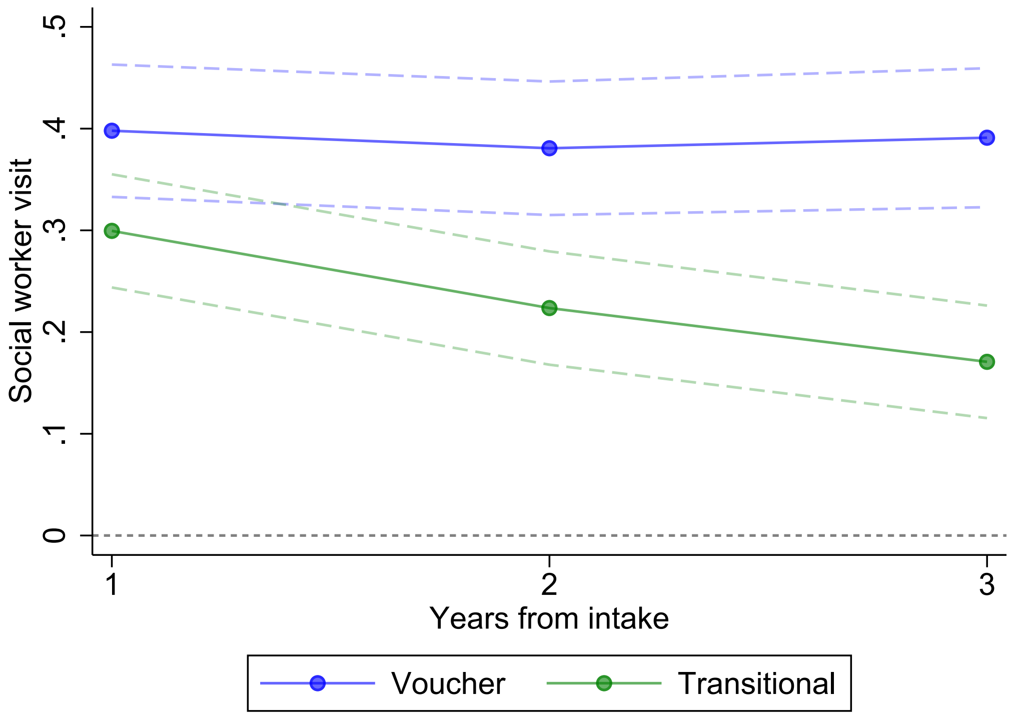
Note: This regression shows mean complier predicted mortality for C_0^v, C_0^t, C_t^v, and C_v^t compliers, respectively.C_0^v compliers are those induced into v from a counterfactual of 0, C_0^t compliers are those induced into t from a counterfactual of 0, C_t^v compliers are those induced from t to v, and C_v^t compliers are those induced from v to t.

Figure 1.4: Effect of voucher and transitional housing on housing status over time



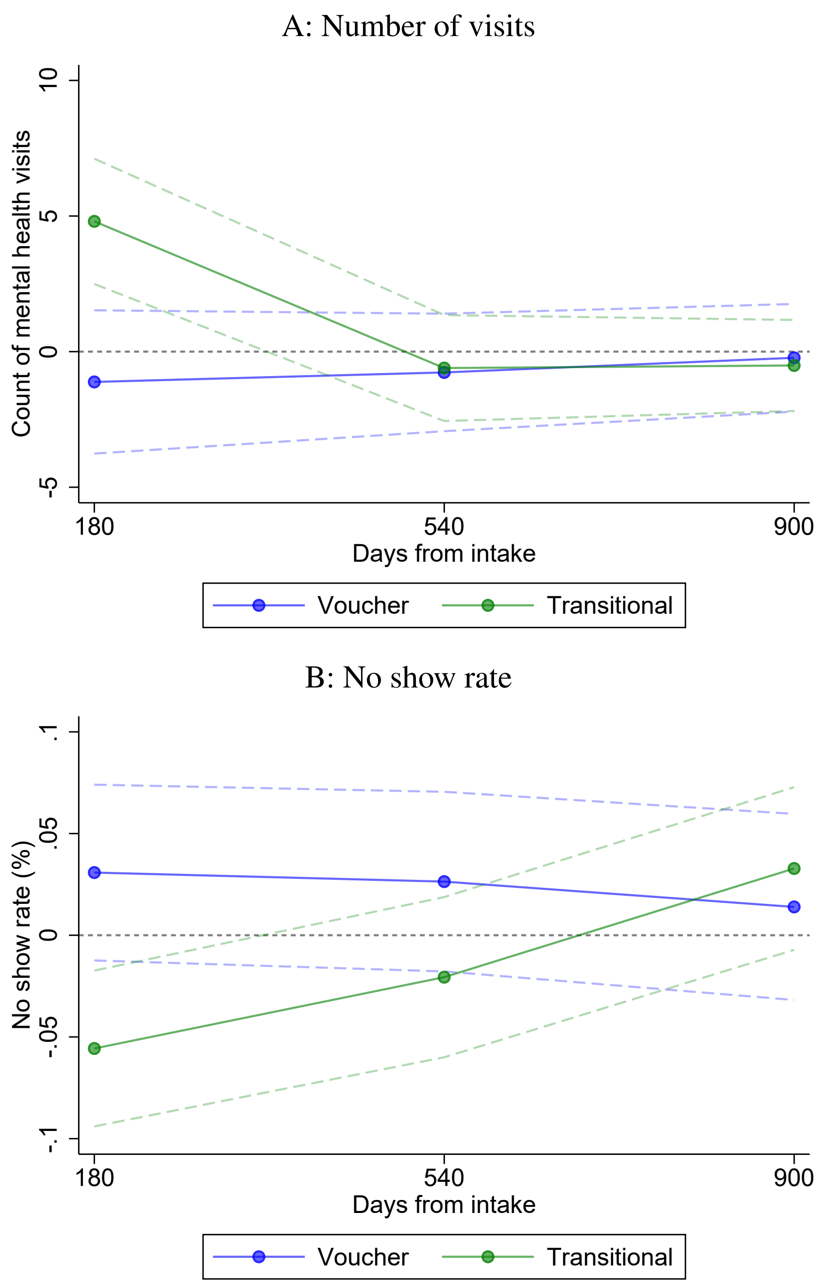
Note: This figure shows 2SLS regression results for the effect of the voucher program and transitional housing on housing status over time. Housing status is broken into quarters from intake. Panel A shows the effect on the probability of being stably housed within that quarter. Panel B shows the effect on the probability of having unknown housing status that quarter, meaning that housing status was not clearly recorded by a social worker. Results for the voucher program are shown in blue, and results for transitional housing are shown in green. Dotted lines show upper and lower bounds.

Figure 1.5: Engagement with VA homelessness services



Note: This figure shows 2SLS regression results for the effect of the voucher program and transitional housing on engagement with homelessness services in years 1, 2, and 3 following intake. The dependent variable is having any visit with a social worker in that year. Results for the voucher program are shown in blue, and results for transitional housing are shown in green. Dotted lines show upper and lower bounds.

Figure 1.6: Engagement with VA mental health services



Note: This figure shows 2SLS regression results for the effect of the voucher program and transitional housing on engagement with mental health care. Panel A shows the effect on the number of attended mental health visits. Panel B shows the effect on the no show rate, defined as the percent of visits not attended over the total number scheduled. Results for the voucher program are shown in blue, and results for transitional housing are shown in green. Dotted lines show upper and lower bounds.

Chapter 2

How do mental health treatment delays impact long term mortality?

2.1 Introduction

The mental health crisis is increasingly a point of concern in public discourse. Those suffering from severe mental illness die an average of 10 to 20 years earlier than the general population (Liu et al. 2017). Mental health treatment has been increasingly destigmatized in recent years, with the number of patients seeking new therapy continuing to rise (APA 2021). Medical research has shown that behavioral therapy and medication are successful at achieving their primary goal – reducing depressive symptoms and improving general mental health (Cuijpers et al. 2016, Gartlehner et al. 2017, Holmes et al. 2018) – but what if patients most in need have difficulty accessing care?

There continue to be reports of a major shortage of mental healthcare providers across geographic regions in the US, often resulting in long appointment delays for patients seeking care (Weiner 2022). Provider shortages are exacerbated by a healthcare system that has traditionally put physical health before mental health. In 2008, the US government passed the Paul Wellstone and Pete Domenici Mental Health Parity and Addiction Equity Act, intended to address treatment disparities between mental and physical health conditions, especially related to insurance coverage. Nevertheless, patients continue to report more difficulty accessing in network providers for mental health treatment, resulting in either higher copayments or longer wait times (Xu et al. 2019, Caron 2021). Given this context, waiting times for mental health treatment are of notable policy concern. The National Committee for Quality Assurance (NCQA) collects data on mental health wait times following ED visits – the setting of my analysis – considering them to be a key quality metric.

Issues surrounding access to and engagement with mental health services are particularly pertinent to veterans, who suffer in high rates from PTSD, substance use disorder (SUD), and other related issues. Of all government sponsored healthcare programs, the Department of Veterans Affairs (VA) provides the most comprehensive mental health services; treatment is free for the majority of low-income veterans. Despite comprehensive insurance covering mental health, veterans still report difficulty accessing behavioral therapy, notably due to long wait times (NASEM 2018). With its detailed demographic data and large patient population vulnerable to mental health crises, VA data provide a unique opportunity to study the impact of mental health treatment delays on long term mortality.

In this paper, I study patients who arrive at any Emergency Department (ED) across the VA system with a mental health-related condition. I collect wait times for these patients, defined as the difference between their ED visit date and their first follow-up psychology or psychiatry appointment. A patient's actual wait time is likely to be correlated with other characteristics affecting survival – notably, the patient's severity level and willingness to schedule future appointments. Thus, in order to isolate the causal effect of treatment delays, I do not use a patient's own wait time directly. Instead, I construct an *exogenous* measure for each patient: the "leave-out" mean wait time – the average wait time of all other patients who arrive in same two-week period and hospital as a given patient and were assigned the same type of follow up appointment (psychiatry, psychology, substance use disorder clinic, etc.).[1] As mean wait time is assigned conditionally based on hospital, time, and appointment type, I control for time in two-week intervals, mental health clinic type, and hospital × year-month fixed effects in all analyses.

Conceptually, one can think of mean wait time as a measure of mental health clinic congestion, which affects both providers' and patients' ability to schedule timely follow up visits. I first show that mean wait time is highly correlated with a patient's actual wait time, even after controlling for the design controls. Moreover, using leave-out mean wait time rather than a patient's own wait time eliminates the potential concerns that either 1) patients play some role in appointment scheduling, or 2) physicians schedule follow up appointments based on a patient's risk level, rather than just appointment availability. I find that mean wait time has a very small and statistically insignificant correlation with predicted patient mortality. In other words, patients who have higher and lower predicted mortality are equally likely to be assigned a high mean wait time, indicating quasi-random assignment.

As mean wait time is uncorrelated with patient risk, my measure of congestion enables me to isolate the causal impact of mental health treatment delays on long term mortality. I show that a one standard deviation increase in mean wait time increases two-year mortality by 0.113 percentage points (s.e. = 0.0041), or about 1.5% of baseline mortality risk. This effect is highly robust to controlling for a detailed range of patient characteristics – including demographics, ICD-9 codes from the ED visit, and prior mental health utilization, among others. Moreover, the mortality effect is long-lasting, persisting for at least five years after the index ED visit. The results are almost entirely driven by patients with substance use disorders (SUD) or psychosis, rather than those with anxiety, depression, or PTSD.

How do relatively small treatment delays have a impact on long term mortality? Prior research has shown that those who are placed into timely mental health care are more likely to stay in consistent treatment (Kreyenbuhl et al. 2009). Similarly, I find that my measure of congestion, mean wait time, strongly predicts whether a patient attends their follow up mental health visit: a one standard deviation increase in mean wait time (11.7 days) leads to a 2.0 percentage point reduction in the probability of attending one's follow up appointment (t=24). Thus, I argue that the mortality effect is primarily driven by patients' engagement with mental health treatment.

However, I also consider competing explanations. I first show that mean (mental health) wait time is not correlated with other hospital wait times or mortality of non-mental health patients. These results imply that my measure of mental health clinic congestion only affects health out-

[1] The VA often has separate clinics and providers for different types of specialty mental health appointments. Therefore, patient-specific wait times are highly dependent on appointment type.

comes via mental health care itself, rather than being correlated with overal hospital quality or crowding. Additionally, I find that mental health clinic congestion does not predict the length of follow up appointments, indicating that patients do not receive more rushed care when appointments are delayed.

Nevertheless, I do find some changes in care patterns on the supply side, which may also contribute to the mortality effect. First, patients who arrive when mental health clinics are congested are assigned providers with slightly lower experience. Moreover, physicians do appear to be aware of and respond to clinic congestion. They compensate for appointment delays by prescribing more mental health drugs at the ED visit itself; a one standard deviation increase in leave-out wait time (11.7 days) increases the probability of being prescribed a drug at the ED visit by 1.2 percentage points (s.e.=0.08) or 4.0%. As prescriptions mitigate mortality risk, prescribing patterns at the ED partially counteract the impact of missing future behavioral therapy appointments.

Rather than being the result of a single missed appointment, the mortality effect is likely a product of disengagement from the mental health care system as whole. In fact, patients with higher mean wait times are far more likely to permanently disengage from both outpatient and inpatient mental health care in the long run. Overall, my results suggest that shorter hospital wait times can be vital in nudging patients into appointment attendance, and that missing a single mental health visit can influence long run engagement with therapy and consequently, mortality.

This study contributes to the literature on the effects of mental health treatment. Previous research has considered outcomes such as employment and career earnings (Biasi et al. 2020, Bhat et al. 2022), high school completion and human capital development (Cuellar and Dave 2015, Angelucci and Bennett 2022), crime (Blattman et al. 2022), suicide (Lang et al. 2013), and mortality (Holliday 2018). Most similar to this paper, a recent medical study finds that quasi-random decreases in mental health staffing increases suicide rates among VA patients (Feyman et al. 2022). These results hint at the issues congestion may cause; however, the authors cannot identify the downstream effects of low provider supply. I show that delays to care have a direct impact on whether patients stay engaged with mental health services. Additionally, in investigating all-cause mortality, rather than suicide alone, this study takes a broader look at the way mental health treatment delays can influence one's overall health status.

Moreover, this study contributes to work addressing the efficiency effects of longer wait times in the health care system. In exploiting quasi-random variation in wait times, I build on a similar empirical strategy as that of Williams and Bretteville-Jensen (2022) and Gï¿œdoy et al. (forthcoming), who study wait times in SUD treatment and orthopedic surgery, respectively. Wait times are a common rationing mechanism for healthcare; unsurprisingly, there is evidence that delays have negative impacts on health outcomes (e.g., Reichert and Jacob 2018, Gruber et al. 2018, Singh and Venkataramani 2022, Turner et al. 2022). Concerns over rationing care may be particularly relevant in cases where cost-sharing is low, as reduced cost barriers to accessing healthcare may tip demand for healthcare services to exceed supply. Low cost sharing is both relevant to the VA, where care is essentially free for low income veterans, as well as for public, single payer healthcare systems, such as the National Health Service (NHS) in the United Kingdom. Cutler (2002) describes how this creates a trade off between efficiency and equity: although removing financial incentives likely improves access for low income patients, it may lead to longer waiting times and more inefficient healthcare delivery. This study provides additional evidence that longer wait times as a result of clinic crowding worsens health outcomes.

Related to the use of wait time as a rationing tool in healthcare is the broader idea that non-price ordeals affect targeting. In the healthcare system, both longer wait times and higher copays can both potentially act as a demand-reducing mechanism; however, these mechanisms likely deter different types of patients from seeking care. In in a preventative health context, Dupas et al. (2016) actually show a case where increasing nonmonetary costs actually improves targeting. In contrast, Deshpande and Li (2023) find that closing Social Security field offices, thus increasing the nonmonetary costs of applying for disability, has negative effects on targeting. Analagously, in the VA setting, clinic congestion deters those on the margin from seeking care, even though these patients may see large health benefits from attending mental health treatment.

The rest of this paper proceeds as follows: Section 2.2 describes the setting and data. Section 2.3 describes the empirical approach. Section 2.4 presents the results, including examining the validity of the wait time measure, as well as exploring heterogeneity and mechanisms. Section 2.5 concludes.

2.2 Setting and Data

2.2.1 The Department of Veterans Affairs

The US Department of Veterans Affairs (VA) is the only public, fully integrated healthcare system in the US, acting as both an insurer and provider of healthcare. The agency provides low-cost medical care to veterans meeting certain criteria, including those with low incomes and/or a a service-connected disability. Veterans are a unique patient population with specific needs. Notable for this study, veterans suffer in high rates from substance use disorder, post-traumatic stress disorder (PTSD), and other mental health conditions. Moreover, veterans face a suicide risk more than 50% higher than that of the general population (VA Office of Mental Health and Suicide Prevention 2022). Veterans are also more likely to suffer from chronic diseases such obesity, diabetes, and heart disease compared to non-veterans, even when controlling for age (Betancourt 2023).

Mental health treatment is free for all veterans with a behavioral health condition connected to their service; all other veterans pay small copays for treatment. Thus, fees generally are not a barrier for this population. However, wait times for mental health treatment are perceived as a signficant obstacle, with less than half of veterans requesting urgent appointments reporting being able to get appointments when they requested. A 2015 VA policy aims to get all new patients seeking behavioral therapy an appointment within a 30-day time frame (NASEM 2018).

2.2.2 Patient encounter, characteristics, and mortality data

The Corporate Data Warehouse, the VA's detailed health record system, contains patient information on all inpatient and outpatient encounters, International Classification of Diseases Clinical Modification, 9th Revision (ICD-9) diagnosis codes connected with each appointment, and demographic characteristics such as date of birth, race, ethnicity, and gender. Relevant to my study design, the VA also records all appointment scheduling requests, and whether these appointments were ultimately cancelled or no-showed.

The primary outcome measure for this study is two-year mortality. Data on mortality from the VA is particularly reliable in that it combines data from the Veterans Benefits Administration (VBA), Centers for Medicare and Medicaid Services (CMS), and Social Security Administration

(SSA) for a combined estimated sensitivity of 98%. Since the effect of treatment may take longer to appear than that of other medical interventions, using a longer time horizon to assess mortality is typical of mental health studies (Appleby et al. 1999, Bjorkenstam et al. 2011). Matching on Social Security Numbers, I also link patient data from the VA with cause of death data, coded by ICD-9 or ICD-10 code, from the National Death Index. As intermediate outcomes, I assess prescriptions for mental health drugs as well long term engagement with both mental and physical health services, as falling out of care is a major concern for this population of interest (Nosï¿œ et al. 2003, Kreyenbuhl et al. 2009).

2.3 Empirical approach

2.3.1 Sample selection

The primary sample consists of all patients with a VA Emergency Department (ED) visit for a mental health condition between January 1, 2001 and February 28, 2015, allowing me to follow all patients up to five years prior to the start of the COVID-19 pandemic. I define a mental health ED visit as any ED visit assigned a mental health diagnosis code (ICD-9 code). Reasons for visiting the emergency room range from thoughts of hurting oneself to an upper GI bleed due to alcohol use. The most common mental health diagnoses at the ED in my sample are general depressive disorder (19% of cases), general anxiety (16% of cases), alcohol use (14% of cases), and adjustment reaction (10% of cases). Additionally, 39.9% of the sample also has a *physical* health diagnosis connected to their ED visit, the most common of which are hypertension, sleep disturbances, and respiratory distress. Some of these may be directly connected to mental health – for instance, those with anxiety may experience breathing issues or difficulty sleeping.

In order to follow patients over a long time horizon, I use only the first ED mental health visit for each patient. Additionally, I drop patients who have no scheduled mental health visit within 90 days of their initial ED visit, my definition of a follow up appointment.[2] These patients are dropped as a result of the empirical design, where mean wait time (described in Section 2.3.2) is constructed according to follow up appointment type (psychiatry, psychology, etc.). Appendix Table B.1 describes patients with no scheduled follow up appointments in more detail, who are have higher rates of SUD and have less prior attachment to VA mental health services. In Table 2.2 and Appendix Section A3.1, I also perform robustness checks where I add these patients back to the sample using a slightly modified version of the main treatment variable and controls.

Figure 2.1 describes the sample selection process. The baseline analytical sample consists of 621,289 patients across 128 hospitals.

2.3.2 Measuring congestion

To measure mental health clinic congestion, I collect wait times for all patients with a mental health ED visit. I define a patient's wait time as the time between their initial ED visit and their follow up mental health appointment. I define a follow up appointment as the patient's first scheduled

[2] In VA data, outpatient encounters with mental health services are not explicitly linked to ED visits as follow ups. Thus, follow ups are likely measured with some error using this time window definition.

encounter with a mental health specialist following their initial ED visit. All appointments within a 90 day window are considered follow up appointments.[3]

While there is likely variation in scheduling practices across clinics, ED doctors work directly with scheduling staff to coordinate follow up appointments. These appointments may be scheduled on site before a patient is discharged from the ED, or they may be scheduled with outpatient scheduling staff via phone following their index visit. The length of time between a patient's ED visit and their follow up appointment date is likely a product of various factors, including provider availability, patient scheduling preferences, and patient severity level. Thus, to construct a measure of clinic congestion orthogonal to a given patient's risk level, I do not use a patient's own wait time. Instead, I group together patients at the same hospital who arrive during the same two-week period and are assigned the same type of follow up appointment (psychology, psychiatry, substance use disorder clinic, etc.). Wait times are appointment specific in that different specialty mental health clinics treat different types of patients, and thus, congestion is specific to appointment type. I assign each patient a mean wait time value, W_{-i}, according to the mean leave-out wait time of all patients in their respective hospital-time-appointment type group, excluding the patient's own wait time. More formally, for patient i and hospital-time-appointment group j, W_{-i} is defined as follows:

$$W_{-i} = \frac{1}{|I_{j(i)}| - 1} \sum_{i' \in I_{j(i)}} \mathbf{1}(i' \neq i) W_{i'} \tag{2.1}$$

where W_i is a patient's wait time and $I_{j(i)}$ is the set of patients in the same hospital-time-appointment type group as patient i. W_{-i} is a measure of mental health clinic congestion at the time of a patient's arrival to the ED. To more accurately reflect congestion, rather than restricting to the first visit for each patient (as is the case for the primary sample; see Figure 2.1), I use all mental health ED visits and their corresponding wait times to form W_{-i}. The average cell size of $j(i)$ is 22 patients; mean W_{-i} is 17.7 days (standard dev. = 11.7 days). I standardize W_{-i} in regression analyses so that coefficients can be interpreted as the impact of a one-standard deviation increase in mean wait time.[4]

While a patient's own wait time may be influenced by patient specific-scheduling factors that are correlated with risk level, mean wait time captures potential appointment delays purely due to congestion. Mean wait time is assigned conditionally based on hospital, time, and mental health clinic type. Thus, I control for two-week period, year-month $\times$ hospital fixed effects, and appointment type in all analyses. Additionally, I add four-year age buckets to the baseline controls, necessary to achieve balance (see Section 2.4.1). With these baseline controls, the standard deviation of residualized mean wait time is approximately 7.2 days. Across analyses, I cluster standard errors by hospital-month.

Ultimately, the study design relies on the assumption that conditional on the design controls, particularly the detailed time $\times$ hospital controls, the level of congestion a given mental health

[3] In other words, if the patient's first mental health encounter following their ED visit occurs after 90 days, I consider this visit to not be a true follow up appointment.

[4] Note that I use only patients with an ED mental health visit to form W_{-i}. Although there are many patients with scheduled mental health visits who do not have an ED visit, my definition of wait time relies on calculating the number of days between the ED visit and follow up visit date. Moreover, timing of follow ups for emergent patients are more likely to reflect congestion, whereas visits for non-emergent patients engaged with behavioral therapy are more likely to follow a recurrent schedule.

clinic experiences is as good as random. More specifically, for a given patient that arrives to the ED when mental health clinic congestion is particularly high, I compare outcomes for this patient with other patients who arrive to the *same* ED in the *same* month (e.g., the Palo Alto VA in November 2013), accounting for patient selection to certain hospitals in particular months. Moreover, I use two-week period fixed effects to adjust for the fact that particular time periods may be systematically busier. Finally, certain appointment types (such as psychiatry) may tend to have both sicker patients and longer wait times. Thus, I control for appointment type indicators to adjust for any systematic correlation between patient selection and type. I further reassure the reader that selection into certain appointment types does not drive the results in Section 2.4.3.

2.4 Results

2.4.1 Design validity

Does mean wait time predict actual wait time?

In order for W_{-i} to be a useful measure of mental health treatment delays, it must be highly correlated with a patient's actual wait time. To assess the relationship between mean wait time and patient-specific wait time, in Figure 2.2 Panel A, I show a binned scatterplot of the relationship between W_i and W_{-i}, conditional on the baseline controls. I run the following regression to assess the predictive power of the treatment variable:

$$W_i = \pi_1 W_{-i} + \zeta_{1,t(i)} + \phi_{1,a(i)} + \gamma_{1,h(i)m(i)} + \mathbf{X}_i\delta_1 + \epsilon_{1,i} \quad (2.2)$$

where W_i is actual wait time, W_{-i} is mean wait time, $\zeta_{1,t(i)}$ are two-week period fixed effects,[5] $\phi_{1,a(i)}$ are appointment type fixed effects, $\gamma_{1,m(i)h(i)}$ are hospital × year-month fixed effects, and $\mathbf{X}_i$ are additional patient demographic controls, which include age buckets (in four-year intervals) in the baseline specification. Appointment types are categorized as follows: general mental health visit (individual, group, or phone visit), psychiatry, psychology, SUD specialty clinic (individual or group), PTSD specialty clinic (individual or group), mental health integrated care, or other specialty clinic type.[6] Excluding reference categories, there are 19,137 hospital-year-month fixed effects, 368 two-week period fixed effects, 10 appointment type fixed effects, and 20 age buckets.[7]

Even with the design controls, which should eliminate non-random correlation between W_i and W_{-i} due to shared hospital, appointment type, and time period, I find that W_i and W_{-i} are highly correlated: a one standard deviation increase in mean wait time (11.7 days) increases actual wait time by about 1.44 days (s.e.=0.046), with a t=32. A patient's own wait time depends on various factors – including patient scheduling preferences and health severity, in addition to congestion levels. Thus, one should not expect mean wait time to perfectly predict patient-specific wait time. Nevertheless, the above analysis demonstrates that W_{-i} is able to pick up mental health crowding that is highly predictive of true patient treatment delays.

[5] These refer to exact two-week period, e.g. the first two weeks of January, 2011.

[6] In Section 2.4.3, for simplicity, I group these into general, psychiatry, PTSD, SUD, or other. My baseline results hold using both the more and less granular versions of the appointment type categories (for both the construction of W_{-i} and controls); see Section 2.4.3 for details.

[7] There are no observations in some hospital-year-month bins due to emergency departments not operating in those locations at particular times.

Is mean wait time as good as random?

Next, I assess independence, or whether W_{-i} appears to be quasi-randomly assigned, conditional on the baseline controls. Importantly, since the main outcome is mortality, high and low risk patients must be equally likely to be assigned a high mean wait time. Following Chan et al. (2022), I construct a predicted mortality for each patient, $\hat{Y}_i$, using detailed patient characteristics and encounter information available in the VA data. Specifically, $\hat{Y}_i$ is formed using a logistic regression with the following covariates: age, gender, race, ethnicity, 3-digit ICD-9 diagnosis codes associated with the ED visit, high risk for suicide flag,[8] priority status,[9] whether the patient arrived by ambulance, prior Elixhauser comorbidities,[10] mental health utilization in the prior year, including prior inpatient and outpatient visits, no shows, and cancellations, and physical health utilization in the prior year, including prior primary care visits, prior specialty appointments, and prior hospitalizations for physical disease. To assess the quality of my mortality prediction, in Figure 3, I plot coefficients and standard errors when I run a logistic regression of two-year mortality on these characteristics.[11] I report a chi-squared statistic of 72,008 and a psuedo-R^2 of 0.210. As age is controlled for in my baseline analysis, I also report a partial R^2 (after conditioning on age) of 0.0704.[12] Secondly, in Appendix Figure B.1, I show a binscatter plot of observed versus predicted mortality. The model is able to predict a mortality risk of over 40% for the top 5% of observations and a mortality risk of over 20% for the subsequent 5%.

I then run the following regression to assess balance:

$$\hat{Y}_i = \pi_2 W_{-i} + \zeta_{2,t(i)} + \phi_{2,a(i)} + \gamma_{2,h(i)m(i)} + \mathbf{X}_i \delta_2 + \epsilon_{2,i} \tag{2.3}$$

I find no relationship between $\hat{Y}_i$ and Z_i, with $\hat{\pi}_2$ = 0.00014 (0.00011) (t = 1.22). In other words, leave-out wait time is uncorrelated with a patient's risk level, indicating quasi-random assignment.

As the VA flags patients deemed high risk for suicide, a natural alternative balance test is whether this high risk flag is correlated with W_{-i}. Using this flag as the left hand side variable, I again find no relationship with W_{-i}. Additionally, another, broader metric indicative of patient severity is whether a patient has a prior hospitalization. Let I_i be an indicator that a patient was hospitalized in the year prior to the index ED visit (for any cause). With I_i as the dependent variable, I again find that the coefficient on W_{-i} is reverse-signed and insignficant; Appendix Section B.2 and Figure A2 provide further details.

[8]Suicide coordinators at the VA can assign patients this flag if patients are screened as being high risk for suicide. Patients deemed high risk for suicide receive specialized review and care at the VA (Veterans Health Administration 2008).

[9]Priority status is a system the VA uses to determine eligibility for VA healthcare and copay responsibilities, based on both disability status and income.

[10]This a standard way of grouping health comorbidities (see Elixhauser et al. 1998).

[11]One interesting note here is that Black patients actually have a lower mortality risk than white patients in this population. This finding has been replicated in other work on mortality for patients with severe mental illness (Das-Munshi et al. 2017).

[12]Note that one would not expect a very high psuedo-R^2 here, as 2-year mortality is a binary 0/1 variable and mortality predictions are probabilities between 0 and 1.

2.4.2 Mortality effect

Baseline effect

I run the following reduced form regression to determine the relationship between W_{-i} and Y_i, or two-year mortality:

$$Y_i = \pi_3 W_{-i} + \zeta_{3,t(i)} + \phi_{3,a(i)} + \gamma_{3,h(i)m(i)} + \mathbf{X}_i\delta_3 + \epsilon_{3,i} \tag{2.4}$$

I find a strong positive relationship, with $\hat{\pi}_3 = 0.00113$ (0.00041). In other words, a one standard deviation increase in mean mental health wait time increases mortality by 0.11 percentage points, or 1.5% of baseline risk. In Figure 2.2 Panel B, I show binned scatterplots of the balance and reduced form relationships (corresponding to Equations 3 and 4), residualizing by the baseline controls. Not only is the balance coefficient insignficant, but the mortality effect is an order of magnitude larger.

Table 2.1 Panel A displays the baseline mortality results. In column 1, I show mortality effect estimate only using baseline controls. In columns 2 through 4, I show the estimate when I add additional patient controls, including gender, race, mental health diagnosis codes associated with the ED visit, and prior mental health utilization. Finally, column 5 shows the estimate when I control for characteristics of other patients who form a given patient's value of W_{-i}, including (leave-out) mean age, mean prior visits, and mean predicted mortality. The rationale behind these latter controls is as follows: leave-out mean wait time should be uncorrelated with patient characteristics, which could be violated if similar types of patients arrive at the hospital at similar times. The baseline results are highly stable across a wide range of controls; I discuss additional robustness checks in more detail in Section 2.4.4.

Additionally, in Panel B, I show heterogeneity results by mental health condition, using mental health ICD-9 diagnosis codes attached to the index ED visit.[13] The effects are by and large driven by patients with SUD or psychosis, such as schizophrenia – the two groups that also have the highest baseline mortality risk.

Finally, in Figure 2.4, I plot the mortality effect over the five years following the ED visit. Given that a course of therapy often lasts several months, intuitively, the coefficient at 30 days is essentially 0 (coeff. = 0.00013 (0.00010)). However, the effect accrues over time, reaching 0.00141 percentage points (s.e.=0.00055) five years after the initial ED visit.

2.4.3 Endogeneity concerns

Is follow up appointment type endogenous?

As described in Section 2.3.2, W_{-i} conditions on what type of follow up appointment the patient is assigned. Visit types include general mental health, which can be a visit with a social worker or psychologist, psychiatry, and specialized SUD and PTSD clinics. This conditioning is necessary to capture true queue lengths, which can vary substantially by appointment type: the mean wait time for a psychiatry appointment is 24.5 days, while the mean wait time for a SUD specialty appointment is 12.0 days. However, one potential concern with this specification is that appointment

[13] Visits at the ED are always assigned ICD-9 diagnosis codes. These codes represent the reason for the ED visit, rather than necessarily being a new diagnosis for the patient.

type is endogenous in cases where a patient may be appropriate for various types of care.

Since I control for appointment type in the baseline specification, my analysis accounts for the fact that certain types of patients select into certain appointment types, on average. However, the correlation between appointment type and patient severity could still present an issue for the analysis if ED staff adjust referral decisions based on queue length–in other words, if selection into a given appointment type varies with residualized W_{-i}. In this case, quasi-random variation in wait times could also induce variation in the types of patients assigned appointment type a, leading to a correlation between W_{-i} and a patient's health status.

More concretely, if wait times are particularly high for appointment type a in a given hospital-time period, are ED staff more or less likely to assign high risk patients to that appointment type? It is not clear which direction the bias may go: on the one hand, patients with higher severity may be most appropriate for appointment types with high wait times (notably, psychiatry). If this match effect drives the selection of appointment type, high risk patients may be more likely to assigned a longer wait time. On the other hand, if ED staff know that a patient has particularly urgent needs, they may decide to send the patient to a provider with shorter wait times, regardless of that provider's specialty.

To directly test whether there is time-varying selection into appointment type that also varies with W_{-i}, I run the following specification:

$$\mathbf{1}(a_i = k) = \beta_1 \tilde{W}_{k,-i} + \beta_2 \hat{Y}_i + \beta_3 \tilde{W}_{k,-i} \times \hat{Y}_i + \epsilon_i \tag{2.5}$$

where $\mathbf{1}(a_i = k)$ is an indicator for patient i choosing appointment type k, $\tilde{W}_{k,-i}$ is the residualized mean wait time for appointment type k (for all patients arriving in the same hospital-two-week period as patient i),[14] and $\hat{Y}_i$ is predicted mortality. The key coefficient is β_3, which answers the question: if wait times for a given appointment type are (randomly) higher, do high risk patients select into or out of appointment type k?

In Appendix Table B.2, I show results for β_1 and β_3 across appointment type categories. Coefficients on β_1 are consistently negative, implying that when the queue is particularly high for a specific appointment type, patients are less likely to be assigned that appointment category. This may either be due to physicians responding to wait times by rushing patients to providers with higher availability, or it may be mechanical: if a patient is referred to multiple providers, they will mechanically be assigned the appointment type he attends first. More importantly, β_3 is either negative or statistically insignificant for all appointment types. These negative coefficients suggest some "reverse-selection": if the queue is especially long for a general mental health visit, high risk patients are assigned to other appointment types. Since I find that high wait times increase mortality, this reverse-sorting suggests that if anything, my baseline results are an underestimate of the true effect of treatment delays.[15]

Nevertheless, to ensure the reader that selection into appointment type is not an important driver of my results, in Table 2.2 Panel A, I show results for two alternative versions of the baseline specification, both of which are plausibly more exogenous. First – in the "saturated" specification – I control for hospital × appointment type and appointment type × two-week period, accounting

[14] I residualize $W_{k,-i}$ by all baseline controls, excluding appointment type.

[15] For simplicity, Table B.2 shows results separated into five major appointment categories: general, psychiatry, PTSD, SUD, and other. These results also hold qualitatively with the more granular appointment type categories used in the main specification, as described in Section 2.4.1.

for more granular selection into appointment type within hospital and time period. Second, as a truly exogenous measure, I create a new, predicted measure of congestion – $\hat{W}_{-i}$ – which does not rely on assigned appointment type at all. Instead, I predict appointment type for these patients using a multinomial logit model. I run predictions hospital-by-hospital, as different hospitals are likely to have different assignment mechanisms to appointment types. Within each hospital, I use the following characteristics to predict appointment type: an indicator for the patient having a prior mental health visit, indicators for most common prior mental health visit type (e.g., if the patient saw the psychiatrist twice in the last year and no other mental health specialist, he would receive a psychiatrist indicator), and indicators for ICD-9 diagnosis code. Using this model, I extract $\mathbf{p}_i$, a vector of the probabilities of being assigned each appointment type.

Then, I use a given patient's vector of predicted appointment types to construct a weighted average predicted leave-out wait time, $\hat{W}_{-i}$, as:

$$\hat{W}_{-i} = \sum_{a} p_{a,i} \bar{W}_{a,h(i)t(i)} \tag{2.6}$$

where $p_{a,i}$ is the probability that patient i is assigned appointment type a and $\bar{W}_{a,h(i)t(i)}$ is the (leave-out) mean wait time for patients arriving at the same hospital and in the same two-week period as patient i who are assigned appointment type a.

Both the aforementioned specifications remove some of the predictive power of the congestion measure. Thus, to improve power in these specifications, in lieu of hospital × year-month controls, I use slightly less granular hospital × year × quarter controls. Additionally, to lessen the computational intensity of the prediction, I group appointment types into major categories – general, psychiatry, SUD, PTSD, and other. In Table 2.2 column 1, I show the baseline mortality result. In column 2, I show the baseline result when I use the slightly less granular hospital × time and appointment type categories; the coefficient is nearly identical.[16] Finally, in columns 3 and 4, I show the saturated and predicted specifications, respectively. Although these specifications are slightly noisier, they yield coefficients highly aligned with the baseline results.

Do patients have discretion over which VA hospital they visit?

Another potential concern with the main regression specification is that patients may use discretion in which hospital they visit. Therefore, conditioning on hospital in the construction of W_{-i} may be an issue if hospital choice is also endogenous to crowding. However, in constructing W_{-i}, hospitals are only divided into broad geographic regions (such as the Cleveland VA), known as VA stations. These stations are usually comprised of a main medical center with an ED, as well as several affiliated outpatient clinics. Patients are highly unlikely to travel to an entirely different geographic region for medical appointments, especially for an emergency visit. Thus, one should think of assigned hospital as the general geographic catchment area for the patient (e.g., Los Angeles, Chicago, etc.), for which there is an associated ED and outpatient clinics; W_{-i} captures the mean mental health clinic congestion in that region. Secondly, I construct W_{-i} based on ED location. Therefore, if initial ED facility choice is unrelated to congestion, even if patients adjust which outpatient clinic they attend in response to wait times, W_{-i} will still be orthogonal to this future optimization.

[16] I also reconstruct W_{-i} using these less granular appointment type categories in this specification.

2.4.4 Additional robustness

In Table 2.2 Panel B, I consider whether sample selection choices are driving my results. In constructing the sample (described in Section II.A), I exclude patients with no scheduled follow up visit, as W_{-i} depends on type of appointment scheduled. To eliminate concern that dropping these patients has implications for my findings, in column 2, I add these patients back to the sample, increasing the sample size to 935,843. This larger sample has a somewhat higher mean two-year mortality of 9.6%. I provide more details on sample characteristics in comparison to the baseline sample in Appendix Section B.1.2 and Table B.1.

Since patients with no scheduled follow up appointment do not have an assigned appointment type, I let $W_{-i} = \hat{W}_{-i}$ for these patients, as described in Section 2.4.3.[17] One can think of $\hat{W}_{-i}$ as a weighted average mean wait time, weighted by the probability of the patient being assigned a particular appointment type. As appointment type is also a baseline control in my main analysis, in this specification, I control for either appointment type directly (for the baseline sample) or the predicted probability of being assigned each appointment type (for patients with no appointment type). In Table B.4, I also show additional robustness checks using this larger sample; results are very similar to those of the baseline analysis sample.

Moving to column 3, about 34% of the sample is immediately admitted as an mental health inpatient following the ED visit. These patients are likely to see a mental health specialist during their inpatient stay (which is considered to be their follow up appointment in the main sample specification). However, due to inconsistencies in the definition of follow up between these patients and others, I modify the definition of a follow up appointment (in both the definition of W_{-i} and regressions) to only include true outpatient encounters. In this specification, I essentially consider one's inpatient stay to be a "long" ED visit. Following this logic, I define a patient's first outpatient mental health visit following discharge as their follow up appointment.[18]

Relevant to inpatient stays, a small literature touches on the implications of the decline in inpatient psychiatric beds overtime (Gibbons et al. 2017, McBain et al. 2022). In hospitals across the US, including at the VA, mental health care has primarily shifted from inpatient to outpatient care, and the consequent decline in beds has been of some policy concern (Raphelson 2017). Relatedly, in this analysis, higher psychiatric capacity could potentially shift patients from outpatient to inpatient care. However, I find that excluding inpatient mental health stays has little effect on the reduced form coefficient, implying that this mechanism is not a primary driver of my results.

In column 4, I remove patients with non-VA insurance, which I describe in more detail in Section 2.4.5. Finally, in Appendix Figure B.3 and Table B.5 I perform various additional robustness checks of interest. These include adding additional control variables and considering alternative forms of clustering, among others. I find that the mortality effect is highly robust.

[17] W_{-i} for observations also in the baseline sample are held fixed in these analyses.

[18] In this specification, wait time is more naturally defined as the difference between the discharge date and follow up appointment date for those with inpatient stays.

2.4.5 Follow up appointment attendance

Impact of treatment delays on follow up appointment attendance

Why do mental health treatment delays have a long term impact on mortality? One explanation is that patients with severe mental illness often have difficulty staying engaged with treatment, which can be exacerbated by treatment delays. In a meta-analysis, Nosï¿œ et al. (2003) find that an estimated 24% of patients with psychosis do not attend mental health visits as scheduled. In the VA setting, Fischer et al. (2008) show that in a cohort of veterans with schizophrenia or bipolar disorder, almost one quarter of the cohort drop out of care for over a year during a 5-year follow up window. Similarly, attrition from mental health care is a major concern among those in my sample. While all patients in my sample have a *scheduled* follow up visit, the no show or cancellation rate for these appointments is about 37%.

In a population who has difficulty staying engaged with care, small system-based nudges have been associated with improved attendance. Specifically, Klinkenberg and Calsyn (1996) find that minimizing wait time to the first appointment following ED discharge is associated with higher engagement with therapy – the exact setting of my analysis. This pattern appears to hold in my sample as well: in Figure 2.5 Panel A, I show the CDF of actual wait times, separating patients who do and do not attend their follow up appointment. Those who attend their appointments tend to have shorter wait times on average.

To formally evaluate whether appointment delays have a causal effect on appointment attendance, I assess whether mean wait time is correlated with the probability patients attend their follow up mental health visit. I run the following regression:

$$\mathbf{1}(V_i = 0) = \pi_4 W_{-i} + \zeta_{4,t(i)} + \phi_{4,a(i)} + \gamma_{4,h(i)m(i)} + \mathbf{X}_i\delta_4 + \epsilon_{4,i} \tag{2.7}$$

where V_i is an indicator for attending a follow up mental health appointment. I find a large and significant relationship, with $\hat{\pi}_1 = 0.0203$ (0.0008), suggesting that a one standard deviation appointment delay (11.7 days) increases the probability of missing one's follow up visit by 2.0 percentage points or 5.6% (t=24). Figure 2.5 Panel B shows a binned scatterplot of the probability of attending a follow up mental health visit versus mean wait time, or W_{-i}, residualizing by the baseline controls.

Overall, my analysis suggests that mental health treatment delays have a major impact on follow up appointment attendance, which I consider to be the primary mechanism driving the mortality effect. Thus, one way to think about these results is to consider treatment delays to be an instrument for attending mental health treatment. Conceptually, patients who arrive during busy periods may have longer delays to obtaining a follow up appointment, pushing those on the margin of returning for a follow up visit ("compliers") to disengage from mental health care. In turn, disengagement from mental health services impacts long term mortality. Specifically, in an instrumental variables framework, π_4 from Eq. 6 acts as the first stage, and π_3 from Eq. 4 – or the effect of mean wait time on mortality – acts as the reduced form. Then, using the jacknifed-instrumental variables estimator, the causal effect of not attending a mental health follow up appointment on two-year mortality is given by $\hat{\beta}_{IV} = \hat{\pi}_3/\hat{\pi}_4$. With the baseline controls, I find that $\hat{\beta}_{IV} = 0.056$ (0.020); in other words, not attending a follow up mental health visit increases 2-year mortality by 5.6 percentage points.

In this setting, both providers and patients may adjust on other margins in response to treatment delays – in other words, there may be violations to the exclusion restriction that make using mean wait time as an instrument imperfect. Therefore, I consider this IV analysis to primarily function as a scaling exercise, giving the reader a sense of how detrimental congestion may be for compliers who disengage from mental health care. In Section 2.4.7, I further evaluate how missing one's first follow up visit affects intermediate outcomes, particularly engagement with long run therapy. First, however, in Section 2.4.6, I consider alternative explanations for the mortality effect.

Could patients be attending mental health services outside the VA?

One important note in interpreting the prior results is that my data only captures patient encounters at VA hospitals. This fact leads to the natural question: could patients be attending mental health services outside the VA? If so, this complicates the counterfactual: patients who miss their appointment at the VA could still be seeing a therapist at another hospital.

Since healthcare in the US is incredibly expensive, it would be unusual and likely infeasible for veterans in my sample to get mental health treatment at another, private hospital without outside insurance. Overall, about 30.3% of the baseline sample has some kind of outside insurance, 55.4% of which is Medicare coverage (Appendix Table B.1). Medicare is public insurance provided by the federal government for all those about 65, as well as individuals who qualify due to disability status (including mental illness).

To assess whether the 30.3% of patients with outside insurance could be driving result, I drop these patients from the baseline sample to create a no-insurance subsample. Described in Appendix Table B.1, this sample is demographically similar to the main sample with a few differences: 1) the mean age is slightly lower, and 2) two-year mean mortality is lower. These latter two facts are consistent with these patients being non-Medicare recipients, so they are likely to be younger and suffer from less severe mental illness. I find that W_{-i} is also highly predictive of missing a follow up visit for this subsample, with a similarly-sized coefficient of 0.022 (0.001). Additionally, in Table 2.2 Panel B and Appendix Table B.4, I rerun all primary specifications for this sample. The baseline mortality effect is 0.075 p.p. (s.e.=0.044) or 1.4% of the mean for this sample– highly consistent with the baseline effect. On the whole, these results suggest that substitution to the private sector is not an important channel in my analysis.

2.4.6 Alternative explanations

Are mental health wait times correlated with general crowding?

An alternative explanation for the mortality effect is that congestion is not specific to mental health visits; in other words, mental health appointment delays may also predict delays throughout the hospital or clinic, indicating poorer overall quality. Notably, Feyman et al. (2022) find significant correlation in waiting times across specialties at the VA. Thus, one hypothesis is that those with high W_{-i} receive poorer overall medical care, and this is driving the mortality effect. Note that while certain VA hospitals are likely always more crowded (for example, VAs in dense metropolitan areas) and that certain time periods may also be associated with higher crowding across specialties, my baseline controls guarantee that all congestion comparisons are done *within* hospital-year-month. In other words, I control for the fact that some hospitals are more crowded overall at certain times, and instead isolate congestion not explained by these controls, which I argue reflect some randomness.

Beyond controlling for these baseline differences in crowding across hospital-time, I conduct two additional tests to assess whether there could still be correlation between W_{-i} and general delays in care across other specialties. First, I directly plot wait times for other common types of ED follow up visits versus residualized W_{-i}, residualizing by the baseline controls. This approach allows me to determine whether the "quasi-random" component of mental health clinic congestion is correlated with other clinic wait times. Second, as a more downstream test, I construct a placebo sample of patients who arrive to the ED with non mental health diagnoses and assess whether W_{-i} predicts mortality for this group. If W_{-i} can neither predict wait times directly nor mortality of non mental-health patients, these tests suggest W_{-i} has an impact on mortality only through mental health treatment itself, rather than through care delivered at the hospital in other specialties.

First, I document the five most common specialties visited following an ED visit: primary care, labs or pathology, podiatry, opthamology, and internal medicine.[19] Then, I calculate mean wait times (as the time between the ED visit date and the follow up visit date) for each of these specialties within hospital-two-week bins.[20] Next, I assign these wait times to W_{-i} by matching on hospital-two-week period. Finally, in Figure 2.6 Panel A, I plot a binscatter of these other wait times versus residualized W_{-i}. I find that all five non-mental health speciality wait times are uncorrelated with residualized W_{-i}, confirming the fact that my measure of mental health clinic congestion does not predict general crowding across specialties.

As a second test, I construct a new analytical sample of 1,377,834 patients who visit the ED during the same time period as the primary sample, but are diagnosed with a circulatory condition at their ED visit, such as stroke or heart attack.[21] I assign these patients a value of W_{-i} according to their time and hospital of arrival.[22] This alternative sample acts as a placebo test, as the mortality of stroke and heart attack patients should not be affected by mental health clinic congestion. However, if mental health clinic congestion indicates poorer overall hospital quality as a result of crowding, the leave-out wait time measure should also predict higher mortality for this group of patients.

In Figure 2.6 Panel B, I show a binned scatterplot of W_{-i} versus two-year mortality, residualizing by two week period, year-month $\times$ hospital fixed effects, and age buckets. I find no significant relationship between W_{-i} and mortality for this placebo sample of non-mental health ED patients. Overall, I conclude that W_{-i} is only a measure of mental health clinic congestion and is not correlated with general delays in care.

[19]ED visits are identified by an ED visit code plus an ICD-9 diagnosis code, while outpatient follow up visits may only be identified by a specialty visit code (since a patient only receives a diagnosis if they attend their appointment). Therefore, it is not straightforward to link all ED visits to follow up visits using a combination of specialty codes and diagnosis codes. Instead, I use a simpler definition of follow up here: the outpatient appointment must be scheduled on the same day as the ED visit.

[20]This level of variation is equivalent to the variation of W_{-i}, except that W_{-i} also varies by mental health clinic type.

[21]Patients in this sample must have a diagnosed mental health condition during the study period in order to be a part of the initial sample construction (in the VA data request). However, these patients arrive to the ED with a circulatory condition, rather than a mental health condition.

[22]As patients in this alternative sample do not have a follow up mental health visit, I group together all appointment types to assign them a mean wait time.

Do longer wait times affect the quality of mental health care?

While the prior tests suggest that W_{-i} is not associated with non-mental health quality of care, what about quality of care within the mental health clinic itself? For example, if appointments are more delayed due to high congestion, it is possible that physicians are either more rushed or stack more appointments closer together in time. Thus, crowding could lead to poorer outcomes independent of appointment attendance.

To assess this hypothesis, I precisely calculate the length of follow up mental health visits using treating provider identifiers. Specifically, I document the psychologist or psychiatrist assigned to a given patient for their follow up appointment and the precise start time of their visit (up to the minute). Then, I follow the same physician to their next appointment with a new patient and again document this visit's start time. I define the difference between these two start times as the follow up appointment length. Using this method, I censor any appointment lengths that are greater than six hours, as I assume this captures the doctor ending and starting a new shift. With censoring, I capture 299,958 appointment lengths (of 394,292 patients who attended their follow up appointment).

The median appointment length is exactly 60 minutes, while the 5th and 95th percentile lengths are 7 minutes and 223 minutes, respectively. This method for calculating appointment lengths is likely noisy: a provider may see multiple patients in the same time window or could have a shift break of only a few hours. Therefore, I assume any extreme outliers are data errors and either 1) focus on the median appointment length within W_{-i} bins or 2) winsorize observations at the 5th and 95th percentiles. In Appendix Figure B.4, I show a binscatter of either the winsorized mean or median follow up appointment length versus W_{-i}, residualizing by the baseline controls. I find that appointment length is uncorrelated with W_{-i} in both cases, mitigating some concern that congestion also leads to rushed care.

Finally, one concern with the above analysis is that I can only capture appointment lengths for those who attend their follow up visit. As demonstrated in Section 2.4.5, visit attendance is non-random and is highly correlated with W_{-i}. Thus, one story is that those who did not attend their appointment due to high congestion also would have had a shorter appointment, and I cannot observe this fact in the prior regressions. To address this non-random attrition problem, I use a Lee bounds approach. Since those with high W_{-i} are essentially censored here, Lee (2009) recommends censoring either the left tail or right tail appointment lengths of patients assigned low W_{-i} to construct worst-case scenario bounds.[23] Using median appointment lengths and worst-case bounds, I can still reject appointment lengths shorter than two minutes for those with mean wait times longer by one standard deviation.

Although I do not find evidence of more rushed care, it is possible that crowding changes delivery of mental health treatment in more subtle ways. For instance, perhaps crowding leads to treatment by lower-skill providers or worsens match quality between provider and patient if certain providers fill appointment slots more quickly. I do find some evidence of this phenomenon: as a second test of the quality of mental health care, I match treating providers (for follow up visits) to number of years of experience that provider has at the VA. Using provider experience as the left-hand side variable, I find that a one stand. dev. increase in wait time decreases provider expe-

[23] Since Lee conceptualizes the problem using a binary control and treatment group, for simplicity, I convert W_{-i} to an indicator for above and below median to censor observations.

rience by about 0.16 years (s.e.=0.023), or about 1.5% of the mean.[24] Does provider experience contribute to the mortality effect? Although not a causal analysis, in an OLS regression, I find that two-year mortality is uncorrelated with provider experience, conditional on the baseline controls (coeff.=0.000059; s.e.=0.000044). Nevertheless, I cannot rule out that other components of mental health treatment are changing when clinics are more congested. In fact, in the next section, I find evidence that providers marginally adjust treatment decisions in response to crowding. Thus, while I argue that appointment delays primarily affect long term health outcomes via *patient* behavior – by increasing the likelihood of missing their follow up visits – supply-side adjustments may also play a role in contributing to the mortality effect.

Do ED doctors anticipate delays with compensatory behaviors?

What if higher W_{-i} is actually correlated with better mental health care delivery? This could be the case if there is any compensatory behavior; in other words, physicians are aware that mental health treatment delays could lead to poor patient outcomes, and therefore, they are compensating for these delays by modifying other care practices. In this case, lower appointment attendance among those with high wait times may be mitigated by improved care delivery.

One potential way to compensate for longer appointment delays is to prescribe mental health medications at the ED visit itself, rather than waiting for a follow up consultation. Overall, 30.2% of patients are prescribed a mental health drug on the day of the ED visit, 54.5% of which are antidepressants. Interestingly, a one standard deviation increase in mean wait time increases the probability of being prescribed a drug by 1.2 percentage points (s.e.=0.08) or 4.0% (Table 2.3). Specifically, lithium, antidepressants, analgesics, and anticonvulsants all appear to be marginal drugs, with physicians being more likely to prescribe them if one's follow up appointment is delayed.

About 64% of patients were prescribed a mental health drug in the year prior to their index ED visit. Thus, another interpretation of the correlation between mean wait time and increased precriptions are that patients come to the ED to refill or update their prescriptions, and may be especially likely to do so if it difficult to get an outpatient mental health appointment. Notably, among the 30% of patients prescribed a drug at the ED visit, about 61% of prescriptions are for repeat drug classes (e.g., the patient had been prescribed an antidepressant in the prior year and are again prescribed one at the ED). Thus, to shed light on this interpretation, in column 2, I show that around half of the effect comes from prescriptions in entirely new drug classes, which is inconsistent with a patient requesting a refill. Secondly, I drop patients who were prescribed a mental health drug in the prior year from the main reduced form analysis; the resulting mortality effect is 0.00130 (0.00063), highly consistent with the baseline results.

Given the compensatory behavior of physicians, an additional concern is that wait time and prescriptions are jointly determined; for example, patients who are prescribed a drug at the ED may be able to wait longer for a follow up appointment if they are already receiving treatment through their medication. Since I am not using a patient's wait time directly, W_{-i} is not influenced by whether patient i himself is actually prescribed a mental health drug. However, other patients' wait times, which are used to form a given patient's W_{-i}, may be affected by prescribing patterns.

[24] Similar to the analysis on appointment durations, I cannot link visits that were no showed or cancelled to a specific provider. Using a conservative Lee bounds approach to correct for attrition, I cannot rule out that treating provider experience is lower or higher for those with high wait times (lower bound: -0.52; upper bound: 0.42).

This in turn could bias results if the rate of prescriptions differs when the ED is crowded. To address this, I drop patients that are prescribed a drug at the ED visit from the construction of W_{-i}. The resulting mortality effect is similarly 0.00130 (0.00039).

These results provide some indication that physicians are of aware of and compensating for appointment delays with more immediate treatment at the ED. Drugs prescribed at the ED could potentially have long run effects: the average days' supply for a given medication is 26 days, with an average of one refill allowed. Beyond refills, even if patients do not return to mental health care, primary care providers may re-prescribe medications to these patients. Since medications should mitigate mortality risk, prescribing patterns at the ED likely partially counteract the impact of missing future behavioral therapy appointments. As evidence of this point, returning to the IV specification in Section 2.4.5, I directly control for whether a mental health drug was prescribed at the ED visit itself. As expected, this increases the coefficient to 0.063 (0.020), making the effects of therapy attendance look even larger.

2.4.7 Intermediate outcomes

Engagement with mental health care

In Section 2.4.5, I showed evidence that mental health treatment delays make patients more likely to miss their follow up appointment. What services are patients actually missing as a result of appointment non-attendance, and how does congestion then affect long term engagement with mental health care? First and foremost, in my sample, 61% of follow up appointments are for general behavioral therapy, 15% are for addiction-related treatment, 7% are psychiatry appointments, 6% are for specialized PTSD treatment, and 11% are other specialty appointment types (such as mental health integrated care). During the two year follow up window, patients engage in a range of treatments, from medication to group therapy to long-term residential stays.

In Table 2.4 Panel A, I first evaluate how congestion affects long run engagement with behavioral therapy. Not only are patients with high mean wait time far more likely to miss their first follow up visit (column 1), as described in Section 2.4.5, but they are far more likely to disengage *entirely* from mental health care in the two years following the index ED visit. In column 2, I show that a one standard deviation increase in mean wait time following the ED visit makes patients 0.67 percentage points (s.e.=0.04) less likely to attend *any* outpatient mental health appointment up to two years following the ED visit. Moreover, in column 3, I show that higher mean wait time decreases the likelihood patients attend any inpatient form of mental health health treatment, including long term residential stays. One interpretation of this latter result is that outpatient therapy attendance may be an important mechanism in connecting patients with more intensive forms of treatment.

Another potential explanation for lack of future mental health engagement is a competing risks model, where patients die before they are able to attend therapy appointments. However, this concern is mitigated by the fact that the majority of the mortality benefit appears only over a long time horizon (see Figure 2.4). On the whole, my results imply that treatment delays following an ED visit have effects that go far beyond a single appointment – instead, they have a long run impact on future engagement with mental health care.

What about medications? The vast majority of patients – around 91% – take some kind of mental health drug during the two-year follow up window. Medications for substance use specifically are somewhat less common, as some disorders do not have FDA-approved medications – notably stimulant use disorder. However, for opioid and alcohol use, medications are recommended as first-line treatments. Among my substance use subsample, around 54% of patients are prescribed a drug related to alcohol or drug use specifically, such as suboxone or naltrexone. Prescriptions may not go hand in hand with therapy attendance – only physicians, not psychologist or social workers, can prescribe medications. While psychiatrists have traditionally prescribed mental health drugs, in the current mental health landscape, primary care providers are increasingly writing these prescriptions (Faghri et al. 2010).

In Appendix Figure B.5, I show results on the effect of mean wait time on use of mental health drugs over time. As described in Section 2.4.6, those with high mean wait time are about 1.2 percentage points more likely to be prescribed a drug at the ED visit itself, which I argue is a compensatory response to treatment delays. While this effect lessens over time as more patients are prescribed drugs at follow up visits, those with high mean wait time are actually still marginally more likely to be using mental health drugs in the long term (including both prescriptions and refills).[25]

With these results, I can rule out that those with high wait times are less likely to take mental health drugs overall. Although patients with high W_{-i} are more likely to drop out of specialty mental health care, they are also more likely to attend a primary care appointment during the follow up period (see evidence in Section 2.4.7). Thus, these results suggest that those with high wait times are more likely to be prescribed drugs by non-mental health specialists (or primary care providers). I cannot rule out that medications play no role in the mortality effect: one potential story is that primary care providers are less skilled at prescribing appropriate drugs to patients with severe behavioral health needs. This explanation is in line with evidence from Currie and MacLeod (2020), who show wide variation in provider skill in prescribing antidepressants. I comment more extensively on the role of primary care doctors in mental health care in Section III.G.iv.

Cause of death

A natural question is whether the mortality effect is being driven by mental health casues of death, such as suicide and overdose, or whether effects are driven by physical disease, which can be exarcerbated by poor mental health (APA 2014). Using standard ICD-10 code classifications for suicide from the National Death Index, I find that only 4.8% of deaths in my sample are directly attributable to suicide (Cornes et al. 2021). If I include drug overdose and suspected suicide events, this percentage increases to 9.6%. There is evidence that suicide deaths may be highly undercounted in both the veteran population and in general, likely due to both stigma and financial incentives related to life insurance policies (Ellis 2022, Snowdon and Choi 2020).

However, to get at the distinction between physical diseases and suicide or overdose-related causes of death, in Appendix Table B.6, I divide deaths into those caused by physical disease versus those caused by accident, suicide, or overdose. Within physical disease, I further divide diseases into major categories, such as heart disease, respiratory diseases, and cancer. Let M_i be an indicator for mortality due to a specific cause. To determine cause-specific death rates among those with high wait time, I run regressions with M_i as the left-hand side variable as follows:

[25]The slight increase in medication use holds true for both all mental health drugs and substance use drugs specifically.

$$M_i = \pi_5 W_{-i} + \zeta_{5,t(i)} + \phi_{5,a(i)} + \gamma_{5,h(i)m(i)} + \mathbf{X}_i \delta_5 + \epsilon_{5,i} \tag{2.8}$$

Then, $\bar{M}_i + \hat{\pi}_5$, where $\bar{M}_i$ is the mean death rate, yields the cause-specific death rate for those with a mean wait time one stand. dev. above the mean ("high wait time patients"). Overall, I find that 6.5% of the sample died from a physical health condition, while only 1.0% died from other causes. This decomposition is similar among those with high wait times, with over 87% of deaths being due to physical causes. Specifically, the primary cause of death among those with high mean wait time is heart disease; these patients have a 1.69% chance of dying from heart disease, about a 3.4% increase from the sample mean of 1.63%. Moreover, high mean wait time patients have increased risk of dying from respiratory diseases, liver disease, and infections.

While perhaps surprising, these results corroborates previously established relationships in the medical and psychology literature. Those with poorer mental health are likely to also suffer from poorer physical health, and severe mental illness can worsen the course of chronic disease (APA 2014). Notably, the fact that most preventable deaths among the severely mentally ill are due to physical rather than mental conditions is consistent with existing evidence (Kisely et al. 2013, Liu et al. 2017). The medical literature has specifically documented the association between psychiatric conditions and excess mortality due to heart disease, respiratory diseases, and infections (Lawrence et al 2003, Newcomer and Hennekens 2007, Liu et al. 2017).

Moreover, one might expect that death from physical disease is particularly likely for those suffering from substance use disorder. Prior research has shown that addiction treatment is associated with decreased incidence and progression of alcohol-associated liver disease (Vannier et al. 2022). In addition, illicit opioid use is associated with a range of physical diseases, including chronic obstructive pulmonary disease, HIV, endocarditis, and viral hepatitis, which can also lead to liver disease (Lewer et al. 2021, Nishimura et al. 2019).

In a similar vein to this study, Kisley et al. (2013) find that patients with orders to attend psychiatric treatment had an all-cause mortality risk nearly half that of the control group at two years, with the strongest effects being for death from physical diseases. Analagously, I find that higher clinic congestion leads to disengagement from mental health care, also resulting in higher death rates from physical disease.

Engagement with non-mental health medical care

Increased mortality from physical causes could have various underlying mechanisms: first, those who do not seek mental health care may be more likely to become detached from the healthcare system at large, and lack of engagement with physical health may be driving the mortality result. For simplicity, call this the "physical health engagement" channel. Alternatively, improved mental health alone could result in improved physical health, without being driven specifically by use of other types of medical care. For simplicity call this the "mental health engagement" channel. Finally, an intermediary between these two mechanisms could be contributing to the mortality effect: those with untreated mental illness could engage differently, and perhaps less effectively, with physical health care, even if they continue attending appointments. For example, those with untreated mental illness may have more difficulty adhering to lifestyle changes associated with improved cardiovascular and respiratory health, such as diet, exercise routines, and smoking cessation. Alternatively, physical health providers may have more difficulty diagnosing and treating physical health conditions in patients with untreated mental illness (Liu et al. 2017). Call this the "physical-mental intersection" channel.

As a point of reference, my sample is actually comprised of a highly VA-attached population: 97% of the sample returns to a VA hospital for at least some type of appointment within the two-year follow up period, and 84% return for a primary care appointment. Therefore, very few individuals lose contact with the VA system following the index ED visit.

To directly test the physical health care engagement channel, in Table 4 Panel B, I show that those with high mean wait time are actually *more* likely to attend a primary care appointment during the follow up period. Why is this the case? There are two potential explanations: first, patients may be substituting between mental and physical health care. If it is hard to get a speedy mental health appointment, they may show up to the primary care doctor instead. In fact, the APA suggests that high usage of primary care due to mental distress may be common, with up to 70% of primary care visits being driven by behavioral health (APA 2014). As primary care doctors treat both physical and behavioral health conditions, those who attend primary care more often may be receiving additional physical or mental health care, though delivered by a generalist. I explore delivery of mental health care by primary care doctors in the subsequent section.

Secondly, since my results suggest that high mean wait time patients also exhibit deteriorating physical health, these patients may simply be in more need of medical care. To provide additional evidence that these patients have declining physical health, in column 2, I show suggestive, though underpowered evidence that those with high mean wait time are more likely to be admitted as an inpatient for a physical health condition. Moreover, in column 3, I show that those with a one stand. dev. higher mean wait time are 0.19 p.p. (s.e.=0.08) or 0.3% more likely to be diagnosed with a new physical Elixhauser comorbidity. As these comorbidities are long term, chronic conditions that can worsen over time, it is likely that existing comorbidities, in addition to new diagnoses, contribute to the mortality effect.

In Appendix Table B.7, I also break down engagement with mental and physical health care for SUD patients, as they an important driver of death in my sample (see Table 2.1). For these patients, results on hospitalizations for physical health and new physical comorbidities are even stronger. These results are aligned with the medical evidence surrounding cause of death for SUD patients, where untreated addiction is highly associated with physical health decline.

Overall, I demonstrate that although patients stay attached to the medical system as a whole, they exhibit declining physical health. Thus, I can rule out the "physical health engagement" channel, in which mental health care helps patients stay connected to physical health providers. However, both the "mental health engagement" and "physical-mental intersection" channels may play a role in contributing to the mortality effect.

The role of experts in healthcare

Those with high mean wait times, despite being far more likely to disengage from specialized mental health treatment, are likely to both continue taking mental health medications and continue attending primary care. So why are these patients so much worse off? One way to think about the results is that patients disengage from *specialty* mental health care. However, they may continue attending primary care for mental health conditions, and non-specialist providers may continue prescribing them mental health drugs.

To further elucidate delivery of mental health care by primary care providers, I define any primary care visit with an associated mental health ICD-9 diagnosis code as a primary care appointment for mental health.[26] In my sample, about 65.6% of patients attend at least one primary care visit for mental health. Of these patients, about 3.6% attend *only* primary care for their mental health needs; in other words, they drop out of mental health specialty care. Those with high wait times are far more likely to fully substitute to primary care: a one stand. dev. increase in mean wait time increases the probability of receiving mental health treatment only via primary care by 0.27 percentage points (s.e.=0.03), or about 11.4%.

How often do patients receive any type of mental health treatment, whether via specialists or not? To get at this question, I redefine "returning" to mental health care from Table 2.4 as attending 1) a mental health specialty clinic appointment or 2) attending a primary care appointment with a mental health diagnosis code. Using this broader definition of mental health care, I find that those with high wait times are still less likely to receive any type of mental health care during the two year follow up window. A one stand. dev. increase in mean wait time decreases the probability of returning to any type of mental health care by 0.39 percentage points (s.e.= 0.03), about half of the effect size as that of specialty care alone (shown in Table 2.4). These results imply that patients at least partially substitute to primary care to treat their mental health conditions.

Lastly, how much does disengagement from specialty mental health care, as opposed to all forms of mental health care, contribute to the mortality effect? To shed light on this question, I exploit variation in care practices across hospitals. Notably, some hospitals may utilize primary care providers for mental health treatment more often than others. Moreover, there may be systematic variation in patient engagement across hospitals, even conditional on wait times. For instance, different administrations may have better systems to remind patients to attend appointments, and transportation may be more accessible in certain VA locations. To categorize hospitals, I first run the following set of regressions hospital by hospital:

$$D_{sp,i} = \alpha_{sp,h(i)} W_{-i} + \zeta_{1,h(i)t(i)} + \phi_{1,h(i)a(i)} + \mathbf{X}_i \delta_{1,h(i)} + \eta_{1,i} \tag{2.9}$$

$$D_{all,i} = \alpha_{all,h(i)} W_{-i} + \zeta_{2,h(i)t(i)} + \phi_{2,h(i)a(i)} + \mathbf{X}_i \delta_{2,h(i)} + \eta_{2,i} \tag{2.10}$$

where $D_{sp,i}$ is an indicator for disengaging from specialty mental health care (attending 0 visits) in the two years following the ED visit, and $D_{all,i}$ is an indicator for disengaging from all mental health treatment (or both primary and specialty care). For each hospital h, I collect estimated coefficients $\hat{\alpha}_{sp,h(i)}$ and $\hat{\alpha}_{all,h(i)}$. I then categorize values of $\hat{\alpha}_{sp,h(i)}$ and $\hat{\alpha}_{all,h(i)}$ above the 75th percentile as "high disengagement" hospitals. Conceptually, a high value of $\alpha_{h(i)}$ means that at that hospital, congestion makes patients particularly likely to disengage from mental health care. In particular, high $\alpha_{sp,h(i)}$ indicates high disengagement from specialty mental health care, while high $\alpha_{all,h(i)}$ indicates high disengagement from all forms of mental health care.

Finally, to determine what type of mental health care matters for the mortality effect, I run the following regression:

$$Y_i = \pi_H W_{-i} \times H_{h(i)} + \pi_L W_{-i} \times L_{h(i)} + \zeta_{t(i)} + \phi_{a(i)} + \gamma_{h(i)m(i)} + \mathbf{X}_i \delta + \epsilon_i \tag{2.11}$$

where $H_{h(i)}$ and $L_{h(i)}$ are indicators for high and low disengagement hospitals, respectively. I run

[26] This is a relatively broad definition of mental health care, and includes any visit where mental health is addressed. This could include prescribing a mental health drug, checking in on an existing prescription, or psychotherapy.

this regression twice, first forming $H_{h(i)}$ and $L_{h(i)}$ using values of $\hat{\alpha}_{sp,h(i)}$, and then using values of $\hat{\alpha}_{all,h(i)}$. π_H and π_L represent the mortality effect for high and low disengagement hospitals, respectively. The intuition behind including these interactions is as follows: if disengagement from all forms of mental health care is an important driver of the mortality effect, π_H should be high when $H_{h(i)}$ is formed using $\hat{\alpha}_{all,h(i)}$. Alternatively, if disengagement from specialty mental health care specifically is important, π_H should be high when $H_{h(i)}$ is formed using $\hat{\alpha}_{sp,h(i)}$.

I show regression results from Eq. 10 in Table 2.5. I find that hospitals with high disengagement from specialty care have very large mortality effects. In fact, $\hat{\pi}_H$ is about 10 times the magnitude of $\hat{\pi}_L$ when hospitals are categorized based on specialty care alone (column 1). Alternatively, I find little to no heterogeneity when hospitals are categorized based on all forms of mental health care, including primary care (column 2). On the whole, these results suggest that disengagement from specialty mental health care specifically is a key driver of the mortality effect.

Overall, my analysis provides suggestive evidence that primary care may not be sufficient to meet patients' mental health needs. In other words, mental health expertise is useful. These results may be particularly policy relevant in the current US mental health landscape: there is a growing shortage of mental health specialists, and primary care providers are increasingly responsible for treating patients' psychiatric conditions (Faghri et al. 2010).

2.5 Discussion

This paper is the first to evaluate the causal impact of mental health treatment delays on long term, all-cause mortality. For patients with emergent mental health conditions, I find that delaying follow up appointments by just one standard deviation – about 11.7 days – increases 2-year mortality by 1.5%. Moreover, in analyzing the downstream effects of high wait times, I identify a key barrier to healthcare system engagement: I find that those whose appointments were initially delayed are not only more likely to miss their first follow up visit, but they are far more likely to never engage with mental health services at the VA again.

On the whole, my results suggest that reducing clinic wait times could have important downstream effects on patients' wellbeing. One way to illustrate this point is as follows: if mean wait times across VA clinics fell from 17.7 days (as in the baseline sample) to about 10 days, or about a 0.66 standard deviation reduction, the reduced form results suggest that mortality would decrease by 0.073 percentage points, or about 1% of mean mortality. This effect size may represent an upper bound relative to other patient settings: given the urgent nature of their conditions, these veterans have two-year mortality rates higher than other veterans with mental illness. For instance, Bowersox et al. (2012) finds that VA patients with psychosis have a two-year mortality rate of 4.6%; the mortality rate among psychosis patients in my sample is 11.3%. Moreover, compared to non-veterans, these patients have unique provider accessibility, with veterans often paying no copay for treatment.

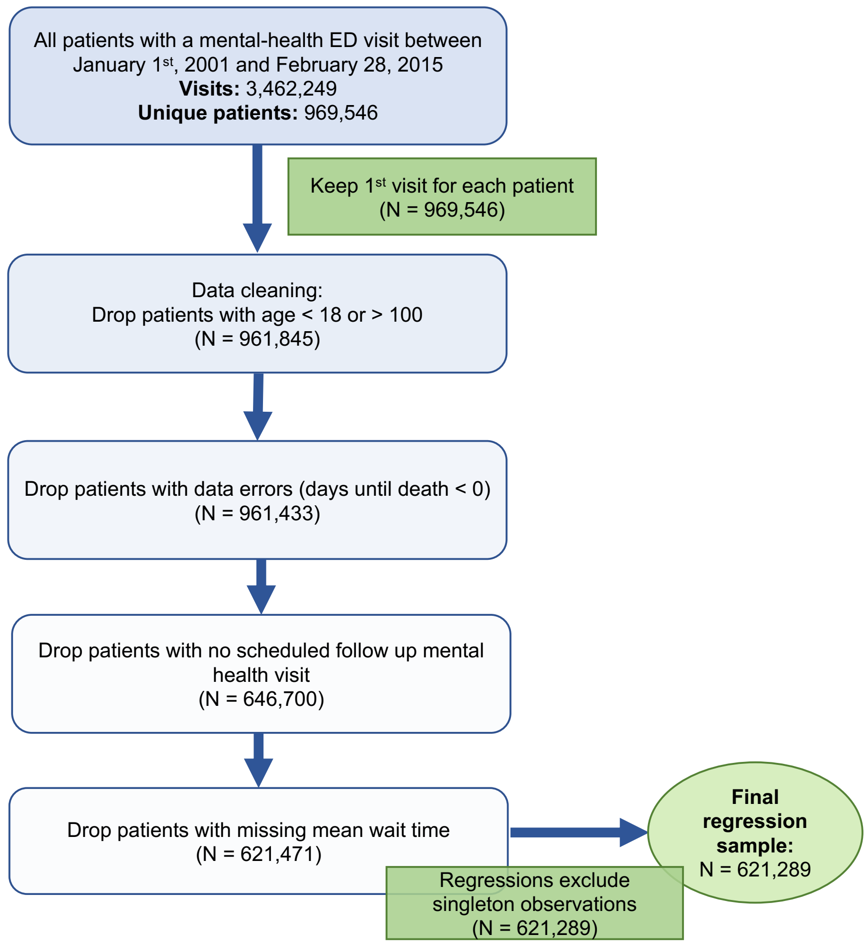
Appointment cancellations and no shows are an important issue not only in this setting, where the combined non-attendance rate was 37%, but among low income and vulnerable populations in general (Sharp and Hamilton 2001). Remarkedly, this study demonstrates that modifying scheduling practices may not only have behavioral implications for a single appointment, but may affect how patients interact with the healthcare system long term. In areas such as mental health, where going lost to follow up from care is a major concern, determining practices that can nudge patients to attend appointments may be key. Preliminary research has shown that massed visits, or scheduling appointments closer together in time (such as triweekly instead of weekly), increases treatment completion rates (Yamokoski 2020). Additionally, incentives to attend therapy sessions, such as vouchers or monetary rewards, are already in use at some VA facilities. These types of interventions have been shown to be highly effective in improving treatment adherence, at least while incentives continue; a randomized-control trial from 2020 finds that financial rewards of just $2-7 per week increase adherence to antidepressents by 21% (Marcus et al. 2020). In comparison, about 63.5% of patients in my sample attend their follow up appointment, and a decrease in wait time of about 12 days increases this probability by 2.0 percentage points – about a 3% increase.

Finally, my results underscore the importance of expertise in medicine. I find that those with long wait times – who are far more likely to disengage from specialty mental health care – are actually more likely to have primary care appointments during the two year follow up period, suggesting that primary care cannot substitute for mental health care. With patients increasingly turning to primary care to treat their mental health conditions, these results further highlight the need for more mental health specialists.

While my analysis provides suggestive evidence of the downstream effects of provider shortages, further research is needed to confirm the mechanisms driving my results. First and foremost, this study did not randomize delivery of mental health care by specialists versus generalists, which would provide more concrete evidence on the effects of mental health experts. Specifically, why do specialists perform better – is it due to differential prescribing patterns, more effective therapy, or some combination of the above? Lastly, this study is limited to mental health patients; future research could investigate whether waiting times have similar effects for patients suffering from physical health conditions.

With mental health treatment making headlines from recent gun legislation to policies surrounding veteran suicide, improving access to and engagement with behavioral therapy is of notable policy relevance. Major new initiatives aim to make mental health care more accessible, both at the VA and at large: in 2022, the VA proposed to eliminate outpatient copayments for veterans flagged high risk for suicide. More broadly, in summer 2022, the US launched 988, an emergency number for those experiencing mental health emergencies. My results suggest that timely care, in addition to more affordable care, may be a key part of the puzzle.

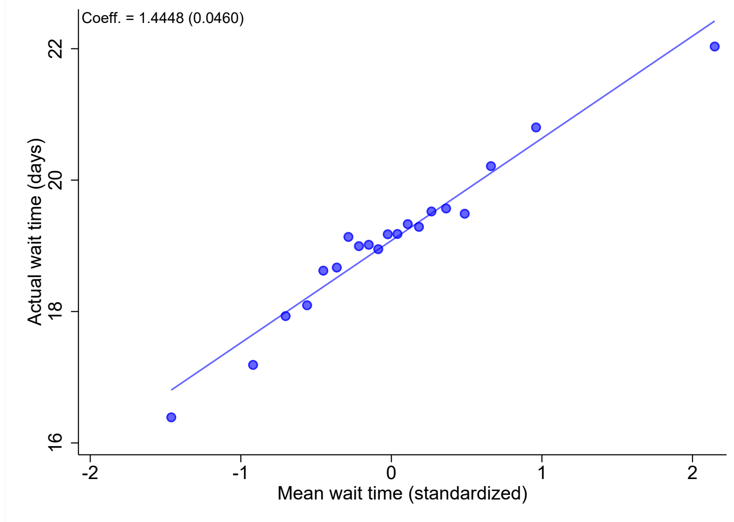
Figure 2.1: Sample selection



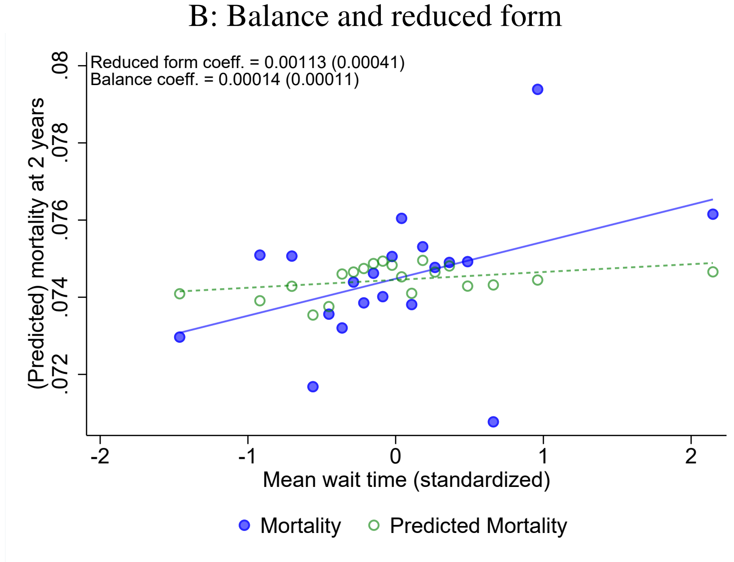
Note: This figure shows the sample selection process using encounter data and patient information from the VA. Patients have a missing mean wait time value if no other patients are in the same hospital-two-week period-appointment type group as the index patient.

Figure 2.2: Relevance, balance, and reduced form

A: Correlation between mean wait time and actual wait time

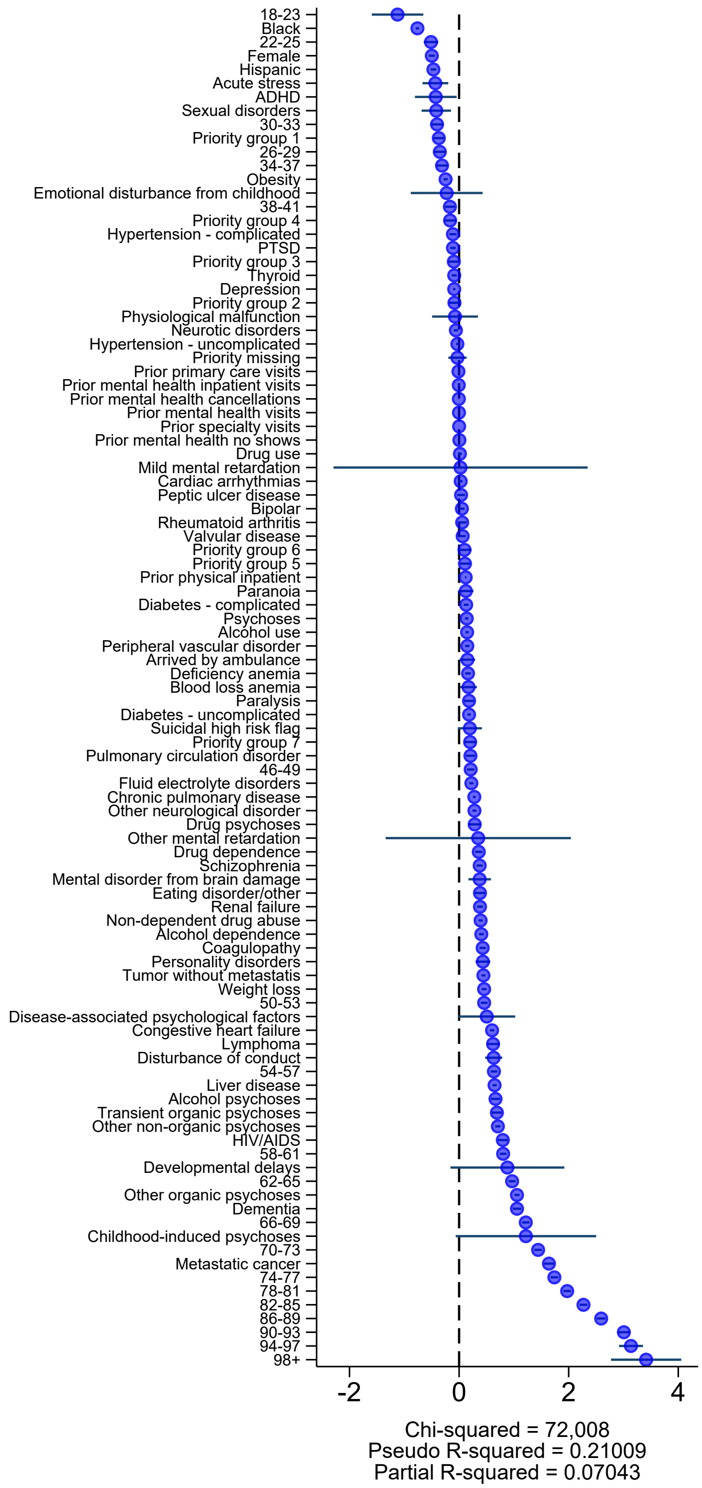


B: Balance and reduced form



Note: Panel A shows a binned scatterplot of actual wait time (in days) versus mean wait time (standardized), or W_{-i}, residualizing by the baseline controls. This represents relationship described in Equation 2. Panel B shows a binned scatterplot of two-year mortality or predicted mortality versus W_{-i}, residualizing by the baseline controls. These represent the balance and reduced form relationships described in Equations 3 and 4. Predicted mortality is plotted using green hollow circles, and mortality is plotted using solid blue circles. Baseline controls include two-week period, year-month $\times$ hospital, appointment type, and age buckets.

Figure 2.3: Mortality prediction



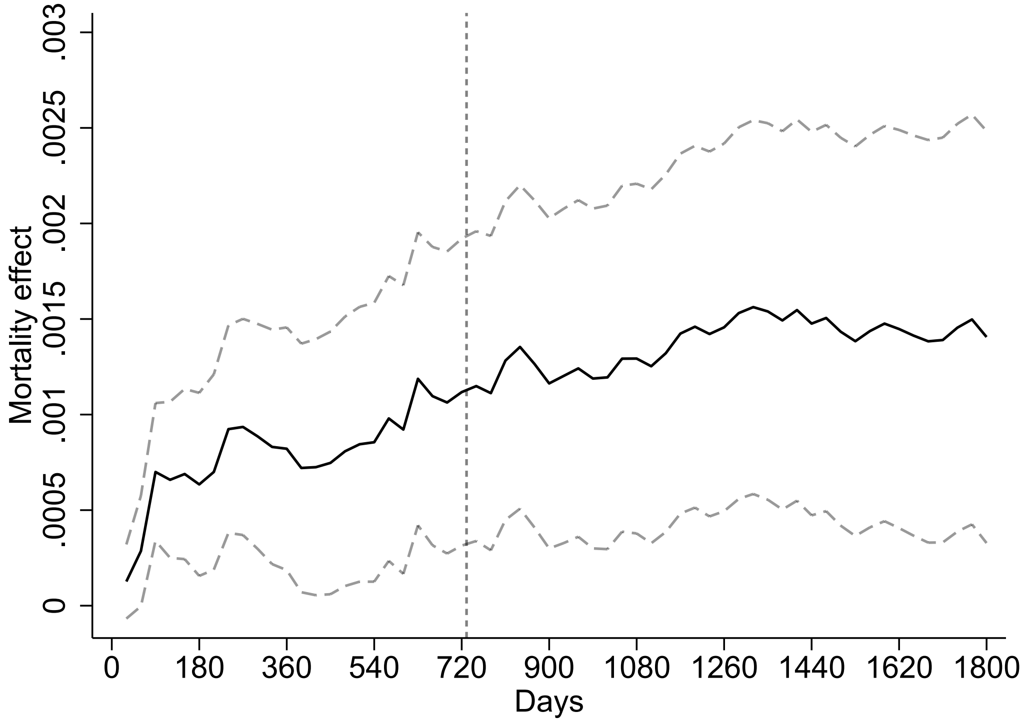
Note: This figure plots regression coefficients and standard errors for a logistic regression of two-year mortality on patient characteristics used to form $\hat{Y}_i$. Physical diagnoses are based on conditions patients were treated for at the VA in the prior year and are divided into Elixhauser comorbidity groups. The excluded group are white males aged 42 to 45 with depression in Priority group 8. The partial R-squared reports the marginal contribution of all other variables after removing the effects of age.

Table 2.1: Effect of mental health wait times on mortality

	A: Overall Sample				
	2-year mortality				
	(1)	(2)	(3)	(4)	(5)
Mean wait time	0.00113	0.00104	0.00107	0.00113	0.00093
	(0.00041)	(0.00041)	(0.00041)	(0.00041)	(0.00041)
Outcome mean	0.0746	0.0746	0.0746	0.0746	0.0746
Observations	621,289	621,289	621,289	621,289	621,289
Baseline controls	Yes	Yes	Yes	Yes	Yes
Demographics	No	Yes	Yes	Yes	Yes
Prior utilization	No	No	Yes	Yes	Yes
Diagnosis codes	No	No	No	Yes	Yes
Leave out controls	No	No	No	No	Yes
	B: Major Diagnosis Category				
	2-year mortality				
	Substance use	Psychosis	Mood disorders	Anxiety	PTSD
Mean wait time	0.00186	0.00360	0.00048	0.00030	0.00048
	(0.00081)	(0.00208)	(0.00074)	(0.00114)	(0.00107)
Outcome mean	0.0707	0.1130	0.0527	0.0527	0.0360
Observations	175,721	52,600	184,812	71,895	60,178

Note: In Panel A, I show the baseline mortality results. Column 1 shows the main estimate, only using baseline controls. Column 2 shows the estimate when I add demographic controls, including gender and race. Column 3 shows the estimate when I also add controls for prior mental health utilization, and column 4 adds 3-digit ICD-9 diagnosis codes (from the ED visit). Finally, column 5 shows the estimate when I control for characteristics of other patients who form a given patient's value of W_{-i}, including (leave-out) mean age, mean prior visits, and mean predicted mortality. In Panel B, I show mortality results in diagnosis group subsamples (based on ICD-9 codes from the ED visit). Baseline controls include month-year $\times$ hospital, appointment type, and age buckets. Standard errors are clustered by hospital-month.

Figure 2.4: Mortality effect over time



Note: This figure plots the mortality effect over time. Solid lines show point estimates, and dashed lines show 95% confidence intervals. The vertical dashed line indicates the main mortality estimate at 2 years.

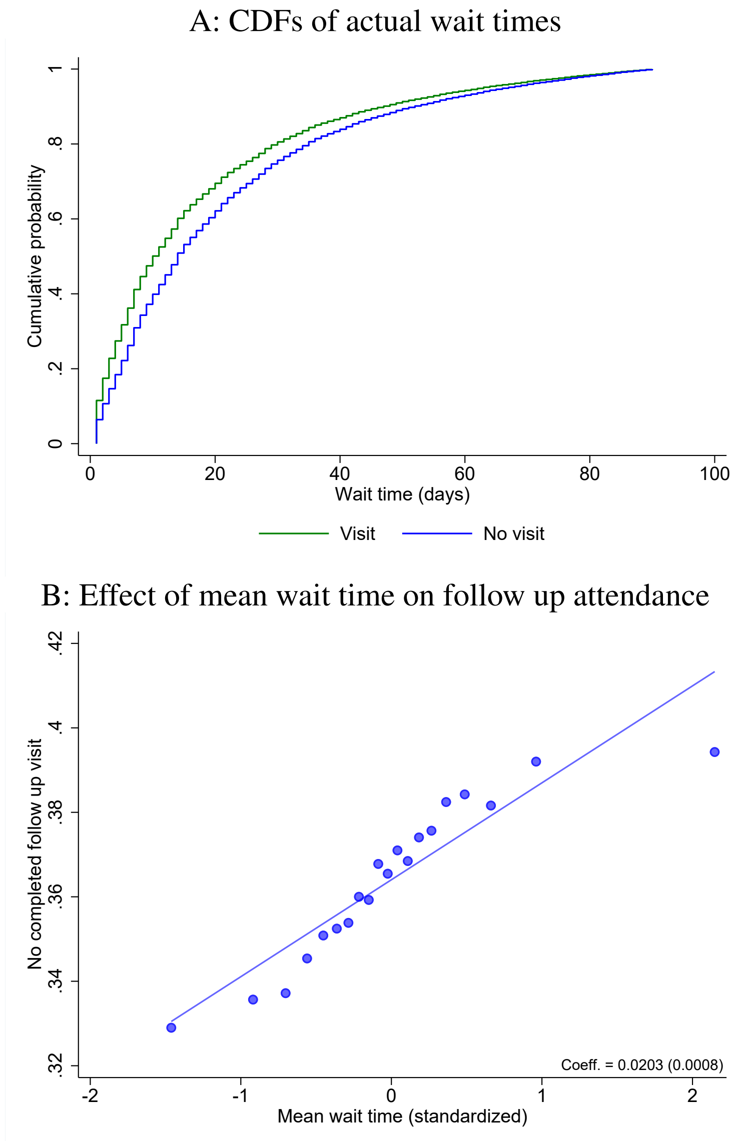
Table 2.2: Robustness checks

	A: Accounting for selection into appointment type			
	(1)	(2)	(3)	(4)
	Baseline	Less granular controls	Saturated	Predicted wait time
Mean wait time	0.00113	0.00124	0.00090	0.00092
	(0.00041)	(0.00041)	(0.00043)	(0.00047)
Outcome mean	0.0746	0.0746	0.0746	0.0746
Observations	621,289	621,289	621,289	621,289
Appointment types	11	5	5	5
Hospital × time controls	Year × month	Year × quarter	Year × quarter	Year × quarter

	B: Alternative samples			
	(1)	(2)	(3)	(4)
	Baseline	+ Patients with no scheduled follow up	Exclude inpatient visits	Only VA insurance
Mean wait time	0.00113	0.00118	0.00129	0.00075
	(0.00041)	(0.00035)	(0.00041)	(0.00044)
Outcome mean	0.0746	0.0956	0.0746	0.0528
Observations	621,289	935,843	621,289	432,863

Note: Panel A shows regression results that account for appointment type endogeneity. Column 1 shows the baseline estimate from Table 2.1, column 1. Column 2 shows the estimate with less granular controls: hospital × year × quarter controls (in lieu of hospital × year × month controls) and less granular appointment type categories (general, psychiatry, PTSD, SUD, and other). In this specification, I also re-construct W_{-i} using these larger appointment type categories. Still using the less granular fixed effects, column 3 adds controls for hospital × appointment type and hospital × 2-week period. Finally, column 4 predicts appointment type instead of using assigned appointment type. Panel B shows the mortality effect using alternative versions of the sample. Column 2 adds back patients with no scheduled follow up visit, column 3 excludes inpatient mental health encounters from the universe of mental health follow up visits, and column 4 excludes patients with outside insurance.

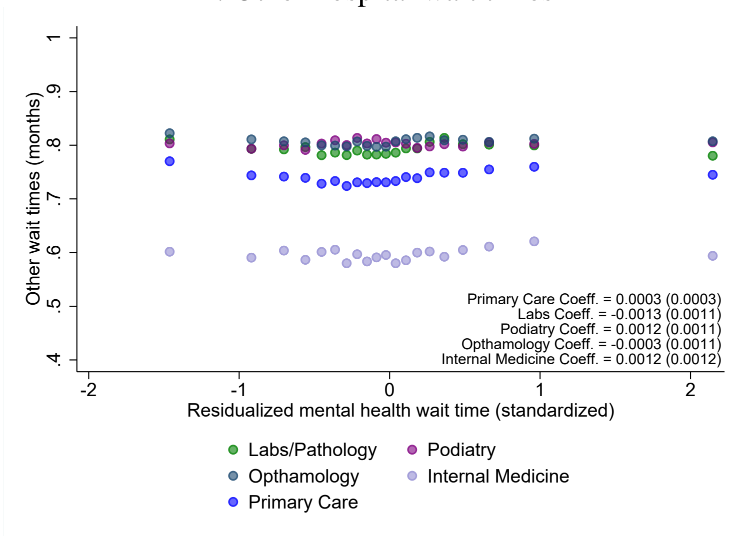
Figure 2.5: Treatment delays and follow up appointment attendance



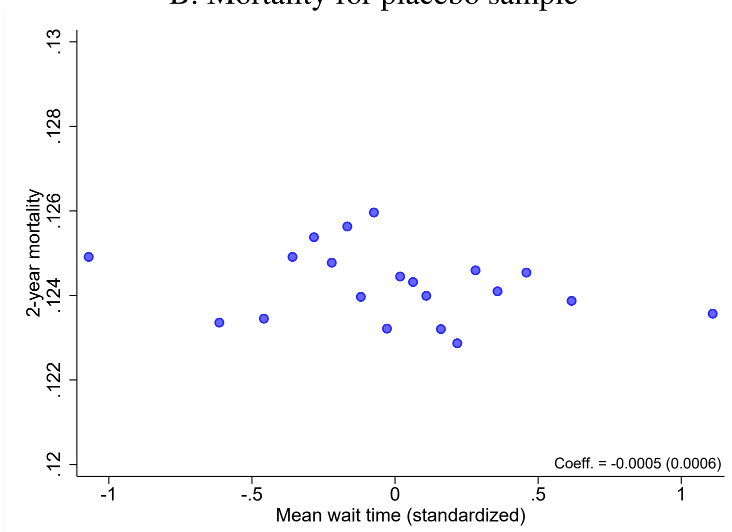
Note: Panel A plots the CDF of actual wait time (in days) for patients who did (green) and did not attend (blue) their follow up appointment. Panel B shows a binned scatterplot of the probability of not attending a follow up mental health visit versus leave-out wait time, or W_{-i}, residualizing by the baseline controls. This represents the relationship described in Equation 6.

Figure 2.6: Correlation between mean wait time and quality of non-mental health care

A: Other hospital wait times



B: Mortality for placebo sample



Note: Panel A shows a binned scatterplot of various other specialty wait times versus residualized W_{-i}. All wait times are calculated as the mean number of months between a patient's ED visit and their follow up appointment day for a given specialty. Panel B shows a binned scatterplot of two-year mortality versus W_{-i} for a placebo sample of patients with circulatory conditions, residualizing by the baseline controls. This is analagous with the reduced form relationship for the main sample described in Equation 4. Baseline controls for this sample include two-week period, year-month $\times$ hospital, and age buckets.

Table 2.3: Effect of mean wait time on prescriptions at the ED

	(1) Prescription (overall)	(2) New prescriptions	(3) Anticonvulsants	(4) Antidepressants
Mean wait time	0.0120	0.0058	0.0014	0.0088
	(0.0008)	(0.0005)	(0.0003)	(0.0006)
Percent increase	4.0%	5.0%	4.6%	5.3%
Outcome mean	0.3016	0.1162	0.0306	0.1643
Observations	621,289	621,289	621,289	621,289
	(5) Analgesics	(6) Sedatives or hypnotics	(7) Anti-psychotics	(8) Lithium
Mean wait time	0.0022	0.0022	0.0015	0.0003
	(0.0004)	(0.0005)	(0.0003)	(0.0001)
Percent increase	4.6%	2.1%	2.8%	7.4%
Outcome mean	0.0476	0.1033	0.0516	0.0042
Observations	621,289	621,289	621,289	621,289

Note: This table shows the effect of mean wait time on the likeliness of being prescribed a mental health drug at the ED visit. Baseline controls include two-week period, month-year × hospital, appointment type, and age buckets. Standard errors are clustered by hospital-month.

Table 2.4: Effect of mean wait time on intermediate outcomes

	A: Mental health outcomes		
	(1)	(2)	(3)
	No follow up visit	Returned to mental health care	Mental health inpatient treatment
Mean wait time	0.0203	-0.0067	-0.0099
	(0.0008)	(0.0004)	(0.0008)
Outcome mean	0.3652	0.9333	0.3153
Observations	621,289	621,289	621,289
	B: Physical health outcomes		
	(1)	(2)	(3)
	Primary care visit	Inpatient visit - non-mental health	New physical comorbidity
Mean wait time	0.0032	0.0012	0.0019
	(0.0006)	(0.0007)	(0.0008)
Outcome mean	0.8404	0.2114	0.6208
Observations	621,289	621,289	621,289

Note: This table presents reduced form results for various intermediate outcomes, following patients for 730 days after the ED visit. Panel A shows results for engagement with mental health care. Returned to mental health care refers to attending any oupatient appointment. Mental health inpatient treatment refers to any mental health inpatient visit, including residential stays (excluding admissions within one day of the initial ED visit). Panel B presents results for non-mental health outcomes. The first column looks at whether a patient attended a primary care appointment, the second column looks at non-mental health inpatient admissions, and the third column looks at whether the patient had any new Elixhauser comorbidity (only including physical conditions) during the two-year follow up period.

Table 2.5: Heterogeneity by level of disengagement from mental health care

	2-year mortality	
	(1) Specialty disengagement	(2) Full disengagement
Mean wait time × high disengagement	0.0031	0.0013
	(0.0007)	(0.0008)
Mean wait time × low disengagement	(0.0003)	0.0011
	(0.0005)	(0.0005)
Outcome mean	0.0746	0.0746
Observations	621,289	621,289

Note: This table shows heterogeneity results for VA hospitals with high and low levels of disengagement from mental health care. "High disengagement" is an indicator for being in the top 25th percentile of hospitals, while "low disengagement" in an indicator for being in the bottom 75th percentile (based on values of $\hat{\alpha}_{h(i)}$ from Eqs. 8 and 9). In column 1, I show heterogeneity results for VA hospitals based on the probability of disengaging from specialty mental health care. In column 2, I show heterogeneity results for VA hospitals based on probabilities of disengaging from all mental health care, including specialty and primary care visits. Disengagement from mental health care is defined as not attending any mental health appointment in the two years following the ED visit. Mental health primary care is defined as a primary care visit assigned a mental health ICD-9 diagnosis code.

Chapter 3

The effect of Red Scare propaganda on US public opinion: Evidence from the newspapers of William Randolph Hearst

3.1 Introduction

Does newspaper slant simply reflect what readers want to hear? If newspapers are profit driven, this should be the case: since consumers have a preference for reading like-minded news, successful newspapers should aim to match political views of readers, rather than views of owners. Leading research corroborates this view: Gentzkow and Shapiro (2011) find that readers explain far more of the variation in slant across US newspapers than owners, and Gentzkow, Shapiro, and Sinkinson (2011) find that explicitly partisan newspapers have little to no effect on voting outcomes. But what if owners are not only ideologically driven, but possess extreme beliefs? As the largest case study in American history, William Randolph Hearst's media empire may be the prime example of the importance of both newspaper ownership and ideology in shaping American political discourse, with ramifications to the present day.

Following the 1917 Bolshevik Revolution in the Soviet Union, the subsequent witchhunt against anyone with suspected communist sympathies in the US marked a period known as the "Red Scare." In historical memory, the face of the Red Scare was Senator Joseph McCarthy, who crusaded to rid the US government of "communist infiltrators" in the 1950s. However, twenty years prior to McCarthy's rise, media mogul William Randolph Hearst may have been the main progenitor of anti-communist discourse. In the 1920s and 30s, Hearst owned the largest media conglomerate in the world, inheriting the *San Francisco Examiner* from his father and buying an additional twenty-eight newspapers across the US. A staunch anti-communist who often promoted his own political views with sensationalized headlines, Hearst reached the height of his influence in the mid-1930s, with his newspapers obtaining a circulation of 20 million. As journalist Jim Tuck states, "If Senator Joe McCarthy was the best-known individual representative of the post–World War II heresy hunt, its leading media representative was the Hearst publishing empire" (Tuck 1995, preface).

American's distaste for socialism, ostensibly due to its link with autocratic regimes historically, has propagated in political debates from the Red Scare era to the present day. Policy proposals advocating for increased public funding for social institutions are often met with strong pushback – from Medicare in the 1960s to the Affordable Care Act in the early 2000s. The US remains an outlier among wealthy nations with regards to both inequality and policies towards redistribution. While many Western European countries have both universal healthcare and education, public opinion surrounding similar policies in the US continue to be divisive. No prior research, however, has sought to relate public debate surrounding redistribution with Red Scare politics.

This paper seeks to connect Red Scare propaganda campaigns, specifically anti-communist rhetoric espoused by William Hearst, with both short and long-run political attitudes towards redistribution. My main results examine the effect of Hearst's reporting on political support for Franklin Delano Roosevelt (FDR) across presidential elections. In the 1930s, FDR launched the New Deal, the widest-scale federal spending program in US history. The New Deal aimed to facilitate recovery from the Great Depression via financial reforms and regulation, as well via the introduction of social programs to help the unemployed, farmers, children, and the elderly. Notably, FDR created the Social Security Administration in 1935, the large scale social program that provides retirement benefits to the elderly. With the rise of the Red Scare, FDR was increasingly labeled as a socialist by right-wing critics.

As Hearst was known for promoting his own staunch anti-communist views, the political slant of his papers was plausibly exogenous to the populations they served. Moreover, Hearst began his career as a Democrat and only split from the party in the mid-1930s, after he had already acquired all of his newspapers. Consequently, it is unlikely that the locations where he chose to purchase papers had any correlation with prior anti-communist or anti-redistributive political sentiment. Notably, Hearst actually supported FDR in the 1932 presidential election, only becoming vehemently anti-communist and anti-FDR in the mid-1930s. Therefore, I consider FDR vote share in the 1932 election to be prior to treatment, allowing me to examine the *change* in vote share across elections.

Exploiting Hearst's sharp change in political attitudes in the mid-1930s in a difference-in-difference design, I find that an increase of 10 percentage points in Hearst circulation per capita decreased support for FDR by 2.14 percentage points (s.e.=0.77), a 3.6% decrease. Since Hearst's papers were primarily located in large cities, across various specifications, I exclude areas without local papers and rural areas from the control group, as well as control for detailed county characteristics interacted with year. I show that results are robust to controlling for non-Hearst newspaper readership, as well as presence of other politically-affiliated newspapers in a county. Moreover, I show that the results hold under various alternative specifications. This includes using a propensity-score matched control group, matching on detailed county characteristics, as well as employing the approach proposed by Callaway Sant'Anna (2021), which avoids potential issues with negative weights in standard difference-in-difference analyses (see Borusyak et al. 2024, Chaisemartin and D'Haultfoeuille 2020).

Beyond historical election results, I also examine the persistence of anti-socialist views through the end of the 20th century using data on public opinion from the American National Election Survey (ANES). Tracking Hearst newspaper entry and exit, I compare respondents in the *same county* who were directly exposed to Hearst's reporting as an adult, those who were children but grew up in a Hearst county, and those who were unexposed to Hearst. Aggregating survey responses between 1970 and 1994, I find that those directly exposed to Hearst as adults continue to harbor more anti-redistributive beliefs: they are more opposed to a public healthcare system and feel more negatively about welfare programs. These results are robust to controlling for a wide array of respondent characteristics, including flexible age and cohort controls, as well as limiting the sample to those old enough to be exposed to Hearst. However, children who grew up in a Hearst county do not share these beliefs, suggesting that anti-redistributive attitudes were not passed down from parents to children.

This study contributes to several strands of literature. First, this paper is not the first to study the effect of Hearst's reporting; Listokin and Snyder (2010) also study the effects of the Hearst media empire on voting outcomes. However, they study an earlier period, 1880-1938, the majority of which Hearst identified as a Democrat. Exploiting the entry of a Hearst newspaper into a given county, the authors find a clear null effect on overall voting behavior, results that are consistent with Hearst's fraught relationship with the Democratic party throughout this period. Though Hearst nominally remained a Democrat until the mid-1930s, he increasingly found himself at odds with the Democratic establishment and eventually sharply diverged to the right (Nasaw 2000). In contrast, I study the effects of Hearst media empire from the 1930s onwards, when Hearst became increasingly radicalized, right-wing, and consistent in his opposition to the Democratic party.

Second, this study contributes to an emerging literature on the effect of media persuasion on public opinion (e.g., Adena et al. 2015, Dellavigna and Kaplan 2007, Enikolopov et al. 2011). Prior studies have shown that media personalities espousing strong political views can influence voting behavior (Wang 2021), and that even subtle changes in language used in newspaper reporting has political ramifications (Djourelova 2021).

Moreover, this study weighs in on how ownership may shape media slant. Gentzkow and Shapiro (2011) conceive of newspapers as a market in which both preferences of readers and owners may influence political bias in newspaper reporting. Using a model of newspaper demand, they argue that slant is actually primarily driven by consumers, who have a preference for like-minded news. Hearst's newspapers actually offer a counterexample to this framework: historians assert that Hearst's papers printed his own, often controversial political views, even leading to a decline in circulation as Hearst veered sharply to the right (Nasaw 2000). Thus, this paper provides some initial evidence that ideology, in addition to profit, may drive media supply.

Finally, this paper contributes to the literature on what drives preferences towards redistribution. Previous literature has considered explanations such as belief in equal opportunity and fairness of social competition (Alesina and Ferrara 2005, Alesina and Angeletos 2005), racial discordance between the rich and poor (Alesina et al. 2001), personal history of hardship, and historical circumstances (Alesina and Giuliano 2011, Bazzi et al. 2020). The paper provides additional evidence that exposure to specific historical events may have long run effects on political attitudes.

3.2 Historical background

3.2.1 The Red Scare, the Depression, and FDR

In 1917, the Bolsheviks overthrew the Russian monarchy, establishing a socialist government in its place. In turn, the success of the Bolsheviks spurred a wave of other communist revolutions throughout Europe and beyond. As international politics surrounding the threat of communist overthrow intensified, American politicians warned the public that there could be a communist plot within the US itself. These fears sparked a domestic witchhunt against anyone with supposed communist or socialist sympathies, including those more broadly on the economic left. This period in US history is famously referred to as the "Red Scare," with red signifying allegiance to the Soviet flag.

While American capitalism boomed during the 1920s, the stock market crash in 1929 marked the start of the Great Depression. Unemployment levels reached a peak of nearly 25% in 1933 (Bureau of Labor Statistics). Franklin Delano Roosevelt, then the governor of New York, won the Democratic party nomination in 1932, pledging to bring America out of the Depression. Bringing small farmers, Southern whites, northern Blacks, Catholics, Jews, labor unions, and intellectuals into his coalition, FDR won the presidential election in 1932 by a landslide, with 57% of the popular vote (Clubb et al. 1987).

In his first days in office, FDR began a series of federal policies and programs that became known as the New Deal, aimed at combatting the Depression. Early initiatives included monetary policy and banking reform. Roosevelt also established organizations to improve public infrastructure and employ the unemployed, including the Public Works Administration and the Civilian Conservation Corps. Other legislation provided relief for farmers and support for home ownership.

Programs enacted 1935 and later were dubbed the "Second New Deal," as they were both more liberal and met with stronger pushback. Notably, with the passage of the Social Security Act in 1935, Roosevelt established the first widespread social safety net program for the elderly, the unemployed, and families with single mothers. The National Labor Relations Act, also passed in 1935, gave workers the right to collective bargaining. Concurrently, FDR increased both individual and corporate tax rates to raise revenue for these projects. These later policies attracted criticism from both the business community and right-wing politicians, with some labeling the New Deal as state-sponsored socialism (Shrecker 2021, Roosevelt 1992, p.51).

3.2.2 Newspapers in the early 20th century

Before both the advent of television and widespread popularity of radio, newspapers were the most important source of information on politics and current events. By 1932, there were nearly 2,000 daily newspapers, encompassing over 1,000 US counties. Combined, these papers had a total circulation of nearly 38 million (Gentzkow et al. 2011). Newspapers were often highly partisan, both declaring explicit party affiliations, as well as endorsing political candidates during elections. Moreover, newspaper distribution was highly local; news hawkers or delivery boys distributed papers within a given city, usually either by selling individual copies on downtown streets or delivering newspapers to subscribers' front porches. As newspaper readership grew, papers were increasingly acquired by large chains – the largest and most famous being the Hearst media empire. Concurrently, newspapers faced rising competition from radio, with around 80% of American households having a radio in their home by 1940 (Treisman 2025).

3.2.3 William Randolph Hearst

Two decades before Senator Joseph McCarthy became the face of America's anti-communist crusade, media mogul William Randolph Hearst captured readers across the US with his so-called "yellow journalism," characterized by sensationalized, exaggerated reporting. Hearst was born in San Francisco in 1863 to millionaire mining engineer George Hearst, who passed down ownership of the *San Francisco Examiner* to his son in 1887. With plans to create a large newspaper chain, by the mid-1920s, Hearst had purchased a string of 28 newspapers across the country, including *Los Angeles Examiner*, theï¿œ*Boston American*, theï¿œ*Atlanta Georgian*, theï¿œ*Chicago Examiner*, and the*ï¿œDetroit Times*. At its peak, readership of the Hearst papers reached 20 million (Nasaw 2000, p. xiv, ch. 24).

Hearst began his political career as a Democrat, serving as a member of the U.S. House of Representatives from 1903 to 1907. He was originally a champion of the working class, continuing to support FDR through the 1932 election. However, as the Red Scare re-entered political discourse, Hearst became increasingly conservative and turned against FDR's later New Deal policies, splitting from the party in the mid-1930s (Nasaw 2000, Schrecker 2021).

Hearst exerted tight control over his papers, printing his own political opinions and dictating the way news should be presented. As historian David Nasaw writes in his biography of Hearst, "The Hearst newspapers had always spoken with one voice, Hearst's" (Nasaw 2000, p.484). He often used bombastic, alarmist rhetoric. In particular, Hearst made connections between the Soviet Union, labor unrest in the US, and New Deal policies, accusing FDR of being a communist. Hearst continued his anticommunist vitriol and attacks against FDR despite declining profits – with his newspaper circulation declining by 8 to 10% in major cities between 1933 and 1937 (Nasaw 2000, ch. 33).

Figure 3.1 displays a *San Francisco Examiner* headline from 1936 which directly implicates FDR with the "Alien Reds." Similarly, Figure C.1 displays other illustrative Hearst newspaper clippings. In addition to calling FDR a communist, Hearst painted the New Deal as both an attack on the American system of government and a recipe for federal waste. His newspapers reported that the New Deal was a result of "flirtation with alien ideas" and that its supporters wanted "to terminate the American system, and subtly put some totalitarian substitute in its place" (*Pittsburgh Sun-Telegraph*, August 23, 1939, *Pittsburgh Sun-Telegraph*, September 22, 1944). On an economic front, Hearst argued that the New Deal would increase debt and taxation while exacerbating poverty, and that FDR's policies had led only to "waste, extravagance, and squandering" (*Pittsburgh Sun-Telegraph*, August 23, 1939, *San Francisco Examiner*, November 2, 1936).

In 1943, Hearst stepped down as president of the Hearst Corporation, his newspaper conglomerate. Nevertheless, his legacy continued to flourish. With rising Cold War tensions, Hearst's denunciations seemed increasingly credible. His successor as head of the Hearst Corporation, Richard Berlin, continued unabashed Hearst's red-baiting practices. Soon, the much younger McCarthy would enter the scene, announcing that he possessed a list of communists in the State Department (Schrecker 2021, Tuck 1995). The Hearst papers not only championed McCarthy's cause, but volunteered to disclose their own files on supposed Communist sympathizers to McCarthy's staff (Nasaw 2000).

3.3 Data

3.3.1 Newspaper data

To determine which newspapers were owned by William Randolph Hearst in the 1930s, I use the Library of Congress US Newspaper directory. Derived via data collection during the United States Newspaper Program (1982-2011), a national effort to catalogue historical papers, the directory contains detailed descriptions of historical papers, including ownership. I search the directory with the keyword "Hearst" and date range 1900-1950. I then check all search results by hand, confirming Hearst ownership with a secondary online source. Additionally, I note dates that a given newspaper was bought, sold, and/or ceased publication. In total, I identify all twenty-eight daily papers owned and operated by William Randolph Hearst in 1932. These newspapers were located in twenty-two unique US counties.[1]

Broader data on daily newspapers comes from Gentzkow, Shapiro, and Sinkinson's (2011) American Newspaper Panel (ICPSR 30261). This dataset contains information on every daily newspaper across US cities starting from 1869. Relevant to this study, data from 1932 and onwards come from the Editor and Publisher Yearbook, a monthly trade magazine covering the newspaper industry. Gentzkow and coauthors collect information on individual papers by county, including circulation, political affiliation, and presidential candidate endorsements. I link my hand-coded dataset on Hearst's papers to this larger newspaper panel by matching on newspaper title and city, with an 100% match rate.

Lastly, to explore newspaper text data directly, I locate Hearst and relevant control newspapers available in either Newspapers.com or NewspaperArchive.com, the two sources with the largest historical coverage (Beach and Hanlon 2022). While coverage of the Hearst papers is incomplete, I locate six Hearst newspapers and eighteen newspapers in propensity-score matched counties (see Section 3.4.2) available within these databases during the 1930s and 40s. To capture Hearst's growing focus on anti-communism directly within news coverage, I scrape the count of total mentions of the word "communist" for each newspaper-year using the Optical Character Recognition (OCR) software within these databases. To approximate coverage during these years, I compare these mentions to mentions of a neutral word, as suggested by Beach and Hanlon (2022).

3.3.2 Public opinion and county characteristics

To analyze public opinion towards redistributive policy, I use both voting data and survey data on public opinion. My primary dependent variable for the baseline results is support for FDR. Following Wang (2021), I use Clubb, Flanigan, and Zingale's data on electoral elections (ICPSR 8611) to determine vote shares for FDR by county in the 1932-1944 elections. Additionally, to assess balance in the pre-period, I collect county-level vote shares for the Democratic party between 1912 and 1928. I match these data to the Newspaper Panel by FIPS county code. As additional control variables, I collect 1930 county demographic information from the ICPSR 2896 dataset (Haines and ICPSR 2010), which collapses decennial census data to the county level.

An alternative way to assess balance would be to look at public opinion surveys in treated and control counties in the years prior to Hearst's switch in political ideology. Unfortunately, the first US public opinion poll, conducted by Gallup, was not administered until 1935. Thus, I rely on both prior vote shares, as well as demographic characteristics, to construct a valid control group.

For longer run data on public opinion, I use the American National Election Studies (ANES) Survey, a nationally representative survey on public opinion first conducted in 1948. Published every two to four years, the survey interviews respondents about attitudes towards elected officials as well as pressing political topics of the day. As individual surveys have only around 1,000

[1] Since Hearst owned multiple newspapers in New York City, I consider all five New York City buroughs to be treated in the baseline specification.

respondents, I focus on questions asked across a wide range of years. The survey contains county information for each respondent, as well as respondent demographics such as race, age, education, and economic status. Additionally, starting in 1968, the survey began asking respondents how long they have lived in a given county, allowing me to track whether respondents were directly exposed to Hearst.

3.4 Sample selection

3.4.1 What predicts Hearst owning a newspaper in a given county?

One relevant concern for the empirical design is why Hearst decided to purchase papers in particular counties. As noted by Gentzkow and Shapiro (2011), newspaper slant is both supply and demand driven; slant could merely reflect what readers want to hear. The idea that slant is demand driven is less concerning in this case study: historical accounts repeatedly emphasize that Hearst's papers reflected his own, often controversial political views, even resulting in a decline in circulation among his working class, pro-labor readers as he became increasingly right wing (Nasaw 2000). Moreover, as Hearst's political views sharply changed over the 1930s, some years after he had acquired all of his papers, the location of Hearst's papers are unlikely to be correlated with prior attitudes towards communism or redistribution.

Nevertheless, even if slant is driven by owners, rather than readers, Hearst counties may look fundamentally different than non-Hearst counties. Hearst purchased papers for both business and political motives, seeking to both create a newspaper empire and increase his prominence in the Democratic party in the early 20th century. Starting with his home city of San Francisco, this primarily brought him to the biggest cities in the country, starting with New York (1895), Chicago (1900), and Los Angeles (1903). In Figure C.2, I more closely examine what characteristics predict Hearst operating in a certain county in 1932 by running the following bivariate regressions:

$$\mathbf{1}[\text{Hearst}_i] = \lambda X_i + \epsilon_i \tag{3.1}$$

where $\mathbf{1}[Hearst_i]$ is an indicator for Hearst operating in county i in 1932 and X_i is a standardized county characteristic. I find that Hearst counties have the largest population sizes, have the highest land value, and are the most urban. These results are consistent with Hearst seeking out the biggest, richest cities to increase his power and influence. Importantly, Hearst's presence in a county is entirely uncorrelated with FDR vote share in 1932, with a coefficient of -0.00003 (0.0003).

Although Hearst ownership is unrelated to political attitudes in the pre-period, there is still concern that Hearst counties, which are densely populated and highly urban, are likely to be on different time trends than smaller, rural counties. To account for this concern, I control for detailed county characteristics interacted with year, as well as construct various versions of the control group that become increasingly more similar to Hearst counties; see section 3.4.2 for details.

3.4.2 Constructing a comparable control group

I start with a sample of all US counties that had at least one daily paper in 1932.[2] Although newspaper circulation during this time was locally concentrated, there is some concern that nearby counties were partially treated. For example, those in a nearby county could have travelled to the city center to purchase a paper from a newspaper stand. Thus, I exclude 55 contiguous border counties from the control group.[3] Table 3.1 compares baseline 1930 demographic characteristics of Hearst vs. non-Hearst counties. Echoing results from Figure C.1, Hearst counties were highly urban, large, and had a higher share of immigrants than other counties.

Thus, to ensure differences in county characteristics are not driving the results, I control for a detailed array of county characteristics interacted with year, as described in Section 3.6.2. In particular, since Hearst counties are particularly large, I flexibly control for log total population interacted with year (using ventile indicators for county size). Additionally, I include the following demographic characteristics: log population located in cities, percent urban, percent Black, percent immigrant, percent farmers, percent illiterate (above the age of 10), percent unemployed, retail sales per capita, log land value per acre, share of farmland with tenant farmers, and whether the county is located in the South.[4]

Demographic $\times$ year controls may be particularly relevant given the political context. The positions of the Democratic party underwent major changes between the late 19th to mid 20th century, with the party initially opposing the abolition of slavery and often promoting laissez-faire capitalism. According to historians and political scientists, the party went through a major re-alignment in the 1928 presidential election, when Democrats nominated progressive candidate Al Smith (Clubb and Allen 1969). Al Smith was a liberal who promoted policies such as workers' rights and racial equality, bringing Blacks, urban blue collar workers, and immigrants into his coalition–voting blocks that would form the basis of support for FDR and the New Deal. Thus, county characteristic $\times$ year controls in this setting allow different demographic groups to have election-specific effects, accounting for any demographic-based realignment in response to the New Deal.

Moreover, radio became an increasingly popular source of political coverage in the 1930s. Thus, to adjust for potential differences in exposure to radio, in my analyses, I add controls for the percent of individuals who own a radio in their household, as well as exposure to Father Coughlin, a prominent right-wing radio host in the 1930s and 40s (Wang 2021).

In addition to employing detailed county controls, I confirm that results hold using alternative versions of the control group that more closely resemble Hearst counties. First, since Hearst's papers were located in large cities, in column 3 of Table 3.1, I restrict the sample to counties where the total population is above the 75th percentile of counties. Moreover, to more closely replicate characteristics of Hearst counties, I also show results using a propensity-score matched sample, matching on all characteristics in Table 3.1. Because of the large number of control counties compared to Hearst counties, I match the twenty nearest neighbors (with replacement) to a given

[2]For analyses at the newspaper level, I analogously use the sample of all US daily papers.

[3]I add these counties back to the sample in a robustness check.

[4]Log population located in cities captures the interaction between log population and percent urban. I include this set of control variables as demographic characteristics, such as percent Black and immigrant, are highly predictive of voting outcomes. Additionally, I include county characteristics and economic variables that are predictive of attitudes towards the Depression and New Deal policies (such as the share unemployed and the share of tenant farmers). On the whole, this set of control variables allow for pre-period balance in the event study design.

Hearst county. In column 4 of Table 3.1, I show that the matched sample more closely resembles the treated sample across baseline characteristics. In Figure 3.2, I show a visual map of treated and control counties for both the sample of large counties and the propensity-score matched sample. For newspaper text analyses, to lessen the computational intensity of the data mining procedure, I focus on text data from newspapers in propensity-score matched counties.

Finally, in Appendix C.4, I also construct a synthetic control for Hearst counties using the method proposed by Abadie (2003). In this approach, I select counties based on the *trend* in Democratic vote share in the pre-period, with the algorithm minimizing the distance between Democratic vote share across time in each Hearst county compared to a weighted average of control counties. In Figure C.5, I show that the synthetic control nearly exactly matches the trend in Democratic vote share of Hearst counties prior to 1936.

3.5 Documenting the shift in Hearst newspaper reporting

3.5.1 Did Hearst's political views affect newspaper endorsements?

To validate the historical narrative regarding Hearst's changing political allegiances and to demonstrate that these views were replicated in his papers, I use Gentzkow et al. (2011) data on which candidate a given paper endorsed for all presidential years between 1932 and 1968. I define a Hearst paper as any paper owned and operated by William Randolph Hearst in 1932. My endorsement measure, E_{nt}, takes on a value of 1 if the newspaper endorsed a Republican candidate in that election year, 0 if they endorsed an independent, and -1 if they endorsed a Democrat.[5] In the raw data, I find that all Hearst papers that reported a party endorsement in 1932 endorsed FDR, consistent with the historical record. Moreover, in every election from 1936-1956, all papers operated by the Hearst Corporation endorsed Republican presidential candidates. Finally, historians generally consider the death of Senator McCarthy in 1957 as marking the end of the Red Scare (Sabin 1999, p.5). Newspapers formerly tightly controlled by Hearst became much less consistently right wing in these later years, endorsing candidates from all parts of the political spectrum.

To formally test whether Hearst paper's exhibited stronger political biases than other newspapers at the time, I run the following specification to examine the effect of Hearst ownership on party endorsement:

$$E_{nt} = \sum_{\tau=1932}^{1968} \delta_\tau \, \text{Hearst}_{n\tau} + \alpha_t + \epsilon_{nt} \tag{3.2}$$

where n indicates newspaper, t indicates election year, $\text{Hearst}_{n\tau}$ is an indicator for being a Hearst paper in year τ, and α_t are year fixed effects. To improve power, I bin the endpoint year to include newspaper endorsements from 1960, 1964, and 1968.[6] All standard errors are clustered at the newspaper level.

Results are shown in Figure 3.3. As demonstrated in the raw data, Hearst newspapers were much more likely than other newspapers to endorse FDR, the Democrat, in 1932; however, in all

[5] Any newspaper with no endorsement reported in that election year is treated as missing.

[6] As many Hearst papers either ceased publication or merged with other newspapers in the mid-20th century, data on endorsements for papers formerly owned by Hearst are more sparse for later years.

years between 1936-1956, Hearst papers were far more likely to endorse a Republican. These results confirm that Hearst exhibited a sharp change in political attitudes in the mid 1930s, and that his own political views were reflected in the slant of his papers. Finally, in years 1960 and onwards, following both Hearst's death and the end of the Red Scare, these same papers were neither more or less likely to endorse Republican political candidates. This null effect confirms that treatment was confined to a historically distinct period.

3.5.2 Did Hearst's anticommunism affect newspaper reporting?

While Section 3.5.1 demonstrates that Hearst's papers were much less likely to endorse a Democrat after his political views veered to the right, how much did anticommunist sentiment actually affect reporting? To analyze the content of Hearst papers directly, I use newspaper text data and OCR software within Newspapers.com and NewspaperArchive.com, the two sources with the largest historical coverage (Beach and Hanlon 2022). While coverage of the Hearst papers is incomplete, I locate six Hearst newspapers (*The San Francisco Examiner, The Oakland Post Enquirer*, *Pittsburgh Sun-Telegraph*, *Washington Times*, *Washington Herald*, and *San Antonio Light*) and eighteen newspapers in propensity-score matched counties available within these databases during the 1930s and 40s.[7]

To capture Hearst's growing focus on anti-communism directly within news coverage, I scrape the count of total mentions of the word "communist" for each newspaper-year.[8] In all analyses, I include newspaper fixed effects to control for the average length and content of a given newspaper. Moreover, as some newspapers might not have full coverage within these directories for a particular year, I compare the number of mentions of "communist" to mentions of a neutral word, either Monday or rain, as suggested by Beach and Hanlon (2022). Let $N_{nt} = \frac{\text{Mentions of "communis"t}}{\text{Mentions of "Monday"}}$ for a given newspaper-year. The mean number of communist mentions per year in a given newspaper is 493, compared to 5,424 mentions of Monday.

I run the following regression to assess newspaper content related to communism:

$$N_{nt} = \sum_{\tau=1929}^{1944} \delta_\tau \, \text{Hearst}_{n\tau} + \alpha_t + \zeta_n + \epsilon_{nt} \tag{3.3}$$

where n indicates newspaper, t indicates election year, $\text{Hearst}_{n\tau}$ is an indicator for being a Hearst paper in year τ, α_t are year fixed effects, and ζ_n are newspaper fixed effects. For ease of interpretation, I scale N_{nt} by the mean so that regression coefficients can be interpreted as the percent increase in communist mentions relative to the sample average.

In Figure 3.4 Panel A, I show estimates of each $\hat{\delta}_\tau$, as well as a pooled estimate where I replace yearly indicators with indicators for Hearst × pre-1934 and Hearst × post-1934.[9] The reference year is 1932. I find that Hearst ownership had no effect on mentions of "communist" prior to 1934, with an average coefficient of -7.5% (s.e.=4.0%). In contrast, in the post-period, Hearst ownership

[7] I drop *Washington Times* and *Washington Herald* from the analysis between 1940-1944, as these newspapers were sold by Hearst in 1939.

[8] Using word mentions as a proxy for political content is an approach commonly used in this literature. For instance, see Ferrara et al. (2022) and Ottinger and Winkler (2020).

[9] In this specification, I estimate the effect in 1934 separately, as I consider this to be the transition year in which Hearst began veering to the right.

predicts a 36.4% (s.e.= 14.8%) increase in communist mentions. In Figure C.3, I show estimates of $\hat{\delta}_\tau$ when I replace the word "Monday" with another neutral word, "rain." The results are very similar.

Next, were Hearst newspaper even more anticommunist than other Republican newspapers of the time? To get at this question, I divide control newspapers in propensity score matched counties to those with a Republican political affiliation (seven papers) and those with a Democratic or Independent political affiliation (eleven papers). I then rerun the content analysis for each group of controls. Results are shown in Figure C.4. I find some evidence of effect heterogeneity: Hearst papers were particularly more likely to contain the word "communist" after 1934 than non-Republican papers. However, Hearst papers were also more likely to mention the word "communist" than other Republican papers, corroborating the historical evidence that Hearst papers were particularly anticommunist.

Lastly, did Hearst explicitly link communism with FDR? In Figure 3.1, I show anecdotal evidence from an issue of *The San Francisco Examiner*, which declares, "Neither is the shameful act to be denied that COMMUNISTS are supporting Franklin D. Roosevelt for re-election in 1936." In Figure 3.4 Panel B, I test this more systematically by exploring how often the words "Roosevelt" and "communist" appear together on the same page.[10] To adjust for both newspaper coverage in that year as well the likeliness that two words randomly appear together on the same page, I normalize joint mentions of "Roosevelt" and "communist" by joint mentions of "Monday" and "rain." Then, I rerun Eq. 3, redefining $N_{nt} = \frac{\text{Mentions of "communist" AND "Roosevelt"}}{\text{Mentions of "Monday" AND "rain"}}$. I again find that Hearst ownership had no effect in the pre-period, with a coefficient of 13.4% (s.e.=10.3%). However, as Hearst turned against FDR, he increasingly implicated him with communism: after 1935, Hearst ownership increased the probability of "Hearst" and "Roosevelt" being mentioned together by 53.4% (s.e.=24.7%).

Although coverage of the Hearst papers is incomplete within these archives, the above analysis provides suggestive evidence that the Hearst papers disportionately reported on communism compared to other newspapers in the same year, as well as disportionately implicated FDR with communism. On the whole, the prior two sections demonstrate that 1) Hearst papers were much more likely to endorse a Republican after 1936 and 2) reporting on communism sharply increased concurrently with this change in endorsements. This evidence provides support for the historical narrative that Hearst turned against FDR in the mid-1930s, and that his turn to the right was driven by anticommunist sentiment.

3.6 Empirical approach

3.6.1 Measuring Hearst penetration

My baseline analysis examines the effect of Hearst newspaper penetration on support for FDR in the 1936, 1940, and 1944 elections. I define my primary Hearst exposure variable as $\frac{\text{circulation}}{\text{population}}$, or Hearst circulation per capita. One concern here is that circulation is likely endogenous within a given market, as more residents will subscribe to a newspaper when it matches their own political

[10]The current OCR software available within Newspapers.com and NewspaperArchive.com cannot differentiate between a word appearing in the same article versus on the same page. Thus, this is likely a noisy proxy of Roosevelt's name being directly mentioned in reference to communism.

views. Thus, I use circulation data from 1932, before Hearst became staunchly anticommunist and anti-FDR. Therefore, Hearst penetration captures how popular his papers were *before* they began printing anti-communist rhetoric, eliminating the potential of political sentiment in the mid-1930s to influence my independent variable. Among Hearst counties, the mean value of Hearst circulation per capita is 0.25, with a standard deviation of 0.13.

As further robustness checks, I use another measure of Hearst penetration: newspaper share, or the count of Hearst papers over total number of daily papers within a county. Within Hearst counties, there were an average of 10.3 daily newspapers operating in 1932 (s.e.=9.0), with a mean of 1.59 (s.e.=0.5) Hearst papers. This alternative measure is plausibly more exogenous, as it does not use circulation data at all. However, my primary measure, Hearst circulation per capita, best reflects how much influence these papers had within individual counties.

3.6.2 Difference-in-difference design

To examine the effect of Hearst penetration on support for FDR, I use a difference-in-difference design. Since Hearst's anticommunist rhetoric began in the mid-1930s, I consider FDR vote share in the 1932 election to be prior to treatment, allowing me to examine the change in vote share over time.

The benefits of difference-in-difference is that it does not require control and treatment groups to be randomly assigned; instead, it requires parallel trends in the absence of treatment. Although the geographic location of Hearst's papers are unrelated to counties' support for FDR in 1932, all of his daily papers were located in highly urban areas, which are not representative of all counties across the US. To account for potential violations in parallel trends, I control for county fixed effects as well as detailed county characteristics interacted with year, allowing these characteristics to have differential effects over time.

Finally, in some specifications, I control for political persuasion of newspapers more broadly. This includes controlling for total newspaper circulation per capita (excluding the circulation of Hearst papers), reflecting how popular newspapers were within a given county. Additionally, I control for the (leave-out) share of newspaper circulation where the newspaper had a Republican political affiliation, as well as the share of newspaper circulation where the newspaper had a Democratic political affiliation.

My main empirical specification is as follows:

$$Y_{it} = \beta H_i \times \text{Post}_t + \mathbf{X}_{it}^0 \gamma + \eta_t + \zeta_i + \varepsilon_{it} \tag{3.4}$$

where Y_{it} is vote share for FDR, H_i is Hearst circulation per capita in county i, Post_t is an indicator for post-1932, η_t are year fixed effects, ζ_i are county fixed effects, and $\mathbf{X}_{it}^0$ are baseline county controls interacted with year. Baseline controls include log total population (in ventiles), log population located in cities, percent urban, percent unemployed, percent Black, percent immigrant, percent farmers, percent illiterate (above the age of 10), percent who own a radio in their household, retail sales per capita, log land value per acre, share of farmland with tenant farmers, and whether the county is located in the South. $\hat{\beta}$ is the estimated difference-in-difference coefficient, which represents the effect of Hearst newspaper penetration on the change in vote share for FDR. I show standard OLS standard errors clustered by county,[11] as well as wild-bootstrap standard er-

[11] I cluster all New York City counties together, as they are essentially one treatment cluster.

rors, which often perform better with a small number of treated clusters (Cameron, Gelbach, and Miller 2008, Canay, Shaikh, Santos 2021).

Additionally, I use an event study approach, where I replace $H_i \times \text{Post}_t$ with interactions between Hearst circulation and year, extracting separate coefficients for all elections. I generalize Y_{it} to represent Democrat vote share in a given presidential year. This latter approach allows me to both assess parallel trends, as well as determine whether Hearst's reporting had a longer run effect on support for the Democratic party.

3.6.3 Within-county design for long run public opinion

Finally, to assess whether anti-redistributive attitudes persisted among those exposed to Hearst, I collect survey data from ANES. I use responses to two survey questions asked across multiple years that are highly related to redistributive beliefs: support for public healthcare and attitudes towards social welfare programs. First, for survey years 1970-1994, ANES asks respondents on a seven-point scale whether the government should provide health insurance for all (1) or health insurance should be fully private (7). I transform responses of 3 or below on the 7-point scale to be indicators that the respondent supports public healthcare. Additionally, from 1992-1994, ANES asked respondents whether spending on welfare programs should be reduced, stay the same, or increase, and from 1976-1994, how negatively respondents feel about those on welfare, from 0 (very negative) to 100 (very positive). To improve power, I combine these latter two variables to create a composite "negative towards welfare" variable, equal to 1 if respondents either feel welfare spending should be reduced or feel negatively towards those on welfare (below the 25th percentile of respondents).

Next, using current county of the respondent and how many years the respondent has been in a said county, I create an indicator, Exposed_r, indicating that the respondent was exposed to a Hearst newspaper. I consider one exposed to Hearst if they resided in a Hearst county while the papers were actively owned by Hearst and promoting anticommunist rhetoric. Based on my prior analysis of newspaper content, I consider this exposure window to start in 1935. For newspapers sold or closed down before 1957, I consider the end exposure year to be the year of sale or closure. Otherwise, I consider the last year of exposure to be 1957, considered to be the end of the Red Scare.[12] I run the following regression:

$$Y_{rt} = \alpha_0 \text{Exposed}_r \times \text{Adult}_r + \alpha_1 \text{Exposed}_r \times \text{Child}_r + \mathbf{X}_{rt}\pi + \zeta_{c(r)} + \gamma_t + \eta_{rt} \tag{3.5}$$

where Y_{rt} is the survey response of interest, Adult_r is an indicator for being exposed to Hearst for at least one year as an adult, Child_r is an indicator for being exposed to Hearst only as a child, $\zeta_{c(r)}$ are county fixed effects, γ_t are survey year fixed effects, and $\mathbf{X}_{rt}$ are respondent characteristics, including birth cohort in ten-year bins, age in five-year bins, gender, race, income binned in percentile buckets,[13] urban, suburban, or rural residence, education buckets (less than high school, high school, some college, or college or higher), major religious group (Protestant, Catholic, Jew-

[12] ANES asks respondents how long they have been in their current county but not for a list of all counties they have lived in. Thus, due to data limitations, I cannot track individuals who were exposed to Hearst but then moved to another county.

[13] As defined by ANES, these are 0-16th, 17th-33rd, 34th-67th, 68th-95th, and 96th-100th.

ish, or other), occupation category[14], and whether the respondent considers themselves to be working, middle, or upper middle class. Importantly, this specification compares respondents in the *same county* who were exposed or unexposed to Hearst, absorbing any other characteristics specific to counties that affect voting outcomes. Moreover, by controlling for detailed age buckets and survey year, I account for the fact that older respondents or respondents surveyed in earlier years may hold generally more conservative beliefs. Again, all standard errors are clustered by county.

3.7 Results

3.7.1 Support for FDR

Table 3.2 shows estimates of $\hat{\beta}$, the effect of Hearst circulation on vote share for FDR in elections 1936 and later. Column 1 shows results for the full sample with baseline controls, column 2 shows results with additional newspaper controls (for other newspapers operating within a given county), column 3 shows results for the sample of large counties, and column 4 shows results for the propensity-score matched sample. Column 2 suggests that a 10 percentage point increase in Hearst circulation decreased FDR vote share by 2.14 percentage points (s.e.=0.77), or 3.6% of the mean vote share for FDR. This result is robust to using more and more selective versions of the control group, as shown in columns 3 and 4. In particular, the point estimate using the propensity-score matched sample is highly aligned with the baseline results, but slightly noisier due to the much smaller sample size.

3.7.2 Event study

In Figure 3.5, I show the results of the event study specification, where the outcome of interest is Democratic vote share over time.[15] I show balance in the pre-period, with coefficients prior to 1932 close to 0 and statistically insignificant across all years. In contrast, I show that coefficients on $H_i \times 1936$, $H_i \times 1940$, and $H_i \times 1944$ are consistently negative and statistically significant, showing that Hearst influenced voting outcomes in all three FDR elections following his turn to the right. This shift in voting outcomes exactly mirrors Hearst's shift in both political endorsements and reporting on communism, as shown in Figures 3 and 4.

Moreover, though the Red Scare died down at the end of the 1950s, I also show some suggestive evidence of persistence: I find that point estimates remain consistently negative–though are insignificant– through 1972. In Section 3.7.5, I directly relate this persistent effect on voting outcomes with attitudes towards redistribution among those exposed to Hearst.

One concern with the event study design are the small number of potentially distinct treated counties. Thus, in the Appendix, I also use a synthetic control design (Abadie and Gardeazabal 2003, Abadie et al. 2010). Rather than relying on parallel trends, this approach matches prior trends in Democratic vote share using a weighted average of control counties, which make up the "synthetic" control. The method relies on the assumption that each Hearst county would have

[14] These are defined by ANES as: 1) professional/managerial, 2) clerical/sales, 3) skilled, semi-skilled, and service workers, 4) laborers, 5) farmers, and 6) homemakers or no occupation.

[15] To improve power in these specifications, I use the full sample of counties with daily papers, but include all control variables, including newspaper controls.

followed the same voting trajectory as its respective synthetic control under the counterfactual of no shift in Hearst's ideology. I find that the synthetic control closely matches the voting pattern of Hearst counties in the pre-period, and I find similar qualitative results to that of the event study. See Appendix C.4 for details.

However, one particular advantage of the event study and difference-in-differences approaches is that they allow me to include detailed county characteristic × year interactions, enabling the effect of a given demographic group to vary in each presidential election. Demographic × year controls may be particularly revelant in this setting, as the New Deal coalition was distinct from that of prior presidential elections, and demographic voting patterns may be correlated with Hearst presence. Most notably, Hearst counties have a very high share of immigrants, as documented in Table 3.1. Prior to the 1928 voter re-alignment, immigrants were far *less* likely to vote for a Democrat; following 1928, they were far more likely. Thus, using the full sample as a control group without these detailed county controls, I would have have falsely concluded that Hearst had no effect (Table C.1 column 1). Limiting the sample to only large counties, I estimate an effect much closer to the baseline estimate (Table C.1 column 3). Nevertheless, adding demographic × year interactions makes the estimate slightly larger and more precise (Table C.1 column 4). I perform additional robustness checks of interest in the following section.

3.7.3 Concerns regarding omitted variable bias

The primary concern with comparing Hearst and non-Hearst counties is that Hearst counties are fundamentally distinct from non-Hearst counties. As shown in Table 3.1, Hearst counties are larger and have a higher share of immigrants than non-Hearst counties. Although both the large county sample and propensity-score matched sample have characteristics that more closely resemble Hearst counties, Hearst bought newspapers in the largest US cities, making finding a comparable control group difficult. In other words, despite the detailed county by year controls used in the analysis, there is still a concern over omitted variable bias.

Although the reader should be reassured by the lack of pre-trends in the event study analysis, I discuss any remaining concerns below. One notable point is that any remaining omitted variable bias is likely to be reverse-signed: FDR was particularly popular among urban voters, and Hearst counties had a much higher share of immigrants, Jews, and Catholics, all of whom were much more likely to support FDR. This point is demonstrated clearly in Table C.1, where I show the results with and without the county characteristic × year controls. The effect of Hearst circulation on vote share for FDR is actually *reverse* signed without these control variables (column 1), and the effect only becomes stronger as I add additional controls (column 2). Moreover, in columns 3 and 4, I show the result with and without controls for the large county sample, which are almost identical. These latter results provide reassurance that these alternative samples – which more closely resemble Hearst counties – suffer from minimal omitted variable bias.

Moreover, I conduct several additional tests to mitigate concerns over omitted variable bias. First, in Table 3.4 (see the subsequent section), I drop the largest Hearst counties from the analysis, and results remain highly robust. Perhaps most convincingly, in Figure 6, I run a placebo test. First, I drop true treated Hearst counties from the analysis. Instead, I consider "treated" counties to be the largest US counties that Hearst did not operate in. I assign these counties artificial values of

Hearst circulation per capita based on those of true treated counties[16], and then I rerun the baseline event study specification. The placebo test answers the question: does Hearst circulation actually have no effect, but instead something about differential trends between large and non-large US counties bias my results? I find insignificant coefficients across the board in the post-period for the placebo sample. This test should reassure the reader that the effect is not driven by Hearst counties being large and urban.

3.7.4 Additional robustness checks

One potential concern with analysis is that newspaper circulation is not exogenous, as it may reflect political opinions of readers. Although readers likely played a smaller role in Hearst's decision to purchase a given paper, an individual chooses whether to have a particular paper delivered. To account for this problem, in the main specification, I use circulation share data from 1932, before Hearst became a hardline anticommunist. Moreover, I use an alternative measure of Hearst penetration: newspaper share, or number of Hearst newspapers in a county over total daily papers in that county. This measure is plausibly more exogenous, as it does not use circulation data. I display results using this alternative definition of the treatment variable in Table 3.3. Treatment effects using newspaper share are highly aligned with the primary results; increasing Hearst newspaper share by 10% decreases support for FDR by 2.02 percentage points (s.e.=0.72).

Secondly, recent innovations to the literature have demonstrated that OLS estimates of difference-in-difference coefficients may place negative weights on event time specific treatment effects, particularly in cases of staggered treatment adoption and treatmet effect heterogeneity (Borusyak et al. 2024, Callaway and Sant'Anna 2021, Chaisemartin and D'Haultfoeuille 2020). In this analysis, the treatment group is fixed and all treatment begins in 1936, making this issue less of a concern here. Nethertheless, to ensure that negative weights are not affecting my results, I employ the method proposed by Callaway and Sant'Anna (2021). I show results in Table C.2 and provide more details in Appendix C.1. The estimates are very similar across approaches.

In Table 3.4, I perform miscellaneous relevant robustness checks. First, one potential concern is that the largest US counties – Manhattan, Cook (Chicago), and Los Angeles county, all of which Hearst operated in – may be fundamentally incomparable to all potential control counties. To eliminate this concern, I drop these counties in column 2.[17] Second, Wang (2021) showed that Father Coughlin, another large media personality in the 1930s and 40s, also influenced vote shares for FDR via his radio program, with Coughlin also turning against FDR in later elections. Thus, to eliminate concern that Coughlin radio signal and Hearst newspaper circulation were spatially correlated, I control for Coughlin radio signal $\times$ post-1932 in column 3.[18] Finally, to account for the fact that error terms may be geographically correlated, I cluster at both the state level (column 4) and report Conley standard errors (column 5), which adjust for spatial correlation (Conley 1999). In Table C.3, I perform various additional robustness checks (see Appendix C.2). These include adding back contiguous border counties to the sample, as well as including more granular control variables, including state $\times$ year fixed effects and controls related to the strength of the Depression

[16] I assign the largest placebo county the value of Hearst circulation per capita corresponding to the largest Hearst county, and then continue matching in order of county size, for a total 22 treated placebo counties.

[17] The results are robust if I also drop all New York City outer boroughs.

[18] I also confirm that my results hold using Wang's group of control counties, which excludes regions in the geographic South.

and New Deal spending. The results are highly aligned and statistically significant across all specifications.

Finally, another relevant question is whether Hearst's other media holdings had an influence on voters. In particular, in the late 1920s, Hearst began to take an interest in radio, buying up local stations to compliment his newspapers. I geolocate these radio stations and add radio exposure as a treatment variable in Appendix C.3 and Table C.4. I find some weak evidence that radio, in addition to newspaper circulation, affected voting outcomes.

3.7.5 Long run outcomes

In Figure 3.6, I plot $\hat{\alpha_0}$ and $\hat{\alpha_1}$, estimated from Equation 5. Across all survey years, 45.9% of respondents stated that they support public healthcare, while 38.1% of survey respondents answered that they feel negatively towards welfare; Hearst-exposed respondents hold even more anti-redistributive beliefs. I find that those exposed to Hearst as adults are less likely to support public healthcare and are more likely to feel negatively towards welfare programs or recipients. Specifically, I find that being exposed to Hearst decreases support for public healthcare by 4.4 percentage points (s.e.=1.3 p.p.), or 9.6% of baseline support, and increases negative attitudes towards welfare by 5.1 percentage points (s.e.= 1.7 p.p.), or 13.4% of baseline. I cannot reject no effects across the board for those exposed to Hearst as children, suggesting that attitudes were not passed from parents to children.

One concern with the prior analysis is that those exposed to Hearst tend to be older than those not exposed to Hearst, and results could be reflecting that older respondents tend to be more conservative. Additionally, those in older cohorts could hold more conservative views, even accounting for age. In Table C.5, I show regression results from Equation 4. In Panel A, I show the baseline results, which controls for detailed respondent characteristics, including five-year age buckets and birth cohorts in ten-year buckets. In Panel B, I restrict the analysis to those old enough to be potentially exposed to Hearst as adults. Finally, as the Red Scare faded out gradually rather than ending abruptly in 1957, in Figure C.6, I conservatively redefine the last year of potential exposure to Hearst-era propaganda as 1960. I find very similar results across all specifications.

3.7.6 Why did Hearst have such a large impact?

Compared with other studies of the media's impact, I find that Hearst had a particularly large effect on public opinion. For instance, Dellavigna and Kaplan (2007) find that the entry of Fox News increased Republican vote share by only 0.4-0.7 percentage points. On the other hand, I find that a 10% increase in Hearst circulation per capita decreased vote share for FDR by 2.1 percentage points. With a mean Hearst circulation per capita of 25% across treated counties, this implies that Hearst presence in a county decreased vote share by about 8 percentage points. However, my estimates are highly aligned with that of Enikolopov et al. (2011), who find that access to the only non-government run TV channel in 1999 Russia decreased support for the party in power by 8.9 percentage points.

What could explain the difference in magnitude between these different settings? One potential explanation is the lack of presence of other media sources. In particular, Durante and Knight (2012) find that highly partisan media causes readers to substitute towards more like-minded news sources, mitigating any potential effects of a single media source. In the case of the Russian media market, in many regions, all other sources of media were run by the government. Similarly, perhaps more limited competition during the FDR era, both due to media consolidation and lack of alternative forms of news coverage, could help explain the large influence of the Hearst papers.

Following this logic, in counties with little alternative sources of news coverage, one should expect Hearst papers to have had more influence. To test this theory directly, I examine effect heterogeneity by the presence of other media in each county. First, during this era, radio was gaining increasing popularity. Thus, I divide the sample into counties with above and below median radio ownership (defining the median based on Hearst counties). Second, I test the diversification of the newspaper market directly by dividing the sample into counties with an above and below median number of (non-Hearst) newspapers. I show results in Table 3.5. The results are striking: in diversified media markets, Hearst newspapers seem to have had no impact at all. Thus, this table provides convincing evidence that the impact of the media highly depends on market concentration and individuals' ability to sustitute to alternative forms of coverage.

3.8 Discussion

The rise of the Soviet Union spurred a domestic Red Scare in the US in the early to mid-20th centuries, with real political ramifications. From William Randolph Hearst's "yellow journalism," to McCarthy's effort to purge communists from government, to fear of the atomic bomb, anti-Soviet and anticommunist sentiment played a prominent role in American political discourse for decades. This paper examines how this rhetoric influenced American households, particularly in their voting behavior and support for redistributive policy, which is frequently associated with socialism. I find that exposure to Red Scare propaganda via the newspapers of William Randolph Hearst decreased support for FDR in the 1936-1944 presidential elections. Moreover, I find evidence that these effects are persistent, decreasing support for public healthcare and welfare programs through the late 20th century.

In contrast with previous work which finds little effect of media ownership, this case study offers insight on when owners may be important in shaping both partisan slant and political outcomes. In particular, Hearst possessed extreme beliefs, controlled up to 60% of newspaper circulation in the areas he purchased papers, and ruthlessly promoted his own ideology even when it nearly led to bankruptcy. In line with my results, Larcinese et al. (2011) and Durante and Knight (2012) also provide case studies where ideologically-motivated owners played a key role in shaping news coverage. However, Durante and Knight (2012) find that highly partisan media causes readers to substitute towards more like-minded news sources, mitigating ownership effects. My results add to the evidence that in low substitutability markets, individuals owners have more power to influence voters.

The widespread and long-lasting influence of Hearst's anticommunist rhetoric adds strong support to the growing evidence that both media personalities and newspaper slant can influence public opinion, with real effects on voting behavior. Thus, this study has important implications with regards to populists' use of media in contemporary elections. In fact, Hearst is sometimes described as the originator of fake news, publishing afactual, sensationalized stories that corroborated his personal viewpoints. Ultimately, this paper both emphasizes the role of journalism in influencing political outcomes, as well as demonstrates the long run legacy of anticommunist sentiment in American politics.

Figure 3.1: Article from the *San Francisco Examiner,* October 9, 1936

Deport the Alien Reds! Rescind the Recognition Of Soviet Russia!

THE fact that Communism—nurtured by Soviet Russia—is seeking to destroy the free institutions of America is not to be denied.

Neither is the shameful fact to be denied that COMMUNISTS are supporting Franklin D. Roosevelt for re-election in 1936 IN ORDER TO BUILD A REVOLUTIONARY "PEOPLES FRONT" IN 1940.

Mr. Roosevelt has said that he repudiates this support.

But what has he done to prove that he repudiates it?

Russian Communism in America can be officially repudiated only by two courses of executive action:

1—By immediately deporting all alien communists, as the law requires; and

2—By immediately rescinding the recognition of Soviet Russia.

Source: Schrecker (2021), PBS.

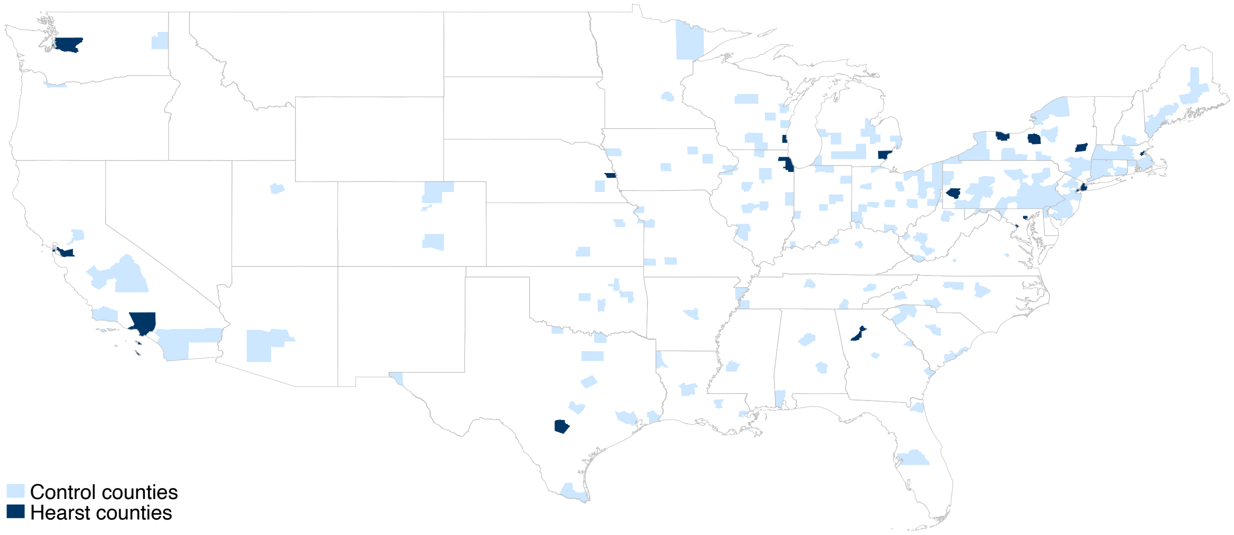
Table 3.1: Population characteristics for Hearst and non-Hearst counties

	Hearst counties	Non-Hearst counties		
		All counties with daily papers	Large counties	Matched counties
Log Population	13.47	10.59	11.78	12.39
Urban	0.92	0.46	0.65	0.75
Black	0.06	0.08	0.09	0.05
Unemployed	0.06	0.04	0.05	0.06
Immigrant	0.19	0.06	0.09	0.17
Farmer	0.02	0.32	0.14	0.07
Male	0.50	0.51	0.50	0.51
Native white	0.72	0.84	0.81	0.75
Above 65	0.05	0.06	0.06	0.06
Catholic	0.22	0.10	0.15	0.27
Occupational income score	0.11	0.08	0.09	0.11
Manufacturing	0.25	0.17	0.27	0.28
Agriculture	0.02	0.30	0.13	0.06
Illiterate	0.03	0.03	0.03	0.03
Own a radio	0.55	0.33	0.41	0.48
Log farm size	0.06	0.07	0.07	0.07
Log land value per acre	0.06	0.04	0.04	0.05
Tenant farmers	0.30	0.32	0.28	0.22
South	0.09	0.26	0.22	0.15
Democratic news share	0.05	0.20	0.14	0.06
Republican news share	0.11	0.19	0.19	0.19
Newspaper circulation per capita	0.53	0.51	0.61	0.60
FDR vote share 1932	0.62	0.63	0.57	0.57
Counties	22	940	216	22
Newspapers	28	1,595	607	93

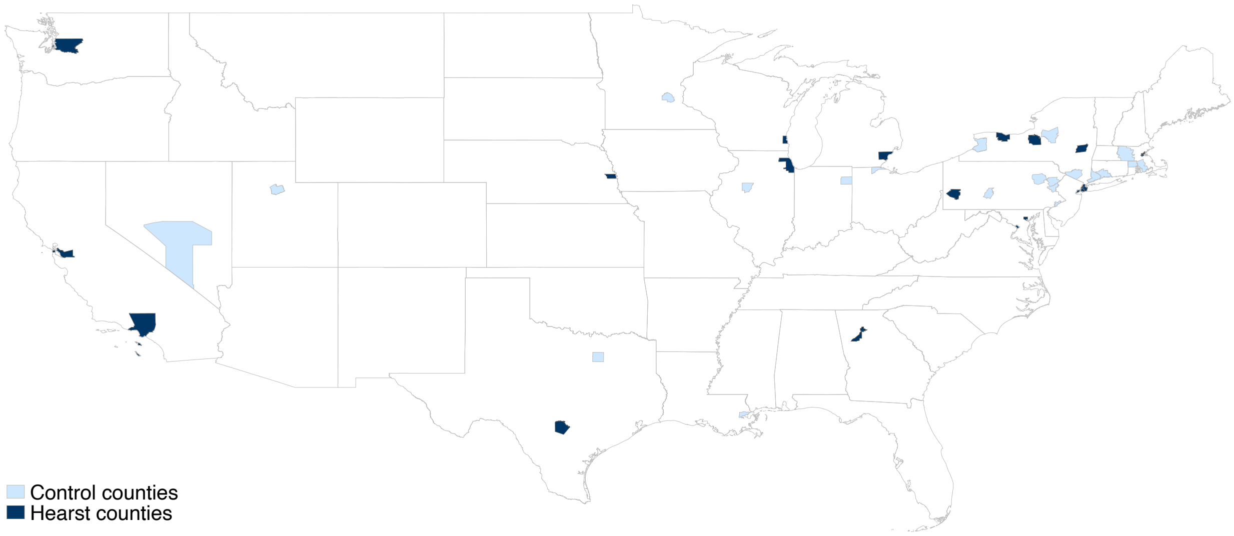
Note: This table compares sample characteristics for Hearst versus non-Hearst counties. Column 1 shows sample means for Hearst counties, column 2 shows sample means for all non-Hearst counties with daily papers, column 3 shows sample means for non-Hearst counties with daily papers where the total population is above the 75th percentile, and column 4 shows means for the propensity-score matched sample. In column 1, newspapers is a count of newspapers owned by Hearst (although there are also non-Hearst papers in those counties). In column 4, I list unique counts of newspapers and counties, but matching is done with replacement. Democratic and Republican news shares do not sum to 1, as many newspapers were either politically unaffialiated or listed their political affiliation as Independent.

Figure 3.2: Map of Hearst and control counties

A: Large counties

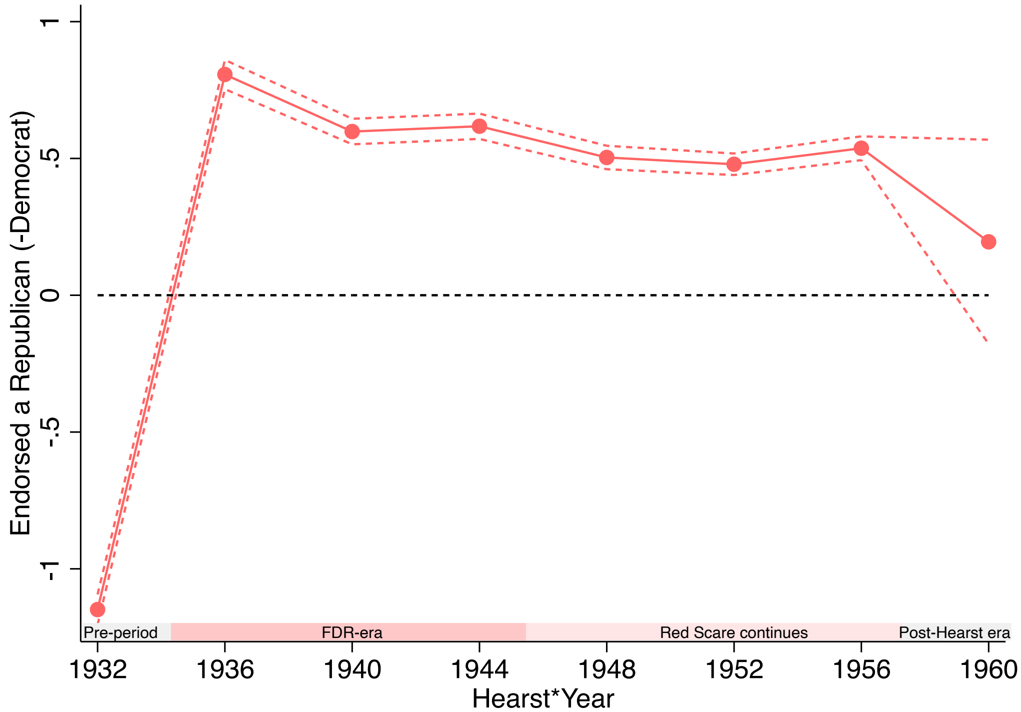


B: Propensity-score matched counties



Note: This figure shows treatment and control counties used for the difference-in-difference analysis. The counties in dark blue are treated Hearst counties, while the counties in light blue are control counties. Panel A shows control counties for the sample of large counties. Panel B shows control counties for the propensity-score matched sample. Matching is done with replacement, so a control county may be matched to several treated counties.

Figure 3.3: Effect of Hearst ownership on presidential party endorsement

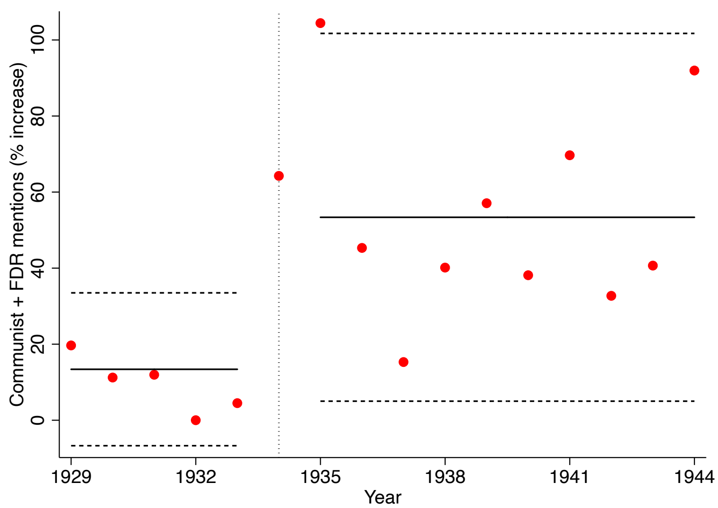


Note: This figure shows the effect of a newspaper being owned by Hearst on the likeliness of that paper endorsing a Republican (or Democratic) candidate for president. 1 indicates endorsing the Republican, 0 endorsing an Independent, and -1 endorsing a Democrat. Dots show coefficients on Hearst × year interactions, and dashed lines show 95% confidence intervals. The endpoint is binned to include years 1960, 1964, and 1968. I control for election year, and standard errors are clustered at the newspaper level.

Figure 3.4: Effect of Hearst ownership on mentions of “communist”

A: Mentions of “communist”

B: “Communist” and “Roosevelt” appearing together



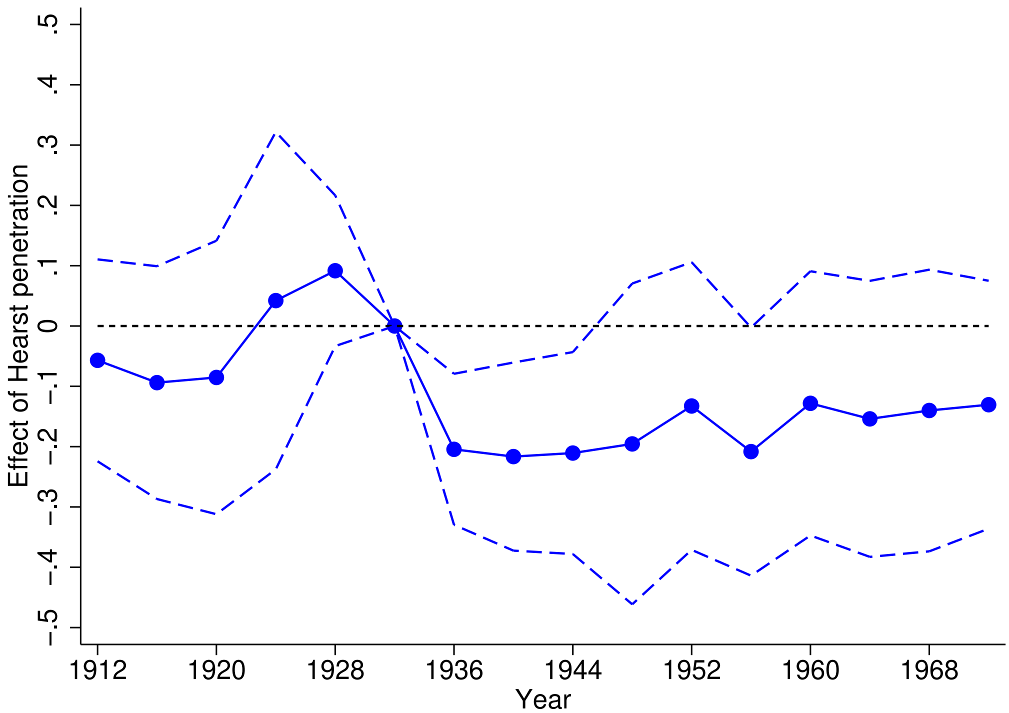
Note: Panel A shows the effect of Hearst ownership on newspaper mentions of the word “communist” each year relative to “Monday” (a neutral word). The dependent variable is the word count ratio between communist and Monday for that year. Panel B shows the effect of Heart ownership on joint mentions of “communist” and “Roosevelt” (relative to joint mentions of “Monday” and “rain”). In both panels, I divide the dependent variable by the sample mean so that coefficients represent the percent increase in mentions. Red dots show yearly point estimates, and the black lines show the average effect between 1929 and 1933 (the pre-period) and 1935 and 1944 (the post-period). The reference year is 1932. Dotted lines show 95% confidence intervals for the mean pre- and post-period effects.

Table 3.2: Effect of Hearst newspaper circulation on support for FDR

	FDR vote share post-1932			
	(1)	(2)	(3)	(4)
Hearst circulation per capita × post-1932	-0.211	-0.214	-0.177	-0.209
	(0.075)	(0.077)	(0.080)	(0.098)
Wild bootstrap confidence interval	[-0.436, -0.079]	[-0.418, -0.076]	[-0.373, -0.018]	[-0.389, -0.018]
Outcome mean	0.596	0.594	0.584	0.579
Observations	3,861	3,757	940	168
Controls				
County fixed effects	Yes	Yes	Yes	Yes
County char. × year	Yes	Yes	Yes	Yes
Newspaper char × year	No	Yes	Yes	Yes
Sample	Full	Full	Large counties	Matched

Note: The table shows difference-in-difference regression results, where row 1 displays $\hat{\beta}$, the effect of Hearst circulation on vote share for FDR in elections 1936 and later. Column 1 is a regression on the full sample using baseline controls, column 2 is a regression on the full sample using additional newspaper controls, column 3 is a regression on the sample of large counties, and column 4 is a regression on the propensity-score matched sample. I show standard errors clustered by county (in parentheses), as well as wild bootstrap confidence intervals. Sample sizes slightly differ between columns (1) and (2) due to missing newspaper circulation data in some counties.

Figure 3.5: Effect of Hearst newspaper circulation on Democratic vote share over time



Note: This figure shows regression coefficients from the event study specification, where Democratic vote share is the dependent variable. Dots show point estimates, and dashed lines show 95% confidence intervals. I use the full sample of counties with daily papers, and controls include baseline county controls interacted with year and newspaper controls interacted with year. Standard errors are clustered by county.

Figure 3.6: Placebo treatment counties event study

Note: This figure shows regression coefficients from the placebo analysis, where I consider "treated" counties to be the largest US counties that Hearst did not operate in. I assign these counties artificial values of Hearst circulation per capita based on those of true treated counties, and then I rerun the baseline specification. Hearst counties are excluded from the analysis. Dots show point estimates, and dashed lines show 95% confidence intervals.

Table 3.3: Effect of Hearst newspaper share on support for FDR

	FDR vote share post-1932			
	(1)	(2)	(3)	(4)
Hearst newspaper share × post-1932	-0.191	-0.202	-0.186	-0.218
	(0.073)	(0.072)	(0.068)	(0.101)
Wild bootstrap confidence interval	[-0.354, -0.046]	[-0.346, -0.068]	[-0.317, -0.046]	[-0.430, -0.032]
Outcome mean	0.596	0.594	0.584	0.579
Observations	3,861	3,757	940	168
Controls				
County fixed effects	Yes	Yes	Yes	Yes
County char. × year	Yes	Yes	Yes	Yes
Newspaper char × year	No	Yes	Yes	Yes
Sample	Full	Full	Large counties	Matched

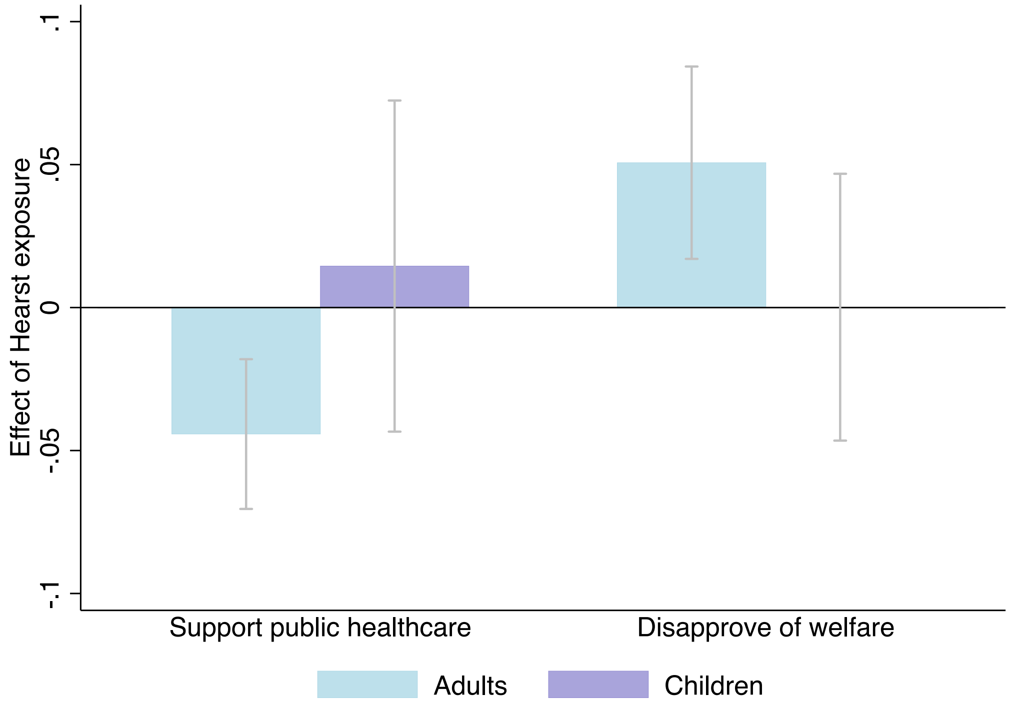
Note: This table shows difference-in-difference regression results using the newspaper share measure of Hearst penetration. Newspaper share is defined as the count of Hearst papers over total number of daily papers within a county. Column 1 is a regression on the full sample using baseline controls, column 2 is a regression on the full sample using additional newspaper controls, column 3 is a regression on the sample of large counties, and column 4 is a regression on the propensity-score matched sample. I show standard errors clustered by county (in parentheses), as well as wild bootstrap confidence intervals.

Table 3.4: Robustness checks

	FDR vote share post-1932				
	(1) Baseline	(2) Drop large Hearst counties	(3) Control for Coughlin signal	(4) Cluster by state	(5) Spatial clustering
Hearst circulation per capita × post-1932	-0.214	-0.218	-0.211	-0.214	-0.214
	(0.077)	(0.076)	(0.078)	(0.070)	(0.052)
Outcome mean	0.594	0.594	0.594	0.594	0.594
Observations	3,757	3,745	3,757	3,757	3,757

Note: This table shows difference-in-difference regression results under various alternative specifications. Column 1 shows the baseline estimate using the full sample and all baseline control variables, including newspaper circulation variables. In column 2, I drop Manhattan, Cook, and Los Angeles counties. In column 3, I control for an indicator of Coughlin's radio signal strength × post-1932, from Wang (2021). In column 4, I cluster standard errors by state, and in column 5, I report Conley (1999) spatial standard errors.

Figure 3.7: Effect of Hearst exposure on long run attitudes



Note: This figure shows the effect of exposure to Hearst as either a child or an adult on attitudes towards government spending and redistribution. The blue bars plot $\hat{\alpha}_0$, and the lavender bars plot $\hat{\alpha}_1$ (both from Equation 5). Gray lines show 95% confidence intervals. "Support public healthcare" comes from ANES survey years 1970-1994, and "Disapprove of welfare" comes from 1976-1994 and is a combined measure of either wanting to reduce spending on welfare or feeling negatively towards those on welfare. All standard errors are clustered by county.

Table 3.5: Heterogeneity by presence of other media

	FDR vote share post-1932				
	(1) Baseline	(2) High radio presence	(3) Low radio presence	(4) High newspaper presence	(5) Low newspaper presence
Hearst newspaper circulation × post-1932	-0.214	0.030	-0.253	-0.039	-0.459
	(0.077)	(0.125)	(0.088)	(0.110)	(0.112)
Outcome mean	0.594	0.495	0.601	0.570	0.597
Observations	3,757	256	3,501	460	3,297

Note: This table shows heterogeneity in the effect of Hearst circulation on vote share for FDR based on the presence of other media in the county. In column 1, I show the baseline estimate. In columns 2 and 3, I divide the sample into counties with above and below median shares of radio ownership (with the median defined based on Hearst counties). In columns 4 and 5, I divide the sample into counties with an above and below median number of non-Hearst newspapers.

Bibliography

Abadie, Alberto. 2003. "Semiparametric instrumental variable estimation of treatment response models." *Journal of Econometrics*, 113(2): 231–263.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of Californias Tobacco Control Program." *Journal of the American Statistical Association*, 105(490): 493–505. Publisher: ASA Website _eprint: https://doi.org/10.1198/jasa.2009.ap08746.

Abadie, Alberto, and Javier Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review*, 93(1): 113–132.

Abed Faghri, Nahid M, Charles M Boisvert, and Sanaz Faghri. 2010. "Understanding the expanding role of primary care physicians (PCPs) to primary psychiatric care physicians (PPCPs): enhancing the assessment and treatment of psychiatric conditions." *Ment Health Fam Med*, 7(1): 17–25.

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya. 2015. "Radio and the Rise of The Nazis in Prewar Germany." *The Quarterly Journal of Economics*, 130(4): 1885–1939.

Alesina, Alberto, and Eliana La Ferrara. 2005. "Preferences for redistribution in the land of opportunities." *Journal of Public Economics*, 89(5): 897–931.

Alesina, Alberto, and George-Marios Angeletos. 2005. "Fairness and Redistribution." *American Economic Review*, 95(4): 960–980.

Alesina, Alberto, and Paola Giuliano. 2011. "Preferences for Redistribution." In *Handbook of Social Economics*. Vol. 1, , ed. Jess Benhabib, Alberto Bisin and Matthew O. Jackson, 93–131. North-Holland.

Alesina, Alberto, Edward Glaeser, and Bruce Sacerdote. 2001. "Why Doesn't the United States Have a European-Style Welfare State?" *Brookings Papers on Economic Activity*, 2001(2): 187–254. Publisher: Brookings Institution Press.

American National Election Survey. 2020. "Time Series Cumulative Data File (1948-2020)."

American Psychological Association (APA). 2014. "Briefing series on the role of psychology in health care: primary care."

American Psychological Association (APA). 2021. "Demand for mental health treatment continues to increase, say psychologists."

Angelucci, Manuela, and Daniel Bennett. 2022. "The Economic Impact of Depression Treatment in India: Evidence from Community-Based Provision of Pharmacotherapy." Job Market Paper.

Appleby, L., J. A. Dennehy, C. S. Thomas, E. B. Faragher, and G. Lewis. 1999. "Aftercare and clinical characteristics of people with mental illness who commit suicide: a case-control study." *Lancet*, 353(9162): 1397–1400.

Aubry, Tim, Paula Goering, Scott Veldhuizen, Carol E. Adair, Jimmy Bourque, Jino Distasio, Eric Latimer, Vicky Stergiopoulos, Julian Somers, David L. Streiner, and Sam Tsemberis. 2016. "A Multiple-City RCT of Housing First With Assertive Community Treatment for Homeless Canadians With Serious Mental Illness." *Psychiatr Serv*, 67(3): 275–281.

Basu, Anirban, Romina Kee, David Buchanan, and Laura S Sadowski. 2012. "Comparative Cost Analysis of Housing and Case Management Program for Chronically Ill Homeless Adults Compared to Usual Care." *Health Serv Res*, 47(1 Pt 2): 523–543.

Bazzi, Samuel, Martin Fiszbein, and Mesay Gebresilasse. 2020. "Frontier Culture: The Roots and Persistence of Rugged Individualism in the United States." *Econometrica*, 88(6): 2329–2368.

Beach, Brian, and W. Walker Hanlon. 2022. "Historical Newspaper Data: A Researchers Guide and Toolkit." *NBER Working Papers*.

Betancourt, Jose A., Diane M. Dolezel, Ramalingam Shanmugam, Gerardo J. Pacheco, Paula Stigler Granados, and Lawrence V. Fulton. 2023. "The Health Status of the US Veterans: A Longitudinal Analysis of Surveillance Data Prior to and during the COVID-19 Pandemic." *Healthcare (Basel)*, 11(14): 2049.

Bhat, Bhargav, Jonathan de Quidt, Johannes Haushofer, Vikram H. Patel, Gautam Rao, Frank Schilbach, and Pierre-Luc P. Vautrey. 2022. "The Long-Run Effects of Psychotherapy on Depression, Beliefs, and Economic Outcomes." NBER Working Paper 30011.

Bhuller, Manudeep, and Henrik Sigstad. 2024. "2SLS with multiple treatments." *Journal of Econometrics*, 242(1): 105785.

Biasi, Barbara, Michael S. Dahl, and Petra Moser. 2021. "Career Effects of Mental Health." NBER Working Paper 29031.

Bjorkenstam, Emma, Rickard Ljung, Bo Burstrom, Ellenor Mittendorfer-Rutz, Johan Hallqvist, and Gunilla Ringback Weitoft. 2012. "Quality of medical care and excess mortality in psychiatric patients–a nationwide register-based study in Sweden." *BMJ Open*, 2: e000778.

Blandhol, Christine, John Bonney, Magne Mogstad, and Alexander Torgovitsky. 2022. "When is TSLS Actually LATE?" NBER Working Paper.

Blattman, Christopher, Sebastian Chaskel, Julian C. Jamison, and Margaret Sheridan. 2022. "Cognitive Behavior Therapy Reduces Crime and Violence over 10 Years: Experimental Evidence." NBER Working Paper 30049.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2024. "Revisiting Event-Study Designs: Robust and Efficient Estimation." *The Review of Economic Studies*.

Bowersox, Nicholas W., Amy M. Kilbourne, Kristen M. Abraham, Brian H. Reck, Zongshan Lai, Amy S. B. Bohnert, David E. Goodrich, and Chester L. Davis. 2012. "Cause-specific mortality among Veterans with serious mental illness lost to follow-up." *Gen Hosp Psychiatry*, 34(6): 651–653.

Brounstein, Jakob, and John Wieselthier. 2024. "The labor market impacts of unconditional housing out of homelessness."

Callaway, Brantly, and Pedro H. C. SantAnna. 2021. "Difference-in-Differences with multiple time periods." *Journal of Econometrics*, 225(2): 200–230.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *The Review of Economics and Statistics*, 90(3): 414–427. Publisher: MIT Press.

Canay, Ivan A., Andres Santos, and Azeem M. Shaikh. 2021. "The Wild Bootstrap with a Small Number of Large Clusters." *The Review of Economics and Statistics*, 103(2): 346–363.

Caron, Christina. 2021. "Mental Health Providers Struggle to Meet Pandemic Demand." *The New York Times*.

Cavallo, Eduardo, Sebastian Galiani, Ilan Noy, and Juan Pantano. 2013. "Catastrophic Natural Disasters and Economic Growth." *The Review of Economics and Statistics*, 95(5): 1549–1561. Publisher: The MIT Press.

Chan, David, and Yiqun Chen. n.d.. "The Productivity of Professions: Evidence from the Emergency Department." *American Economic Review*.

Chan, David C., David Card, and Lowell Taylor. 2023. "Is There a VA Advantage? Evidence from Dually Eligible Veterans." *American Economic Review*, 113(11): 3003–3043.

Chapman, Alec B, Audrey Jones, A Taylor Kelley, Barbara Jones, Lori Gawron, Ann Elizabeth Montgomery, Thomas Byrne, Ying Suo, James Cook, Warren Pettey, Kelly Peterson, Makoto Jones, and Richard Nelson. 2021. "ReHouSED: A novel measurement of Veteran housing stability using natural language processing." *Journal of Biomedical Informatics*, 122: 103903.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review*, 106(4): 855–902.

Clubb, Jerome M., and Howard W. Allen. 1969. "The Cities and the Election of 1928: Partisan Realignment?" *The American Historical Review*, 74(4): 1205–1220. Publisher: [Oxford University Press, American Historical Association].

Clubb, Jerome M., William H. Flanigan, and Nancy H. Zingale. 1987. "Electoral Data for Counties in the United States: Presidential and Congressional Races, 1840-1972: Version 1."

Cohen, Elior. 2024. "Housing the Homeless: The Effect of Placing Single Adults Experiencing Homelessness in Housing Programs on Future Homelessness and Socioeconomic Outcomes." *American Economic Journal: Applied Economics*, 16(2): 130–175.

Conley, T. G. 1999. "GMM estimation with cross sectional dependence." *Journal of Econometrics*, 92(1): 1–45.

Cornes, Michelle, Robert W. Aldridge, Elizabeth Biswell, Richard Byng, Michael Clark, Graham Foster, James Fuller, Andrew Hayward, Nigel Hewett, Alan Kilmister, Jill Manthorpe, Joanne Neale, Michela Tinelli, and Martin Whiteford. 2021. "ICD-10 codes used to define underlying causes of death due to alcohol, drugs or suicide." In *Improving care transfers for homeless patients after hospital discharge: a realist evaluation*. NIHR Journals Library.

Costantini, Sydney. 2025. "How Do Mental Health Treatment Delays Impact Long-Term Mortality?" *American Economic Review*, 115(5): 1672–1707.

Cuellar, Alison Evans, Sara Markowitz, and Anne M. Libby. 2004. "Mental health and substance abuse treatment and juvenile crime." *J Ment Health Policy Econ*, 7(2): 59–68.

Cuijpers, Pim, Ioana A. Cristea, Eirini Karyotaki, Mirjam Reijnders, and Marcus J.H. Huibers. 2016. "How effective are cognitive behavior therapies for major depression and anxiety disorders?" *World Psychiatry*, 15(3): 245–258.

Cunningham, Mary, Devlon Hanson, Sarah Gillespie, Michael Pergamit, Alyse Oneto, Patrick Spauster, Tracey O'Brien, Liz Sweitzer, and Christine Velez. 2021. "Breaking the Homelessness-Jail Cycle with Housing First." Urban Institute.

Currie, Janet M., and W. Bentley Macleod. 2020. "Understanding doctor decision making: the case of Depression treatment." *Econometrica*, 88(3): 847.

Cutler, David M. 2002. "Equality, Efficiency, and Market Fundamentals: The Dynamics of International Medical-Care Reform." *Journal of Economic Literature*, 40(3): 881–906.

Das-Munshi, Jayati, Chin-Kuo Chang, Rina Dutta, Craig Morgan, James Nazroo, Robert Stewart, and Martin J Prince. 2017. "Ethnicity and excess mortality in severe mental illness: a cohort study." *Lancet Psychiatry*, 4(5): 389–399.

de Chaisemartin, Clement, and Xavier D'Haultfoeuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review*, 110(9): 2964–2996.

DellaVigna, Stefano, and Ethan Kaplan. 2007. "The Fox News Effect: Media Bias and Voting." *The Quarterly Journal of Economics*, 122(3): 1187–1234.

Deshpande, Manasi, and Yue Li. 2019. "Who Is Screened Out? Application Costs and the Targeting of Disability Programs." *American Economic Journal: Economic Policy*, 11(4): 213–248.

Djourelova, Milena. 2023. "Persuasion through Slanted Language: Evidence from the Media Coverage of Immigration." *American Economic Review*, 113(3): 800–835.

Dupas, Pascaline, Vivian Hoffmann, Michael Kremer, and Alix Peterson Zwane. 2016. "Targeting health subsidies through a nonprice mechanism: A randomized controlled trial in Kenya." *Science*, 353(6302): 889–895.

Durante, Ruben, and Brian Knight. 2012. "Partisan Control, Media Bias, and Viewer Responses: Evidence from Berlusconi's Italy." *Journal of the European Economic Association*, 10(3): 451–481. Publisher: Oxford University Press.

Elixhauser, A., C. Steiner, D. R. Harris, and R. M. Coffey. 1998. "Comorbidity measures for use with administrative data." *Med Care*, 36(1): 8–27.

Ellis, Ralph. 2022. "Veterans Suicides May Be Highly Undercounted, Study Says."

Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya. 2011. "Media and Political Persuasion: Evidence from Russia." *American Economic Review*, 101(7): 3253–3285.

Evans, William N., David C. Philips, and Krista J. Ruffini. 2019. "Reducing and Preventing Homelessness: A Review of the Evidence and Charting a Research Agenda."

Evans, William N., Sarah Kroeger, Caroline Palmer, and Emily Pohl. 2019. "Housing and Urban Development Veterans Affairs Supportive Housing Vouchers and Veterans Homelessness." *Am J Public Health*, 109(10): 1440–1445.

Ferrara, Andreas, Joung Yeob Ha, and Randall Walsh. 2022. "Using Digitized Newspapers to Refine Historical Measures: The Case of the Boll Weevil." *NBER Working Papers*.

Feyman, Yevgeniy, Daniel A. Asfaw, and Kevin N. Griffith. 2022. "Geographic Variation in Appointment Wait Times for US Military Veterans." *JAMA Netw Open*, 5(8): e2228783.

Feyman, Y., S. M. Figueroa, Y. Yuan, M. Price, A. Kabdiyeva, J. Nebeker, M. Ward, P. Shafer, S. D. Pizer, and K. L. Strombotne. 2022. "Effect of Mental Health Staffing Inputs on Suicide-Related Events." *Health Serv Res*.

Fischer, Ellen P., John F. McCarthy, Rosalinda V. Ignacio, Frederic C. Blow, Kristen L. Barry, Teresa J. Hudson, Richard R. Owen, and Marcia Valenstein. 2008. "Longitudinal patterns of health system retention among veterans with schizophrenia or bipolar disorder." *Community Ment Health J*, 44(5): 321–330.

Fishback, Price V., Shawn Kantor, and John Joseph Wallis. 2003. "Can the New Deal's three Rs be rehabilitated? A program-by-program, county-by-county analysis." *Explorations in Economic History*, 40(3): 278–307.

Gartlehner, Gerald, Gernot Wagner, Nina Matyas, Viktoria Titscher, Judith Greimel, Linda Lux, Bradley N. Gaynes, Meera Viswanathan, Sheila Patel, and Kathleen N. Lohr. 2017. "Pharmacological and non-pharmacological treatments for major depressive disorder: review of systematic reviews." *BMJ Open*, 7(6): e014912.

Gentzkow, Matthew, and Jesse M. Shapiro. 2010. "What Drives Media Slant? Evidence from U.S. Daily Newspapers." *Econometrica*, 78(1): 35–71.

Gentzkow, Matthew, Jesse M. Shapiro, and Michael Sinkinson. 2011*a*. "The Effect of Newspaper Entry and Exit on Electoral Politics." *American Economic Review*, 101(7): 2980–3018.

Gentzkow, Matthew, Jesse M. Shapiro, and Michael Sinkinson. 2011*b*. "United States Newspaper Panel, 1869-2004."

Gibbons, Robert D., Kwan Hur, and J. John Mann. 2017. "Suicide Rates and the Declining Psychiatric Hospital Bed Capacity in the United States." *JAMA Psychiatry*, 74(8): 849–850.

Godoy, Anna, Venke F. Haaland, Ingrid Huitfeldt, and Mark Votruba. n.d.. "Hospital Queues, Patient Health and Labor Supply." *American Economic Journal: Economic Policy*, forthcoming.

Gruber, Jonathan, Thomas Hoe, and George Stoye. 2021. "Saving Lives by Tying Hands: The Unexpected Effects of Constraining Health Care Providers." *The Review of Economics and Statistics*.

Gubits, Daniel, Marybeth Shinn, Michelle Wood, Scott R. Brown, Samuel R. Dastrup, and Stephen H. Bell. 2018. "What Interventions Work Best for Families Who Experience Homelessness? Impact Estimates from the Family Options Study." *J Policy Anal Manage*, 37(4): 735–766.

Haines, Michael R., and Inter-University Consortium For Political And Social Research. 2005. "Historical, Demographic, Economic, and Social Data: The United States, 1790-2002."

Hanson, Devlin, and Sarah Gillespie. 2024. "Housing First Increased Psychiatric Care Office Visits And Prescriptions While Reducing Emergency Visits." *Health Affairs*, 43(2): 209–217. Publisher: Health Affairs Publishing.

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

Heinesen, Eskil, Christian Hvid, Lars Johannessen Kirkeboen, Edwin Leuven, and Magne Mogstad. 2022. "Instrumental Variables with Unordered Treatments: Theory and Evidence from Returns to Fields of Study."

Hendren, Nathaniel, and Ben Sprung-Keyser. 2020. "A Unified Welfare Analysis of Government Policies." *The Quarterly Journal of Economics*, 135(3): 1209–1318.

Holliday, Stephanie. 2018. "The Relationship Between Mental Health Care Access and Suicide." RAND Corporation.

Holmes, Emily A., Ata Ghaderi, Catherine J. Harmer, Paul G. Ramchandani, Pim Cuijpers, Anthony P. Morrison, Jonathan P. Roiser, Claudi L. H. Bockting, Rory C. O'Connor, Roz Shafran, Michelle L. Moulds, and Michelle G. Craske. 2018. "The Lancet Psychiatry Commission on psychological treatments research in tomorrow's science." *The Lancet Psychiatry*, 5(3): 237–286. Publisher: Elsevier.

Hotton, Anna, Mary-Ellen Mackesy-Amiti, and Basmattee Boodram. 2021. "Trends in homelessness and injection practices among young urban and suburban people who inject drugs: 1997-2017." *Drug and Alcohol Dependence*, 225: 108797.

Hoxby, Caroline M. 2000. "The Effects of Class Size on Student Achievement: New Evidence from Population Variation." *Q J Econ*, 115(4): 1239–1285.

Humphries, John Eric, Aurelie Ouss, Kamelia Stavreva, Megan T Stevenson, and Winnie van Dijk. 2025. "Conviction, Incarceration, and Recidivism: Understanding the Revolving Door." *The Quarterly Journal of Economics*, 140(4): 2907–2962.

Kertesz, Stefan G, Kimberly Crouch, Jesse B Milby, Robert E Cusimano, and Joseph E Schumacher. 2009. "Housing First for Homeless Persons with Active Addiction: Are We Overreaching?" *Milbank Q*, 87(2): 495–534.

Kirkeboen, Lars J., Edwin Leuven, and Magne Mogstad. 2016. "Field of Study, Earnings, and Self-Selection." *The Quarterly Journal of Economics*, 131(3): 1057–1112. Publisher: Oxford University Press.

Kisely, Steve, Neil Preston, Jianguo Xiao, David Lawrence, Sandra Louise, and Elizabeth Crowe. 2013. "Reducing all-cause mortality among patients with psychiatric disorders: a population-based study." *CMAJ*, 185(1): E50–E56.

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

Klinkenberg, W. D., and R. J. Calsyn. 1996. "Predictors of receipt of aftercare and recidivism among persons with severe mental illness: a review." *Psychiatr Serv*, 47(5): 487–496.

Koh, Katherine A., Jill S. Roncarati, Melanie W. Racine, James J. O'Connell, and Jessie M. Gaeta. 2022. "Unsheltered vs. Sheltered Adults Experiencing Homelessness: Health Care Spending and Utilization." *J Gen Intern Med*, 37(8): 2100–2102.

Kreyenbuhl, Julie, Ilana R. Nossel, and Lisa B. Dixon. 2009. "Disengagement From Mental Health Treatment Among Individuals With Schizophrenia and Strategies for Facilitating Connections to Care: A Review of the Literature." *Schizophr Bull*, 35(4): 696–703.

Lang, Matthew. 2013. "The Impact of Mental Health Insurance Laws on State Suicide Rates." *Health Economics*, 22(1): 73–88.

Larcinese, Valentino, Riccardo Puglisi, and James M. Snyder. 2011. "Partisan bias in economic news: Evidence on the agenda-setting behavior of U.S. newspapers." *Journal of Public Economics*, 95(9): 1178–1189.

Larimer, Mary E., Daniel K. Malone, Michelle D. Garner, David C. Atkins, Bonnie Burlingham, Heather S. Lonczak, Kenneth Tanzer, Joshua Ginzler, Seema L. Clifasefi, William G. Hobson, and G. Alan Marlatt. 2009. "Health Care and Public Service Use and Costs Before and After Provision of Housing for Chronically Homeless Persons With Severe Alcohol Problems." *JAMA*, 301(13): 1349–1357.

Lawrence, David M., Cashel D'arcy J. Holman, Assen V. Jablensky, and Michael S. T. Hobbs. 2003. "Death rate from ischaemic heart disease in Western Australian psychiatric patients 1980-1998." *Br J Psychiatry*, 182: 31–36.

Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *The Review of Economic Studies*, 76(3): 1071–1102.

Lewer, Dan, Thomas D. Brothers, Naomi Van Hest, Matthew Hickman, Adam Holland, Prianka Padmanathan, and Paola Zaninotto. 2022. "Causes of death among people who used illicit opioids in England, 2001-18: a matched cohort study." *Lancet Public Health*, 7(2): e126–e135.

Library of Congress. 2011. "U.S. Newspaper Directory, 1690-Present."

Listokin, Siona, and Jason Snyder. 2010. "The Impact of Editorial Slant: Evidence from the Hearst Media Empire." UCLA Working Paper.

Liu, Nancy H., Gail L. Daumit, Tarun Dua, Ralph Aquila, Fiona Charlson, Pim Cuijpers, Benjamin Druss, Kenn Dudek, Melvyn Freeman, Chiyo Fujii, Wolfgang Gaebel, Ulrich Hegerl, Itzhak Levav, Thomas Munk Laursen, Hong Ma, Mario Maj, Maria Elena Medina Mora, Merete Nordentoft, Dorairaj Prabhakaran, Karen Pratt, Martin Prince, Thara Rangaswamy, David Shiers, Ezra Susser, Graham Thornicroft, Kristian Wahlbeck, Abe Fekadu Wassie, Harvey Whiteford, and Shekhar Saxena. 2017. "Excess mortality in persons with severe mental disorders: a multilevel intervention framework and priorities for clinical practice, policy and research agendas." *World Psychiatry*, 16(1): 30–40.

Marcus, Steven C., Megan E. Reilly, Kelly Zentgraf, Kevin G. Volpp, and Mark Olfson. 2021. "Effect of Escalating and Deescalating Financial Incentives vs Usual Care to Improve Antidepressant Adherence: A Pilot Randomized Clinical Trial." *JAMA Psychiatry*, 78(2): 222–224.

McBain, Ryan K., Jonathan H. Cantor, and Nicole K. Eberhart. 2022. "Estimating Psychiatric Bed Shortages in the US." *JAMA Psychiatry*, 79(4).

Meyer, Bruce D., Angela Wyse, and Ilina Logani. 2023. "Life and Death at the Margins of Society: The Mortality of the U.S. Homeless Population."

Montgomery, Ann Elizabeth, Lindsay L. Hill, Vincent Kane, and Dennis P. Culhane. 2013. "Housing Chronically Homeless Veterans: Evaluating the Efficacy of a Housing First Approach to Hud-Vash." *Journal of Community Psychology*, 41(4): 505–514. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcop.21554.

Mountjoy, Jack. 2022. "Community Colleges and Upward Mobility." *American Economic Review*, 112(8): 2580–2630.

Nasaw, David. 2000. *The Chief: The Life of William Randolph Hearst.* Houghton Mifflin Harcourt.

National Academies of Sciences, Engineering, and Medicine. 2018. *Timely Access to Mental Health Care.* National Academies Press (US).

National Academies of Sciences, Engineering, Health and Medicine Division, Board on Population Health and Public Health Practice, Policy and Global Affairs, Science and Technology for Sustainability Program, and Committee on an Evaluation of Permanent Supportive Housing Programs for Homeless Individuals. 2018. "Evidence of Effect of Permanent Supportive Housing on Health." In *Permanent Supportive Housing: Evaluating the Evidence for Improving Health Outcomes Among People Experiencing Chronic Homelessness*. National Academies Press (US).

National Committee for Quality Assurance (NCQA). n.d.. "Follow-Up After Emergency Department Visit for Mental Illness."

National Homeless Information Project. 2018. "Veterans Data: National Homeless Information Project."

National Housing Law Project. 2008. "HUD-VASH: Long-Neglected Program Brought Back To Life." 38.

Newcomer, John W., and Charles H. Hennekens. 2007. "Severe Mental Illness and Risk of Cardiovascular Disease." *JAMA*, 298(15): 1794–1796.

Nishimura, Marin, Harpreet Bhatia, Janet Ma, Stephen D. Dickson, Laith Alshawabkeh, Eric Adler, Alan Maisel, Michael H. Criqui, Barry Greenberg, and Isac C. Thomas. 2020. "The Impact of Substance Abuse on Heart Failure Hospitalizations." *Am J Med*, 133(2): 207–213.e1.

Nose, M., C. Barbui, and M. Tansella. 2003. "How often do patients with psychosis fail to adhere to treatment programmes? A systematic review." *Psychol Med*, 33(7): 1149–1160.

Ottinger, Sebastian, and Max Winkler. 2022. "The Political Economy of Propaganda: Evidence from US Newspapers." *IZA Discussion Papers*.

Porter, Justin. 2024. "The Evening: Record Homelessness in the U.S." *The New York Times*.

Raphelson, Samantha. 2017. "How The Loss Of U.S. Psychiatric Hospitals Led To A Mental Health Crisis." *NPR*.

Raven, Maria C., Matthew J. Niedzwiecki, and Margot Kushel. 2020. "A randomized trial of permanent supportive housing for chronically homeless persons with high use of publicly funded services." *Health Serv Res*, 55(Suppl 2): 797–806.

Reichert, Anika, and Rowena Jacobs. 2018. "The impact of waiting time on patient outcomes: Evidence from early intervention in psychosis services in England." *Health Econ*, 27(11): 1772–1787.

Reif, Sharon, Preethy George, Lisa Braude, Richard H. Dougherty, Allen S. Daniels, Sushmita Shoma Ghose, and Miriam E. Delphin-Rittmon. 2014. "Residential Treatment for Individuals With Substance Use Disorders: Assessing the Evidence." *Psychiatric Services*, 65(3): 301–312. Publisher: American Psychiatric Publishing.

Roosevelt, Franklin D. 1992. *FDR's fireside chats.* Norman : University of Oklahoma Press.

Rosenheck, Robert, Wesley Kasprow, Linda Frisman, and Wen Liu-Mares. 2003. "Cost-effectiveness of supported housing for homeless persons with mental illness." *Arch Gen Psychiatry*, 60(9): 940–951.

Roy, A. D. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers*, 3(2): 135–146. Publisher: Oxford University Press.

Sabin, Arthur J. 1999. *Red Scare in Court: New York Versus the International Workers Order.* University of Pennsylvania Press.

Salit, Sharon A., Evelyn M. Kuhn, Arthur J. Hartz, Jade M. Vu, and Andrew L. Mosso. 1998. "Hospitalization Costs Associated with Homelessness in New York City." *New England Journal of Medicine*, 338(24): 1734–1740. Publisher: Massachusetts Medical Society _eprint: https://www.nejm.org/doi/pdf/10.1056/NEJM199806113382406.

Schrecker, Ellen. 2021. "William Randolph Hearst and McCarthyism." PBS.

Sharp, D. J., and W. Hamilton. 2001. "Non-attendance at general practices and outpatient clinics." *BMJ*, 323(7321): 1081–1082.

Silver, David, and Jonathan Zhang. 2026. "Invisible Wounds: How Mental Disability Benefits Shape Veteran Well-Being." *American Economic Journal: Economic Policy*, 18(1): 282–317.

Singh, Manasvini, and Atheendar Venkataramani. 2022. "Capacity Strain and Racial Disparities in Hospital Mortality." NBER Working Paper 30380.

Siskind, Dan, Meredith Harris, Steve Kisely, Victor Siskind, James Brogan, Jane Pirkis, David Crompton, and Harvey Whiteford. 2014. "A Retrospective Quasi-Experimental Study of a Transitional Housing Program for Patients with Severe and Persistent Mental Illness." *Community Ment Health J*, 50(5): 538–547.

Snowdon, John, and Namkee G. Choi. 2020. "Undercounting of suicides: Where suicide data lie hidden." *Glob Public Health*, 15(12): 1894–1901.

Stergiopoulos, Vicky, Stephen W. Hwang, Agnes Gozdzik, Rosane Nisenbaum, Eric Latimer, Daniel Rabouin, Carol E. Adair, Jimmy Bourque, Jo Connelly, James Frankish, Laurence Y. Katz, Kate Mason, Vachan Misir, Kristen O'Brien, Jitender Sareen, Christian G. Schutz, Arielle Singer, David L. Streiner, Helen-Maria Vasiliadis, Paula N. Goering, and At

Home/Chez Soi Investigators. 2015. "Effect of scattered-site housing using rent supplements and intensive case management on housing stability among homeless adults with mental illness: a randomized trial." *JAMA*, 313(9): 905–915.

Treisman, Rachel. 2025. "'Broadcasting' has its roots in agriculture. Here's how it made its way into media." *NPR*.

Trump, Donald. 2025. "Executive Order: Ending Crime and Disorder on America's Streets."

Tsai, Jack, and Robert A. Rosenheck. 2015. "Risk factors for homelessness among US veterans." *Epidemiol Rev*, 37: 177–195.

Tsemberis, Sam, Leyla Gulcur, and Maria Nakae. 2004. "Housing First, Consumer Choice, and Harm Reduction for Homeless Individuals With a Dual Diagnosis." *Am J Public Health*, 94(4): 651–656.

Tsemberis, S., and R. F. Eisenberg. 2000. "Pathways to housing: supported housing for street-dwelling homeless individuals with psychiatric disabilities." *Psychiatr Serv*, 51(4): 487–493.

Tuck, Jim. 1995. *McCarthyism and New York's Hearst Press: A Study of Roles in the Witch Hunt.* University Press of America.

Turner, Alex J, Igor Francetic, Ruth Watkinson, Stephanie Gillibrand, and Matt Sutton. 2022. "Socioeconomic inequality in access to timely and appropriate care in emergency departments." *Journal of Health Economics*, 85: 102668.

Unemployment rate - Civilian labor force 14 yrs. & over. n.d..

US Department of Housing and Urban Development (HUD). 2010. "Costs Associated with First Time Homelessness for Families and Individuals."

US Department of Housing and Urban Development (HUD). 2024. "The 2024 Annual Homelessness Assessment Report (AHAR) to Congress."

Vannier, Augustin G. L., Jessica E. S. Shay, Vladislav Fomin, Suraj J. Patel, Esperance Schaefer, Russell P. Goodman, and Jay Luther. 2022. "Incidence and Progression of Alcohol-Associated Liver Disease After Medical Therapy for Alcohol Use Disorder." *JAMA Network Open*, 5(5): e2213014.

VA Office of Mental Health and Suicide Prevention. 2022. "National Veteran Suicide Prevention: Annual Report." Department of Veterans Affairs.

Veterans Health Administration. 2008. "VHA Directive 2008-036." Department of Veterans Affairs.

Wang, Tianyi. 2021. "Media, Pulpit, and Populist Persuasion: Evidence from Father Coughlin." *American Economic Review*, 111(9): 3064–3092.

Weiner, Stacy. 2022. "A growing psychiatrist shortage and an enormous demand for mental health services."

Williams, Jenny, and Anne Bretteville-Jensen. 2022. "Whats Another Day? The Effects of Wait Time for Substance Abuse Treatment on Health-Care Utilization, Employment and Crime." *Institute of Labor Economics*.

Xu, Wendy Yi, Chi Song, Yiting Li, and Sheldon Michael Retchin. 2019. "Cost-Sharing Disparities for Out-of-Network Care for Adults With Behavioral Health Conditions." *JAMA Network Open*, 2(11): e1914554.

Yamokoski, Cynthia. 2020. "Massed Delivery of Evidence-Based Psychotherapy for PTSD." Department of Veterans Affairs.

Appendix A

Appendix to Chapter 1

A.1 Visualizing treatment shares under instrument additivity

Figure A.3 visualizes treatment shares and the additivity condition in simplex space for the case of simple binary instruments $(Z_{v,i}, Z_{t,i}) \in \{0,1\}^2$. Let $\mathbf{p}(z_v, z_t) = (p_v(z_v, z_t), p_t(z_v, z_t), p_0(z_v, z_t))'$, a point on the simplex, represent the shares of individuals in v, t, and 0 based on the realized values of the instruments. For instance, $\mathbf{p}(0,0)$ represents treatment shares when both instruments are off. Further, let $\Delta\mathbf{p}(z_v, z_t) = \mathbf{p}(z_v, z_t) - \mathbf{p}(0,0)$, or the change in treatment shares from $(Z_{v,i}, Z_{t,i}) = (0,0)$ to (z_v, z_t), represented by an arrow on the simplex. Under this notation, I can rewrite Eq. 3 as follows:

$$\Delta\mathbf{p}(1,1) = \Delta\mathbf{p}(1,0) + \Delta\mathbf{p}(0,1).$$

Under additivity, the four points $\mathbf{p}(0,0), \mathbf{p}(1,0), \mathbf{p}(0,1)$, and $\mathbf{p}(1,1)$ must form a parallelogram — that is, the change in treatment shares induced by $Z_{v,i}$ is the same regardless of the value of $Z_{t,i}$, and vice versa. This geometric condition is equivalent to requiring that the two instruments shift disjoint complier groups (see Proposition 1).

A.2 Understanding complier types: detail

A.2.1 Proof of Proposition 1

Proof. Let:

$$\begin{aligned}\Pr[D_{k,i} = 1 | Z_v = 1, Z_t = 1] - \Pr[D_{k,i} = 1 | Z_v = 1, Z_t = 0]& \\ = \Pr[D_{k,i} = 1 | Z_v = 0, Z_t = 1] - \Pr[D_{k,i} = 1 | Z_v = 0, Z_t = 0]&\end{aligned}$$

$\forall k \in \{0, v, t\}$. Adding and subtracting terms in this expression, the following equalities all must hold (see Table A.2):

$$\underbrace{\Pr[D_0 = 1 | Z_v = 1, Z_t = 0] - \Pr[D_0 = 1 | Z_v = 0, Z_t = 0]}_{-(P(C_0^v) + P(C_0^{v+,t-}) + P(C_0^{v-,t+}))} = \underbrace{\Pr[D_0 = 1 | Z_v = 1, Z_t = 1] - \Pr[D_0 = 1 | Z_v = 0, Z_t = 1]}_{-P(C_0^v)}$$

$$\underbrace{\Pr[D_v=1|Z_v=0,Z_t=1]-P[D_v|Z_v=0,Z_t=0]}_{-P(C_v^t)}=\underbrace{\Pr[D_v=1|Z_v=1,Z_t=1]-\Pr[D_v=1|Z_v=1,Z_t=0]}_{-(P(C_0^{v-,t+})+P(C_v^t)+P(C_t^{v-}))}$$

$$\underbrace{\Pr[D_t=1|Z_v=1,Z_t=0]-\Pr[D_t=1|Z_v=0,Z_t=0]}_{-(P(C_t^v)+P(C_t^{v-}))}=\underbrace{\Pr[D_t=1|Z_v=1,Z_t=1]-\Pr[D_t=1|Z_v=0,Z_t=1]}_{-(P(C_0^{v+,t-})+P(C_t^v))}$$

These equations imply the following equalities hold for complier shares:

$$P(C_0^v)+P(C_0^{v+,t-})+P(C_0^{v-,t+})=P(C_0^v)$$

$$P(C_v^t)=P(C_0^{v-,t+})+P(C_v^t)+P(C_t^{v-})$$

$$P(C_t^v)+P(C_t^{v-})=P(C_0^{v+,t-})+P(C_t^v)$$

These three equalities can only hold if $P(C_0^{v+,t-})=P(C_0^{v-,t+})=P(C_t^{v-})=0$. The remaining complier groups are then point identified from propensity scores as:

$$P(C_0^v)=-(\Pr[D_0=1|Z_v=1,Z_t=0]-\Pr[D_0=1|Z_v=0,Z_t=0])$$

$$P(C_0^t)=-(\Pr[D_0=1|Z_v=0,Z_t=1]-\Pr[D_0=1|Z_v=0,Z_t=0])$$

$$P(C_t^v)=-(\Pr[D_t=1|Z_v=1,Z_t=0]-\Pr[D_t=1|Z_v=0,Z_t=0])$$

$$P(C_v^t)=-(\Pr[D_v=1|Z_v=0,Z_t=1]-\Pr[D_v=1|Z_v=0,Z_t=0])$$

□

A.2.2 Empirical complier shares

Following Proposition 1, to empirically estimate complier shares, I run the following first stage regressions using a binary version of $Z_{v,i}$:

$$D_{0,i}=\alpha_v Z_{v,i}+\alpha_t Z_{t,i}+\alpha_{vt} Z_{v,i}Z_{t,i}+\zeta_{0,h(i)y(i)}+\phi_{0,h(i)m(i)}+X_i\beta_0+\epsilon_{0,i} \tag{A.1}$$

$$D_{v,i}=\pi_v Z_{v,i}+\pi_t Z_{t,i}+\pi_{vt} Z_{v,i}Z_{t,i}+\zeta_{1,h(i)y(i)}+\phi_{1,h(i)m(i)}+X_i\beta_1+\epsilon_{1,i} \tag{A.2}$$

$$D_{t,i}=\psi_v Z_{v,i}+\psi_t Z_{t,i}+\psi_{vt} Z_{v,i}Z_{t,i}+\zeta_{2,h(i)y(i)}+\phi_{2,h(i)m(i)}+X_i\beta_2+\epsilon_{2,i} \tag{A.3}$$

where $D_{0,i}$, $D_{v,i}$, and $D_{t,i}$ are indicators for receiving no intervention, the voucher program, and transitional housing, respectively. In practice, let discrete $Z_{v,i} = 1$ if (residualized) voucher availability is above the 90th percentile. Conditional on the baseline controls, these coefficients yield various complier share mixtures, as described in Table A.2.

In Table A.3, I show coefficient estimates. In all three regressions, the interaction term coefficient is close to 0 and statistically insignificant, implying that the joint values of $Z_{v,i}, Z_{t,i}$ do not affect treatment choice (see proof in Appendix A.2.1). In other words, $P(C_0^{v+,t-}) = P(C_0^{v-,t+}) = P(C_t^{v-}) = 0$. This leaves four complier groups that are actually present empirically: C_0^v, C_0^t, C_v^t,and C_t^v. The shares of these groups are given by $-\alpha_v, -\alpha_t, -\pi_t$, and $-\psi_v$, respectively, corresponding to the equivalent propensity score differences in Proposition 1.

A.3 2SLS bounds with multiple instruments

A.3.1 Decomposing 2SLS

Below, I decompose the 2SLS coefficient, β_v, into a well-defined LATE and selection terms. This proof follows closely those of Heinesen et al. (2022) and Mountjoy (2022). For simplicity, I ignore control variables X_i and consider the case of two binary instruments, $Z_{v,i}$ and $Z_{t,i}$.

First, following the notation in Appendix A2.2:

$$E[D_{v,i}|\mathbf{Z}_i] = \pi_0 + \pi_v Z_{v,i} + \pi_t Z_{t,i} \tag{A.4}$$

$$E[D_{t,i}|\mathbf{Z}_i] = \psi_0 + \psi_v Z_{v,i} + \psi_t Z_{t,i} \tag{A.5}$$

Similarly, for Y_i:

$$E[Y_i|\mathbf{Z}_i] = \gamma + \beta_v \hat{D}_{v,i} + \beta_t \hat{D}_{t,i} \tag{A.6}$$

where β_v is the 2SLS coefficient for the effect of housing first, and $\hat{D}_{v,i}$ and $\hat{D}_{t,i}$ are predicted from first stage models.

Plugging in Eq. A4 and A5:

$$E[Y_i|\mathbf{Z}_i] = \gamma + \beta_v(\pi_0 + \pi_v Z_{v,i} + \pi_t Z_{t,i}) + \beta_t(\psi_0 + \psi_v Z_{v,i} + \psi_t Z_{t,i})$$

$$= (\gamma + \pi_0\beta_v + \psi_0\beta_t) + \underbrace{(\beta_v\pi_v + \beta_t\psi_v)}_{\zeta_v} Z_{v,i} + \underbrace{(\beta_v\pi_t + \beta_t\psi_t)}_{\zeta_t} Z_{t,i}$$

I can then solve for β_v as a function of $\pi_v, \pi_t, \psi_v, \psi_t, \alpha_v$, and α_t as:

$$\beta_v = \frac{\psi_t\zeta_v - \psi_v\zeta_t}{\pi_v\psi_t - \psi_v\pi_t} \tag{A.7}$$

Now, reduced form coefficients ζ_v and ζ_t can be represented as a weighted sum of LATEs and complier shares, as follows:

$$\zeta_v = E[Y_{v,i} - Y_{0,i}|C_0^v]P(C_0^v) + E[Y_{v,i} - Y_{t,i}|C_t^v]P(C_t^v) \tag{A.8}$$

as $Z_{v,i}$ induces C_0^v compliers from 0 to v and C_t^v compliers from t to v. Similarly,

$$\zeta_t = E[Y_{t,i} - Y_{0,i}|C_0^t]P(C_0^t) + E[Y_{t,i} - Y_{v,i}|C_v^t]P(C_v^t) \tag{A.9}$$

Plugging these definitions into Eq. A7:

$$\beta_v = \frac{\psi_t(E[Y_{v,i}-Y_{0,i}|C_0^v]P(C_0^v)+E[Y_v-Y_t|C_t^v]P(C_t^v))-\psi_v(E[Y_{t,i}-Y_{0,i}|C_0^t]P(C_0^t)+E[Y_t-Y_v|C_v^t]P(C_v^t))}{\pi_v\psi_t-\psi_v\pi_t}$$

Based on Table A.2, we also know that $\pi_v = P(C_0^v) + P(C_t^v)$, $\pi_t = -P(C_v^t)$, $\psi_v = -P(C_t^v)$, and $\psi_t = P(C_0^t) + P(C_v^t)$. Plugging these in:

$$\beta_v = \frac{(P(C_0^t)+P(C_v^t))(E[Y_{v,i}-Y_{0,i}|C_0^v]P(C_0^v)+E[Y_v-Y_t|C_t^v]P(C_t^v))+P(C_t^v)(E[Y_{t,i}-Y_{0,i}|C_0^t]P(C_0^t)+E[Y_t-Y_v|C_v^t]P(C_v^t))}{\underbrace{(P(C_0^v)+P(C_t^v))(P(C_0^t)+P(C_v^t))-P(C_v^t)P(C_t^v)}_{W}}$$

We can interpret $E[Y_{v,i} - Y_{0,i}|C_0^v]$ as the local average treatment effect of housing first (compared to no intervention) for C_0^v compliers. To isolate this term, add and subtract $\frac{W}{W}E[Y_{v,i}-Y_{0,i}|C_0^v]$:

$$\beta_v = E[Y_{v,i} - Y_{0,i}|C_0^v] + \underbrace{\frac{P(C_t^v)P(C_0^t)}{W}}_{A}(E[Y_{v,i} - Y_{t,i}|C_t^v] - E[Y_{v,i} - Y_{0,i}|C_0^v] + E[Y_{t,i} - Y_{0,i}|C_0^t])$$

$$+\underbrace{\frac{P(C_t^v)P(C_v^t)}{W}}_{B}(E[Y_{v,i} - Y_{t,i}|C_t^v] + E[Y_{t,i} - Y_{v,i}|C_v^t])$$

Re-expressing the above in terms of a LATE term and selection terms, I have:

$$\beta_v = \underbrace{E[Y_{v,i} - Y_{0,i}|C_0^v]}_{\text{LATE}_v} + A(E[Y_{v,i} - Y_{0,i}|C_t^v] - E[Y_{v,i} - Y_{0,i}|C_0^v]) + A(E[Y_{t,i} - Y_{0,i}|C_0^t] - E[Y_{t,i} - Y_{0,i}|C_t^v]) \tag{A.10}$$

$$+B(E[Y_{v,i} - Y_{0,i}|C_t^v] - E[Y_{v,i} - Y_{0,i}|C_v^t]) + B(E[Y_{t,i} - Y_{0,i}|C_v^t] - E[Y_{t,i} - Y_{0,i}|C_t^v])$$

A.3.2 Defining worst case scenario bounds

Returning to Eq. 11 in the main text, assume $|E[Y_{k,i} - Y_{0,i}|C_i = c] - E[Y_{k,i} - Y_{0,i}|C_i = c']| < \epsilon$ for all complier groups c. Therefore, if all average treatment effects for complier groups differ by at most ϵ, the difference between the true LATE and β_v can be characterized as:

$$|\text{LATE}_v - \beta_v| \leq 2A\epsilon + 2B\epsilon$$

This equation implies I can define bounds on the true LATE as $(\beta_v - 2A\epsilon - 2B\epsilon, \beta_v + 2A\epsilon + 2B\epsilon)$.

A.4 Housing status over time: complier means

To gain a fuller picture of housing status over time across complier groups, I follow the logic of Eq. 9 in the main text to generate complier potential outcomes. I run the following regressions:

$$H_{q,i}D_{0,i} = \sigma_q D_{0,i} + \alpha_q Z_{t,i} + \zeta_{8,s(i)y(i)} + \phi_{8,h(i)m(i)} + X_i\beta_8 + \epsilon_{8,i} \tag{A.11}$$

$$H_{q,i}D_{0,i} = \rho_q D_{0,i} + \gamma_q Z_{v,i} + \zeta_{9,h(i)y(i)} + \phi_{9,h(i)m(i)} + X_i\beta_9 + \epsilon_{9,i} \tag{A.12}$$

where I instrument for $D_{0,i}$ with $Z_{v,i}$ in Eq. A11 and with $Z_{t,i}$ in Eq. A12. The coefficients σ_q and ρ_q give the mean housing rate in quarter q, $\bar{H}_q$, for C_0^v and C_0^t compliers who receive no intervention, respectively. Then, to calculate mean housing rates for these compliers when they recieve a voucher or transitional housing, I add each respective treatment effect (as shown in Figure 1.4) to the means under no intervention.[1] I repeat this analysis for unknown housing status, replacing the outcome with $U_{q,i}D_{0,i}$. Finally, I compute the percent of compliers who are unstably housed in a given quarter as $1 - \bar{H}_q - \bar{U}_q$.

I show means over time for C_0^v and C_0^t complier groups in Figure A.8. Echoing results from Figure 1.4, C_0^t compliers who enroll in transitional housing have the highest rates of stable housing in quarter 1 (days 0-90), as they can immediately enter transitional housing units. Meanwhile, C_0^v compliers who receive a HUD-VASH voucher have high rates of unstable housing in quarter 1, likely reflecting them searching for an apartment that accepts their voucher. Nevertheless, these compliers have the highest long term rates of stable housing. Compliers who recieve no intervention have very high rates of having unknown housing status, consistent with them dropping out of the VA homelessness support system. While some of these individuals likely find their own independent housing and thus may not need social worker services, others may be at particularly high risk of mortality.

[1] It is not possible to separately identify these means using complier characteristic regressions alone, as running regressions with $H_{q,i}D_{v,i}$ as the left hand side variable would yield a mixture between C_0^v and C_t^v compliers, as both of these complier groups are induced into v by Z_v (see Table A.2 and Mountjoy 2022). However, following the logic of Eq. 9, Eq. A11 and A12 yield untreated potential outcome means.

Table A.1: Unhoused Veteran characteristics compared to US unhoused population

	HUD statistics	Veteran sample
Veterans	0.04	1.00
Children	0.19	0.00
Male	0.60	0.90
Black	0.32	0.36
Hispanic	0.31	0.07

Note: This table compares the characteristics of Veterans in my sample to HUD statistics for the US unhoused population in 2024.

Table A.2: First stage coefficients and propensity scores interpretation

Coefficient	Interpretation	Complier shares
α_v	$\Pr[D_0 = 1 \mid Z_v = 1, Z_t = 0] - \Pr[D_0 = 1 \mid Z_v = 0, Z_t = 0]$	$-(P(C_0^v) + P(C_0^{v+,t-}) + P(C_0^{v-,t+}))$
α_t	$\Pr[D_0 = 1 \mid Z_v = 0, Z_t = 1] - \Pr[D_0 = 1 \mid Z_v = 0, Z_t = 0]$	$-(P(C_0^t) + P(C_0^{v+,t-}) + P(C_0^{v-,t+}))$
$\alpha_v + \alpha_{vt}$	$\Pr[D_0 = 1 \mid Z_v = 1, Z_t = 1] - \Pr[D_0 = 1 \mid Z_v = 0, Z_t = 1]$	$-P(C_0^v)$
$\alpha_t + \alpha_{vt}$	$\Pr[D_0 = 1 \mid Z_v = 1, Z_t = 1] - \Pr[D_0 = 1 \mid Z_v = 1, Z_t = 0]$	$-P(C_0^t)$
π_v	$\Pr[D_v = 1 \mid Z_v = 1, Z_t = 0] - \Pr[D_v = 1 \mid Z_v = 0, Z_t = 0]$	$P(C_0^v) + P(C_0^{v+,t-}) + P(C_0^{v-,t+}) + P(C_t^v) + P(C_t^{v-})$
π_t	$\Pr[D_v = 1 \mid Z_v = 0, Z_t = 1] - \Pr[D_v = 1 \mid Z_v = 0, Z_t = 0]$	$-P(C_v^t)$
$\pi_v + \pi_{vt}$	$\Pr[D_v = 1 \mid Z_v = 1, Z_t = 1] - \Pr[D_v = 1 \mid Z_v = 0, Z_t = 1]$	$P(C_0^v) + P(C_0^{v+,t-}) + P(C_t^v)$
$\pi_t + \pi_{vt}$	$\Pr[D_v = 1 \mid Z_v = 1, Z_t = 1] - \Pr[D_v = 1 \mid Z_v = 1, Z_t = 0]$	$-(P(C_0^{v-,t+}) + P(C_v^t) + P(C_t^{v-}))$
ψ_v	$\Pr[D_t = 1 \mid Z_v = 1, Z_t = 0] - \Pr[D_t = 1 \mid Z_v = 0, Z_t = 0]$	$-(P(C_t^v) + P(C_t^{v-}))$
ψ_t	$\Pr[D_t = 1 \mid Z_v = 0, Z_t = 1] - \Pr[D_t = 1 \mid Z_v = 0, Z_t = 0]$	$P(C_0^t) + P(C_0^{v+,t-}) + P(C_0^{v-,t+}) + P(C_v^t)$
$\psi_v + \psi_{vt}$	$\Pr[D_t = 1 \mid Z_v = 1, Z_t = 1] - \Pr[D_t = 1 \mid Z_v = 0, Z_t = 1]$	$-(P(C_0^{v+,t-}) + P(C_t^v))$
$\psi_t + \psi_{vt}$	$\Pr[D_t = 1 \mid Z_v = 1, Z_t = 1] - \Pr[D_t = 1 \mid Z_v = 1, Z_t = 0]$	$P(C_0^t) + P(C_0^{v-,t+}) + P(C_v^t) + P(C_t^{v-})$

Note: This table interprets the coefficients from saturated first stage regressions (Eqs. A1, A2, and A3) and maps propensity scores to complier shares. Column 2 interprets the coefficient in terms of propensity scores, and column 3 shows what complier shares these propensity scores correspond to. Complier groups are defined in Table 1.2.

Table A.3: First stage coefficient estimates

Coefficient	Estimate
α_v	-0.052 (0.005)
α_t	-0.075 (0.003)
α_{vt}	-0.009 (0.007)
π_v	0.062 (0.005)
π_t	-0.014 (0.002)
π_{vt}	0.008 (0.007)
ψ_v	-0.010 (0.003)
ψ_t	0.089 (0.002)
ψ_{vt}	0.001 (0.005)

Note: This figure shows the results of the saturated first stage regressions (Eqs. A1 - A3). Standard errors are shown in parentheses. Given that all interaction terms are close to 0 and statistically insignificant, under my treatment choice model, the shares of C_0^v, C_0^t, C_v^t, and C_t^v compliers are given by $-\alpha_v$, $-\alpha_t$, $-\pi_t$, and $-\psi_v$, respectively.

Table A.4: Robustness related to switching between treatment statuses

	First stage		3-year mortality		
	Voucher (Baseline)	Voucher in 12 months	Baseline	Treatment: voucher in 12 months	Drop those dual-enrolled
Voucher			-0.046	-0.046	-0.049
			(0.017)	(0.018)	(0.017)
Transitional housing			-0.007	0.001	-0.014
			(0.015)	(0.014)	(0.017)
Perc. receiving voucher ($Z_{v,i}$)	0.510	0.445			
	(0.017)	(0.019)			
Tr. liaison available ($Z_{t,i}$)	-0.013	-0.003			
	(0.002)	(0.002)			
	0.118	0.209	0.066	0.066	0.067
Observations	289,932	289,932	289,932	289,932	263,559

Note: This table shows robustness checks related to individuals switching between treatment statuses across time. In column 1, I show the baseline first stage results when $D_{v,i}$ – or initial voucher program enrollment – is the dependent variable. In column 2, I show the first stage when I re-define $D_{v,i}$ as receiving a housing voucher within 12 months of intake. In column 3, I show the baseline 2SLS results. In column 4, I show 2SLS results when I re-define $D_{v,i}$ as receiving a housing voucher within 12 months. In column 5, I drop those who initially enroll in 0 or tr who also receive a housing voucher within a year.

Table A.5: Comparison of effects to other US studies

	(1) Hanson and Gillespie	(2) Raven	(3) Primary Sample	(4) Primary Sample	(5) Prior Jail Sample	(6) Prior Jail Sample
ITT effect	2.0	3.9				
	(1.8)	(3.6)				
TOT effect	2.4	7.7	-0.046	-0.028	-0.043	-0.011
	(2.2)	(7.3)	(0.017)	(0.015)	(0.032)	(0.027)
Time horizon	2 years	2 years*	3 years	2 years	3 years	2 years
Observations	721	423	289,932	289,932	123,371	123,371
p-value (vs. Hanson and Gillespie)				0.051		0.31
p-value (vs. Raven)				0.16		0.26

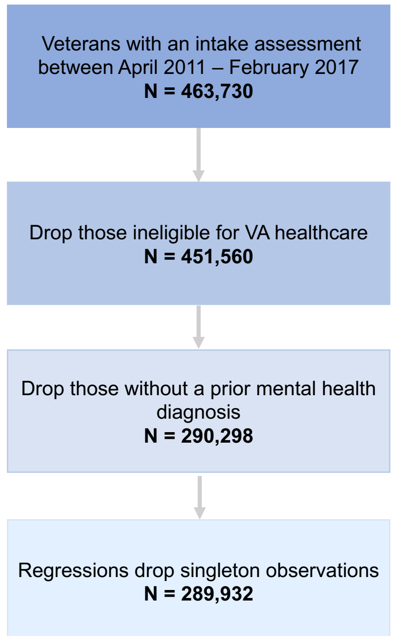
Note: This table compares my mortality effects with those of Hanson and Gillespie (2024) (column 1) and Raven (2020) (column 2). Column 3 shows the results for the primary sample, and column 4 shows the results for the primary sample at a 2 year time horizon. Column 5 shows results for an alternative sample of individuals seeking homelessness services who had been incarcerated for more than one month in their lifetime. Column 6 shows results for this sample at a 2-year horizon. TOT standard errors are approximated as $\hat{SE}_{ITT}/\hat{\pi}$, treating the first stage as fixed. For two year mortality estimates, I show p-values of the difference between the coefficients I estimate and those of Hanson and Gillespie (2024) and Raven (2020). *In Raven (2020), individuals are followed anywhere from 0 to 4 years, so I consider the average time horizon.

Table A.6: Heterogeneity by low demand/bridge transitional housing

	(1) Baseline	(2) Above median low demand/bridge	(3) Below median low demand/bridge
Trans. housing	-0.007	-0.020	0.002
	(0.015)	(0.024)	(0.019)
Dep. var. mean	0.066	0.068	0.063
Observations	289,932	138,527	151,405

Note: This table shows the effect of transitional housing on mortality based on whether GPD beds are classified as low demand and/or bridge. VISNs are classified into above and below median based on the percent of their GPD beds that are low demand and/or bridge. Column 1 shows the baseline IV estimate, column 2 shows the estimate for above median VISNs, and column 3 shows the estimate for below median VISNs. Baseline controls include VA station $\times$ year, VA station $\times$ month, and indicators for having a prior inpatient stay related to homelessness or mental health. Standard errors for all regressions are clustered by VA station $\times$ year $\times$ month.

Figure A.1: Sample selection



Note: This figure shows the sample selection process.

Figure A.2: Change in treatment status given z_v, z_t for 7 complier groups

A: Treatment status given z_v, z_t

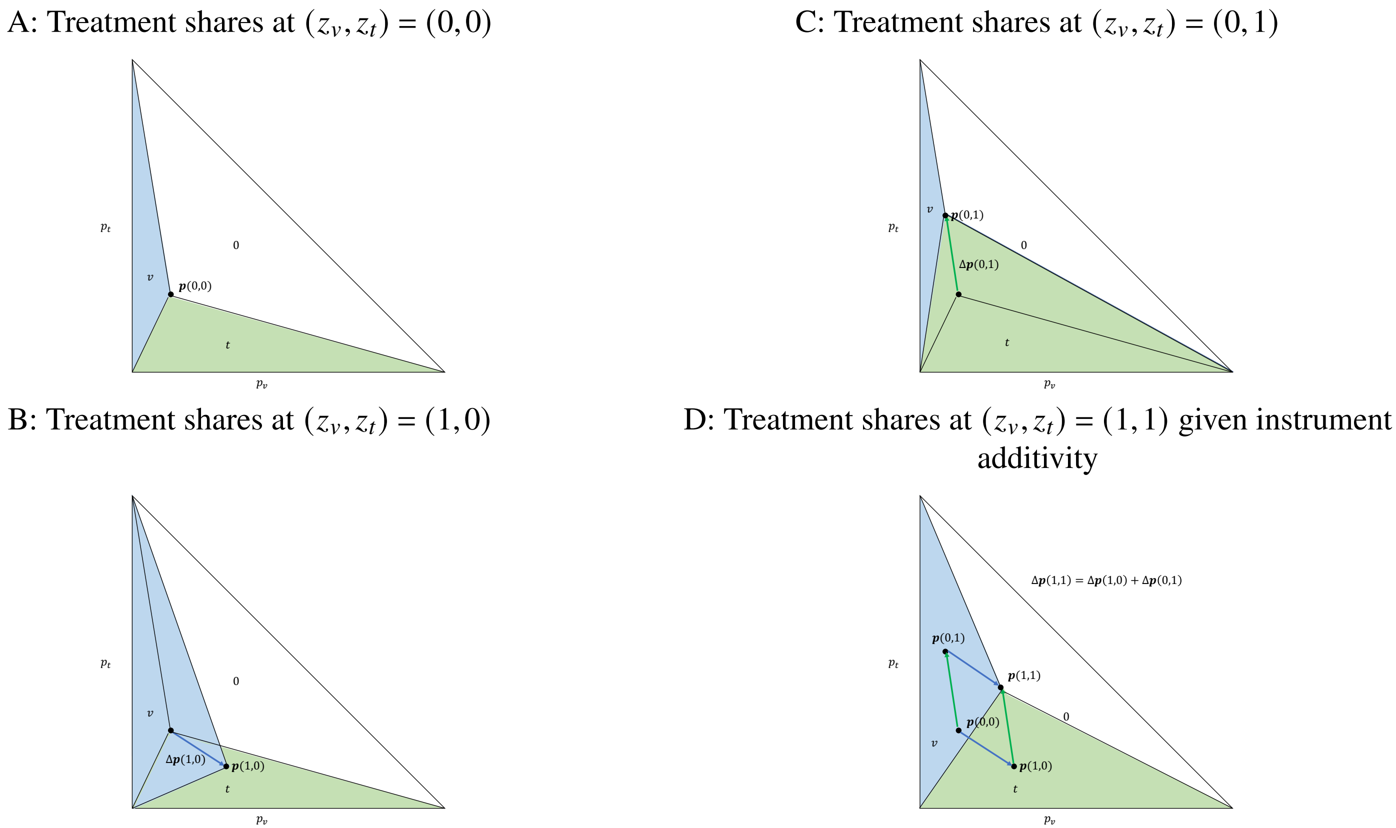
z_v, z_t	C_0^v	C_0^t	$C_0^{v+,t-}$	$C_0^{v-,t+}$	C_t^v	C_t^{v-}	C_v^t
0,0	0	0	0	0	t	t	v
1,0	v	0	v	v	v	v	v
0,1	0	t	t	t	t	t	t
1,1	v	t	v	t	v	t	t

B: Effect of change in z_v, z_t from (0,0) to (1,0)

z_v, z_t	C_0^v	C_0^t	$C_0^{v+,t-}$	$C_0^{v-,t+}$	C_t^v	C_t^{v-}	C_v^t
0,0	0	0	0	0	t	t	v
1,0	v	0	v	v	v	v	v
0,1	0	t	t	t	t	t	t
1,1	v	t	v	t	v	t	t

Note: This figure duplicates information provided in Table 1.2, highlighting how treatment status changes as values of the instrument change. Rows 1, 2, 3, and 4 show potential treatments for each complier group when z_v and z_t are both 0, when $z_v = 1$ and $z_t = 0$, when $z_v = 0$ and $z_t = 1$, and when both instruments equal 1, respectively. Panel B illuminates the effect of a change in (z_v, z_t) from (0,0) to (1,0). With this change, C_0^v, $C_0^{v+,t-}$, and $C_0^{v-,t+}$ compliers are induced from 0 to v, and C_t^v and C_t^{v-} compliers are induced from t to v.

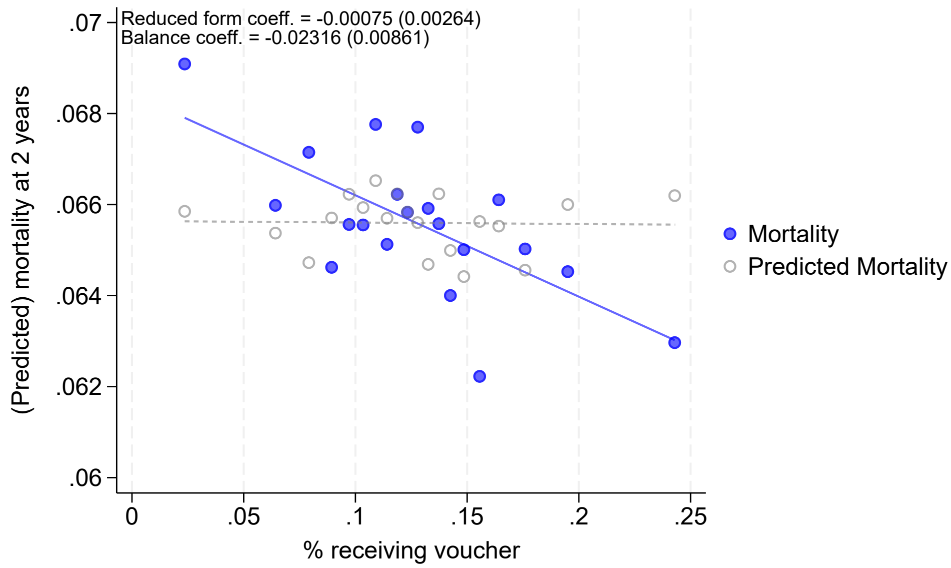
Figure A.3: Simplex visualization of instrument additivity



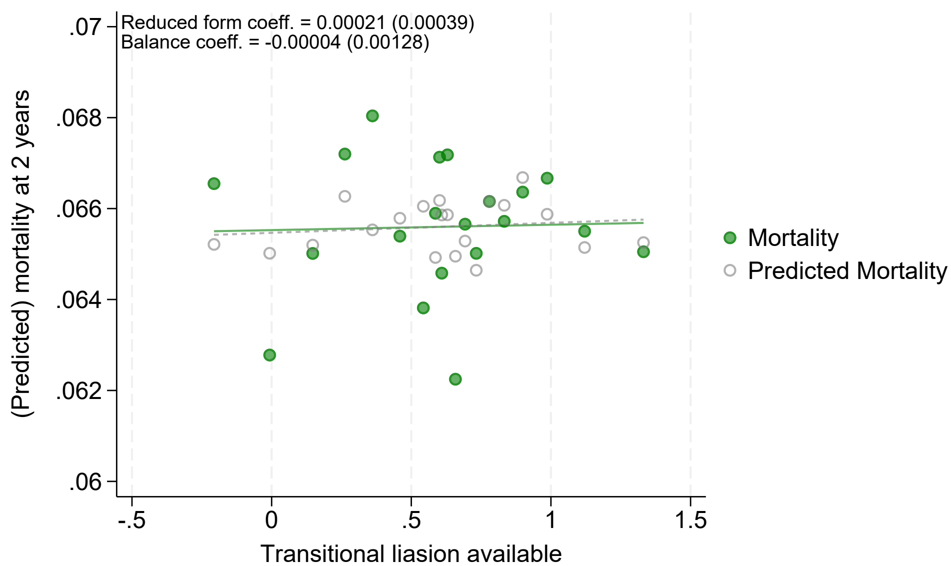
Note: This figure visualizes treatment shares and the additivity condition in simplex space for the case of simple binary instruments $(Z_{v,i}, Z_{t,i}) \in \{0,1\}^2$. Let $\mathbf{p}(z_v, z_t) = (p_v(z_v, z_t), p_t(z_v, z_t), p_0(z_v, z_t))'$, a point on the simplex, represent the shares of individuals in v, t, and 0 based on the realized values of the instruments. For instance, $\mathbf{p}(0,0)$ represents treatment shares when both instruments are off. Further, let $\Delta\mathbf{p}(z_v, z_t) = \mathbf{p}(z_v, z_t) - \mathbf{p}(0,0)$, or the change in treatment shares from $(Z_{v,i}, Z_{t,i}) = (0,0)$. Across all panels, the x-axis of the simplex shows the share of the population in v, and the y-axis shows the share of the population in t. Shaded in blue is the share of the population in v, shaded in green is the share of the population in t, and in white is the share of the population in 0. Panel A visualizes hypothetical treatment shares at $\mathbf{p}(0,0)$. Panel B visualizes the change in shares when $Z_{v,i}$ is turned on, and Panel C visualizes the change in shares when $Z_{t,i}$ is turned on. Finally, Panel D shows where $\mathbf{p}(1,1)$ must lie under instrument additivity (Eq. 3).

Figure A.4: Reduced form and first stage for $Z_{v,i}$ and $Z_{t,i}$

A: $Z_{v,i}$

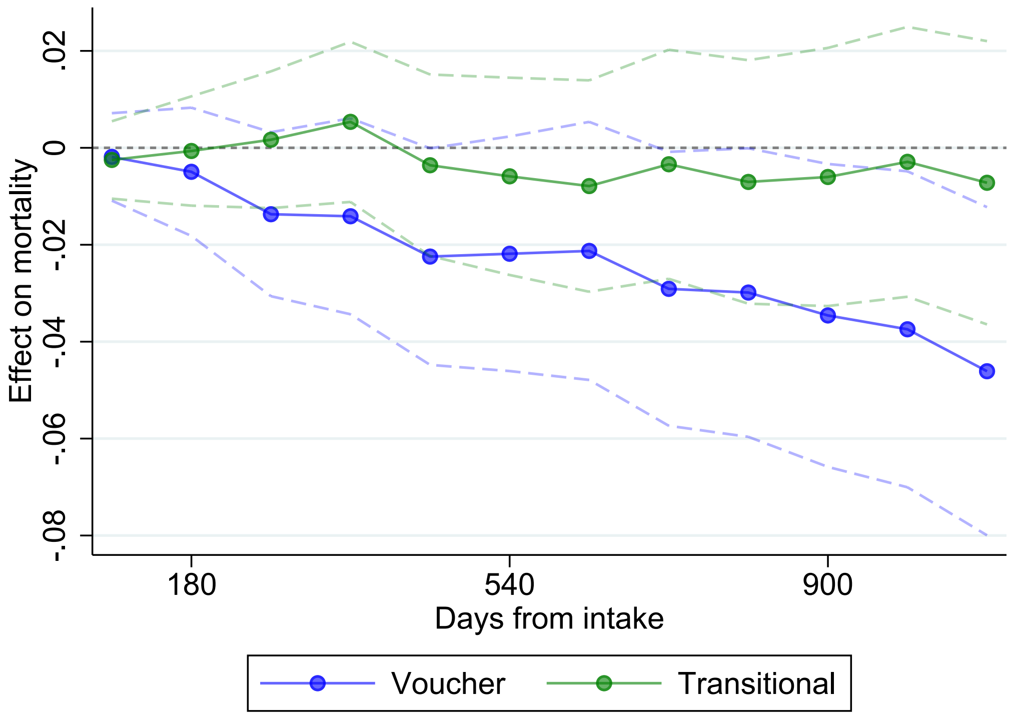


B: $Z_{t,i}$



Note: This figure plots the relationship between $\hat{Y}_i$ (predicted mortality) and the instruments as a balance test, as well as the relationship between Y_i (actual mortality) and the instruments (the reduced form relationship). Panel A shows the relationship for $Z_{v,i}$, and Panel B shows the relationship for $Z_{t,i}$. Coefficients on the instruments when $\hat{Y}_i$ and Y_i are the left hand side variables are shown in the upper left-hand corner.

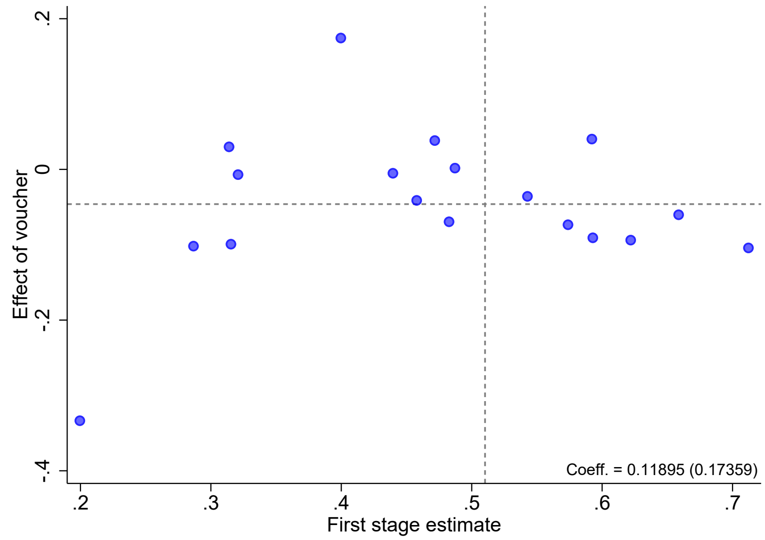
Figure A.5: Mortality effect over time



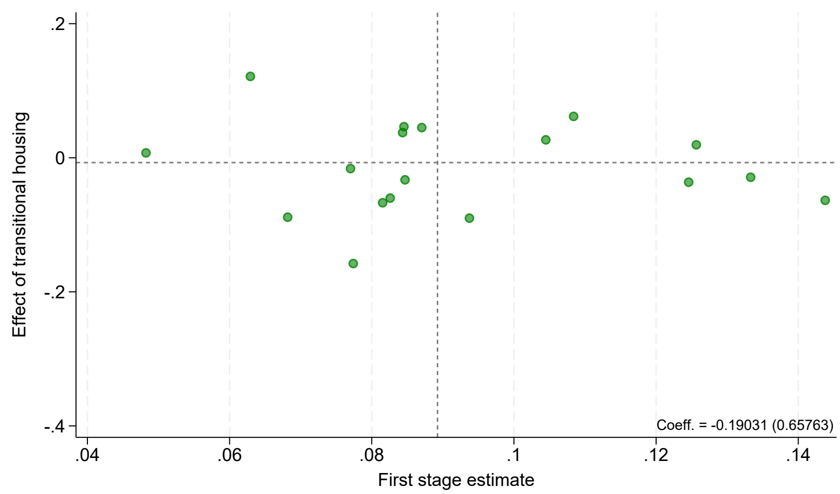
Note: This figure shows 2SLS regression results for the effect of the voucher program and transitional housing on mortality over time. Results for the voucher program are shown in blue, and results for transitional housing are shown in green. Dotted lines show upper and lower bounds.

Figure A.6: Treatment effect heterogeneity by propensity to be treated

A: LATE vs. first stage for housing first



B: LATE vs. first stage for transitional housing

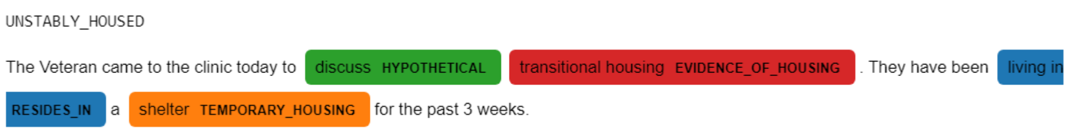


Note: This figure shows heterogeneity in 2SLS estimates for the effect of the voucher program on mortality (Panel A) and the effect of transitional housing on mortality (Panel B). In Panel A, I run first stage regressions for the voucher program separately by Veteran Integrated Service Network (VISN), collecting coefficients for the effect of $Z_{v,i}$ on $D_{v,i}$. Then, I estimate the LATE (for the effect of the voucher program on mortality) for each VISN, and plot the relationship between them. Similarly, for Panel B, I run the first stage for transitional housing separately by VISN and collect coefficients for the effect of $Z_{t,i}$ on $D_{t,i}$. Then, I estimate the LATE (for the effect of the transitional housing on mortality) for each VISN, and plot the relationship. Dotted lines show the overall sample first stage estimate (x-axis) and the overall sample 2SLS estimate (y-axis). The bottom right corner of each panel shows the regression coefficient for the relationship between IV estimates and first stages.

Figure A.7: Natural language processing algorithm examples

A: Example note for stable housing

B: Example note for unstable housing

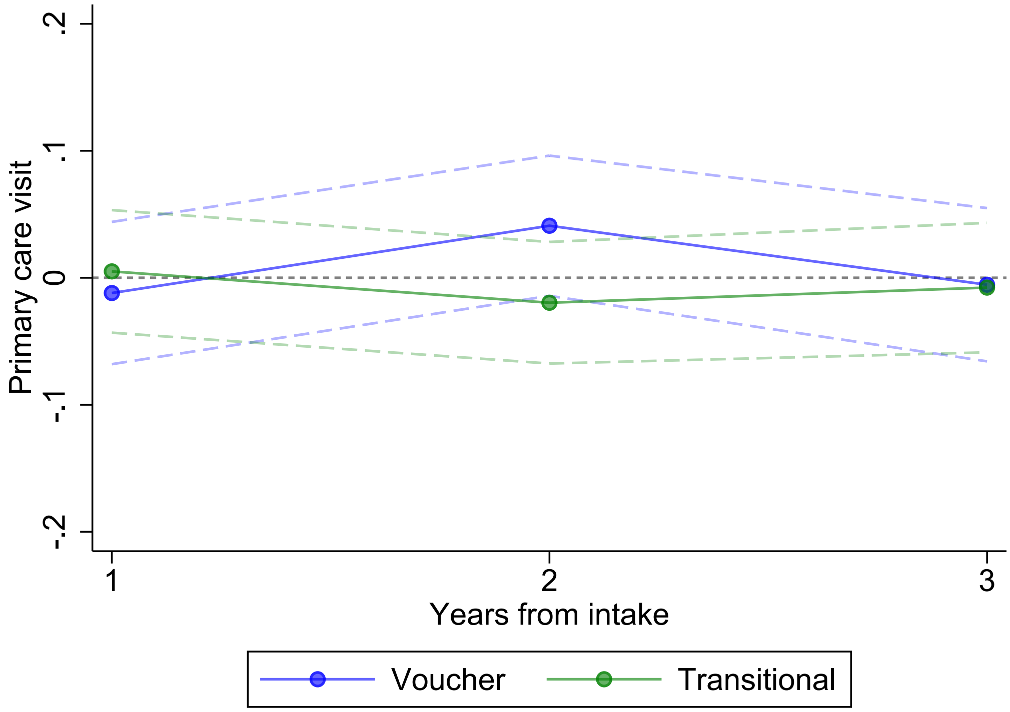


Note: This figure shows two example patient notes processed by the ReHouSED NLP algorithm. The algorithm highlights words either related to housing ("was homeless") or that provide context (such as being hypothetical or historical). Panel A is a note categorized as stably housed, and Panel B is a note categorized as unstably housed. Chapman et al. (2021) provides more precise rules on the logic of the NLP algorithm.

Figure A.8: Housing status over time for compliers

Note: This figure shows mean potential housing status over time for C_0^v and C_0^t compliers who do and do not receieve the intervention. Green indicates the share of compliers who are unstably housed in the quarter, red indicates the share that are stably housed, and grey indicates unknown housing status.

Figure A.9: Engagement with primary care over time



Note: This figure shows 2SLS regression results for the effect of the voucher program and transitional housing on engagement with primary care in years 1, 2, and 3 following intake (each year corresponds to one dot in the figure). Results for the voucher program are shown in blue, and results for transitional housing are shown in green. Dotted lines show upper and lower bounds.

Appendix B

Appendix to Chapter 2

B.1 Alternative Samples

B.1.1 Patients with no outside insurance

In Section 2.4.5, I discuss a potential issue with the interpretability of my results: what if patients also receive mental health care outside of the VA? To address this issue, I create a subsample of patients who have no outside or non-VA insurance, making it highly likely that they receive care only at the VA. This subsample comprises 69.7% of patients in the primary sample. In Table B.1, I show that across characteristics, the no insurance subsample is highly analagous to the main sample, with two exceptions. First, the mean age is slightly younger compared to that of the main sample. Second, two-year mortality in this subsample is 5.3%, as opposed to 7.5% in the main sample. Both of these facts likely reflect the fact that most patients with outside insurance have Medicare, which is only available for patients above 65 or for those who qualify due to disability status.

Mortality results for this sample are presented in Table B.4.

B.1.2 Patient without scheduled follow up visits

In the sample selection described in Section 2.3.1, I exclude patients with no scheduled follow up visit, as W_{-i} depends on type of appointment scheduled. Since this group is comprised of a substantial 314,554 patients, in Table B.1, I also compare characteristics of patients with no scheduled follow up appointments with those of the main sample. As expected, I find that these patients have much lower prior attachment to VA mental health services and are more likely to have Medicare, implying that more of these patients are using care outside of the VA. Additionally, patients with no follow up appointment are older and are much more likely to be diagnosed with SUD, suggesting that these types of patients may be particularly unlikely to return for mental health services.

Finally, these patients have slightly higher mean wait time than that of the main sample. This fact implies that those who arrive during more congested times may be less likely to have a follow up appointment scheduled at all. As this presents a potential non-random attrition problem that could bias results, I add these patients back to the main sample in Appendix Section B.3.1.

B.2 Additional balance checks

As the VA flags patients considered to be high risk for suicide, a natural additional balance test is to see whether these high risk patients are just as likely to be assigned low and high values of W_{-i}. I restrict to patients flagged as high risk before the ED visit, as being assigned this flag at the ED visit itself could potentially be endogeous to my experiment. Overall, only 0.3% of patients in my sample are assigned the high risk flag before the ED visit. In Figure B.2 Panel A, I show a binscatter of percent of patients assigned the high risk flag versus W_{-i}, residualizing by the baseline controls. I find a small and statisitically insignificant coefficient, providing more evidence that mean wait time is approximately randomly assigned.

Secondly, another key metric indicative of patient severity is whether a patient has a prior hospitalization. Let I_i indicate that a patient was hospitalized in the year prior to the index ED visit (for any cause). In Figure B.2 Panel B, I show a binscatter of I_i versus W_{-i}, residualizing by the baseline controls. I find that the coefficient on W_{-i} is both reverse-signed and insignificant, providing further evidence that W_{-i} is uncorrelated with patient severity.

B.3 Additional robustness checks

B.3.1 Patients without scheduled follow up appointments

To eliminate concern that dropping these patients has implications for my findings, I add patients with no scheduled follow up appointments back to the sample, increasing the sample size to 935,843. This combined sample has a somewhat higher mean two-year mortality of 9.6%.

As the primary construction of W_{-i} and controls rely on assigned appointment type, I make a few modifications to the design in order to run regressions using this larger sample. First, since patients with no scheduled follow up do not have an assigned appointment type, I use two alternative approaches to approximate congestion levels for these patients. I start with a simple average approach: I group all appointment types together during two-week-hospital periods to construct W_{-i} for these patients. Alternatively, I let $W_{-i} = \hat{W}_{-i}$, as described in Section 2.4.3. of the main text.[1] In this latter approach, $\hat{W}_{-i}$ is essentially a weighted average mean wait time, weighted by the probability of the patient being assigned a particular appointment type. As appointment type is also a baseline control in my main analysis, in both specifications, I control for either appointment type directly (for the baseline sample) or the predicted probability of being assigned each appointment type (for patients with no appointment type).

In Table B.3, I show regression results using both of the aforementioned methods. I show results both with the baseline controls and all additional controls used in Table 2.1; results are both robust and very similar to those of the baseline analysis sample.

B.3.2 Miscellaneous robustness checks

In Figure B.3, I show that the baseline results are stable when both 1) detailed patient controls are added, and 2) sensible modifications to the sample or treatment variable are made. Under "Additional controls," I first show the main result with the baseline controls– two-week periods,

[1] W_{-i} for observations also in the baseline sample are held fixed in these analyses.

year-month $\times$ hospital, appointment type, and age buckets. Then, I incrementally add groups of additional controls (the next row adds demographics controls, the following row adds demographic and prior utilization controls, etc.).[2] Demographic controls include sex, race, and ethnicity. Prior utilization controls includes count of inpatient and outpatient mental health visits in the previous year, count of mental health appointment cancellations, and count of no shows. Diagnosis controls are 3-digit ICD-9 diagnosis codes associated with the ED visit. Priority status refers to a patient's eligibility for VA healthcare and responsibility for paying copays, which depends on both disability status and income. The second to last row in this section adds leave-out mean characteristics of patients who form a given patient's mean wait time. Leave-out characteristics include mean age, mean prior visits, and mean predicted mortality. Finally, the last row controls for first half versus second half of the month indicators. The motivation behind these final controls is that different types of patients could arrive to the hospital during the beginning or end of the month, perhaps driven by holidays, paychecks, or benefit schedules.[3]

Under "Sample changes", I consider some possible reasonable modifications to the sample and construction of W_{-i}.[4] First, the primary sample considers any appointment within a 90-day window post ED visit to be a follow up appointment. In contrast, in the first row, I only consider follow up appointments within a 60-day window; patients with a first appointment scheduled between days 61 and 90 are dropped. In the second row of this section, related to the concern that holidays may create non-idiosyncratic variation in crowding, I drop all ED visits in December.

Next, I present results using a slightly modified version of W_{-i}. One potential concern is that wait times of patients who arrive during the same time window as a given patient may be influenced by the patient's own wait time (for example, if a high risk patient is pushed to an earlier appointment slot, this may cause more delays for other patients who arrive at a similar time). To mitigate this concern, in this alternative version of W_{-i}, I leave out wait times of not only the patient themselves, but of all patients who arrive on the same day. This robustness check also accounts for the potential of correlated same-day patient shocks, such as multiple veterans ingesting the same contaminated drug.

In Appendix Table B.5, I perform an additional series of robustness checks. To capture potential correlation in treatment assignment within a given hospital, in column 2, I use hospital level rather than hospital-month level clustering. In column 3, I add day of week fixed effects. In column 4, to capture potential seasonal variation in types of patients arriving, I interact month of the year with ICD-9 3-digit diagnosis group.

Additionally, one potential concern is that those with prior mental health utilization are differentially affected by congestion. Those with prior mental health utilization likely have an advantage in obtaining timely follow up appointments, as they may have pre-existing relationships with providers. However, W_{-i} does not use a patient's wait time directly; instead, it relies on the wait times of patients arriving at the same time as the index patient. Consequently, W_{-i} should exogenously capture clinic congestion, rather than capturing anything about a patient's prior utilization. However, as an additional test to ensure that W_{-i} only captures congestion, rather than variation in patients' prior utilization which may affect wait times, in column 5, I perform a robustness check where I construct W_{-i} only using patients with no prior mental health utilization.

[2]These controls overlap with the controls added in Table 2.1.

[3]Note that ability to pay is not a major concern here, since the vast majority of mentally ill patients receive free mental health services at the VA.

[4]When the main sample selection is modified, the construction of W_{-i} is held constant, unless otherwise noted.

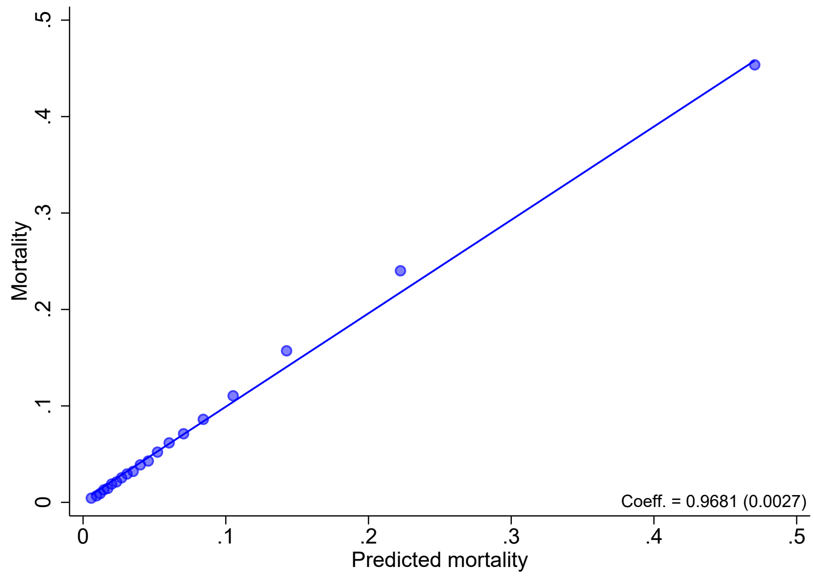
Finally, although all patients in my sample have a mental health diagnosis code connected with their visit, some patients who visit the ED may be experiencing both a mental and physical health emergency. Specifically, 39.9% of the sample also has a *physical* health diagnosis connected to their ED visit, the most common of which are hypertension, sleep disturbances, and respiratory distress. Some of these may be directly connected to mental health – for instance, those with anxiety may experience breathing issues or difficulty sleeping. In column 6, I add controls for physical health diagnoses assigned at the ED visit, including an indicator for no physical health diagnosis.

Table B.1: Patient characteristics across samples

	Main sample	Patients with no other health insurance	Patients with no scheduled follow up visit
Male	0.908	0.905	0.927
Black	0.232	0.254	0.205
Hispanic	0.057	0.059	0.048
Age	51.4	47.9	57.8
Prior outpatient visit	0.573	0.547	0.188
Prior inpatient visit	0.096	0.103	0.027
Substance use disorder	0.285	0.321	0.446
Mood disorders	0.299	0.300	0.142
Anxiety	0.120	0.115	0.140
Psychosis	0.092	0.083	0.070
PTSD	0.103	0.101	0.048
Medicare	0.168	0.000	0.247
Other non-VA insurance	0.135	0.000	0.131
Mean wait time	17.73	17.55	18.30
Mortality	0.075	0.053	0.137
N	621,289	432,863	314,554

Note: This table shows characteristic means for the baseline sample compared to both the subsample of patients with no non-VA insurance and the dropped sample of those who have no scheduled follow up mental health visit. Prior oupatient and inpatient visit are indicators for attending a mental health appointment in the previous year. Mean wait time is equivalent to W_{-i} for the baseline sample. To form mean wait times for patients without a follow up mental health visit, I group all appointment types together, as I cannot observe which type of mental health clinic they would have been scheduled for.

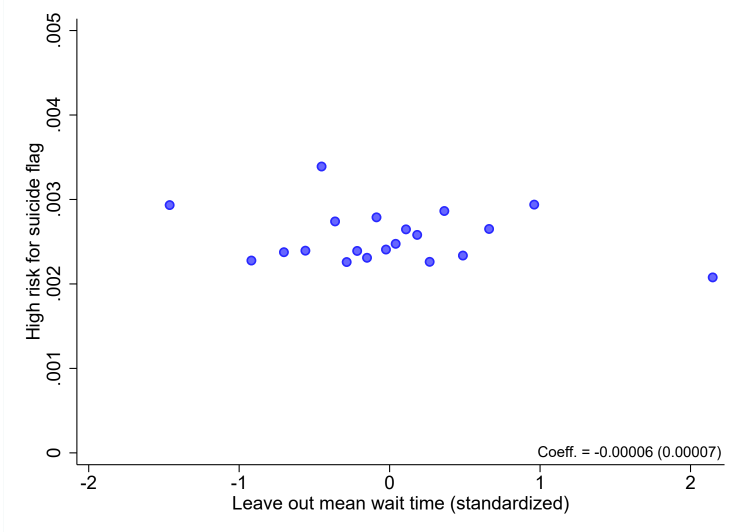
Figure B.1: Predictive power



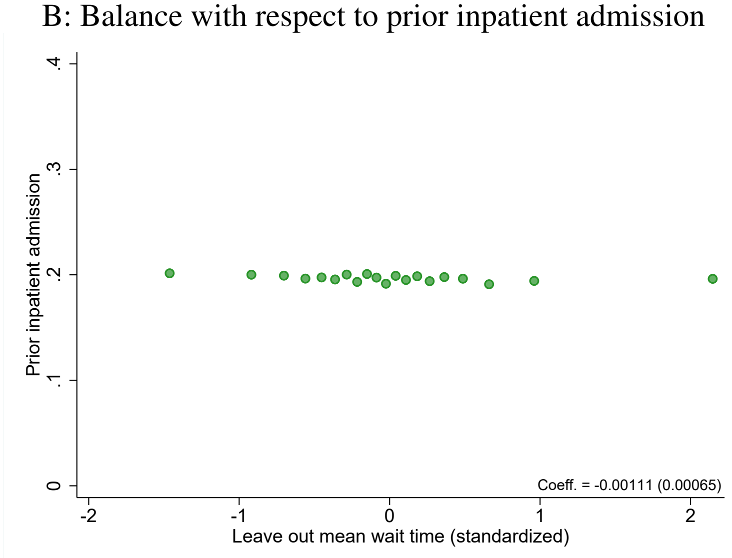
Note: This figure plots observed mortality, Y_i, versus predicted mortality, $\hat{Y}_i$, in twenty equal sized bins. The coefficient shows the regression fit of observed mortality on predicted mortality.

Figure B.2: Alternative balance tests

A: Balance with respect to high risk for suicide flag



B: Balance with respect to prior inpatient admission



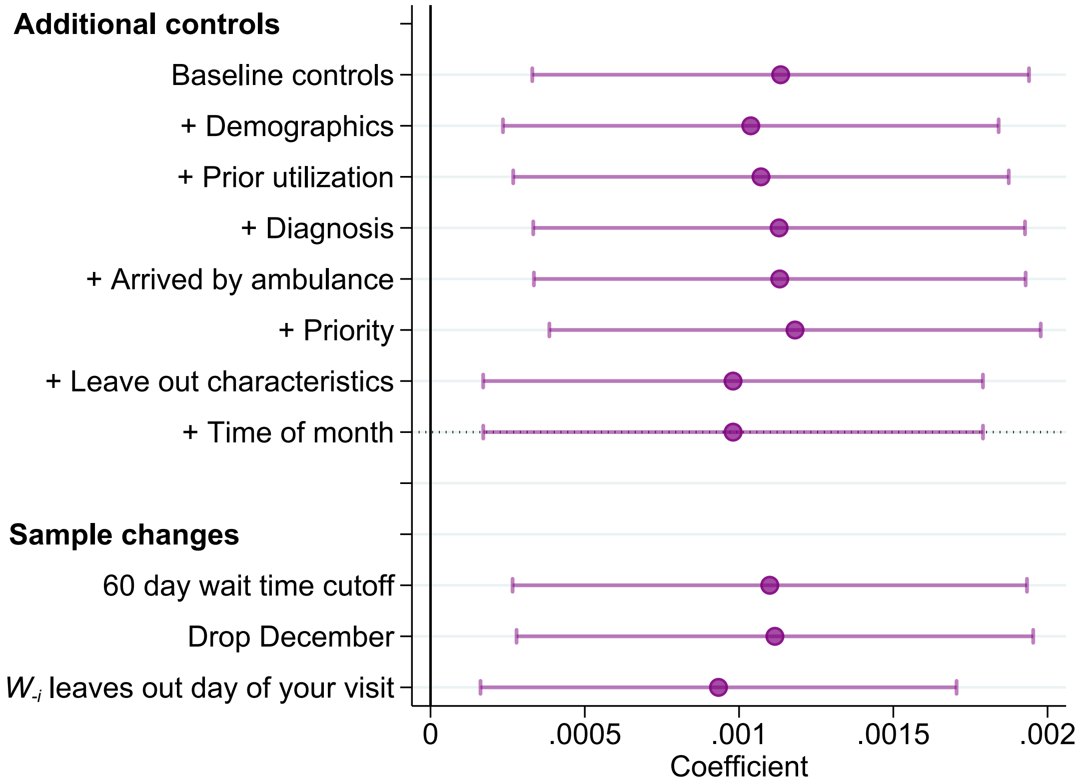
Note: Panel A shows a binscatter of the percent of patients assigned the high risk for suicide flag versus W_{-i}, residualized by the baseline controls. Panel B shows a binscatter of prior inpatient admission, I_i, versus W_{-i}, residualized by the baseline controls.

Table B.2: Selection into appointment type

	General	Psychiatry	PTSD	Substance use	Other
Mean wait time$_k$	-0.0047	-0.0035	-0.0027	-0.0018	-0.0017
	(0.0011)	(0.0007)	(0.0011)	(0.0005)	(0.0007)
Mean wait time$_k \times \hat{Y}$	-0.0043	-0.0053	-0.0013	0.0002	-0.0013
	(0.0011)	(0.0007)	(0.0011)	(0.0006)	(0.0008)
Outcome mean	0.612	0.108	0.069	0.059	0.152
Observations	621,289	621,289	621,289	621,289	621,289

Note: This table shows regression results corresponding to Eq. 5 in the main text. The left hand side variable is an indicator for the patient being assigned a given appointment type (general, psychiatry, PTSD, substance use, or other specialty appointment). The first row shows β_1, the coefficient on residualized wait time for appointment type k, and the third row shows β_3, the coefficient on the interaction between residualized wait time and predicted mortality. Wait times are residualized by hospital $\times$ month, two week period, and age buckets. $\hat{Y}_i$ is standardized for ease of interpretation. General includes therapy visits with psychologists and social workers. All standard errors are clustered by hospital-month.

Figure B.3: Robustness to additional controls and sample changes



Note: This figure shows the effect of not attending a mental health appointment on 2-year mortality when various changes to the controls, sample, or wait time measure are made. Dots are point estimates; bars show 95% confidence intervals. Under "Additional controls," I first show the main estimate using the baseline controls, two-week period, year-month × hospital, appointment type, and age buckets. I then incrementally add additional groups of controls. Demographics includes sex, ethnicity, and race. Prior utilization includes prior mental health outpatient visits, no shows, and cancellations, as well as prior inpatient mental health visits. Diagnosis refers to 3-digit ICD-9 diagnosis codes from the ED visit. Priority status, a designation between 1 and 8 assigned by the VA based on income and disability status, indicates eligibility for VA healthcare and copay responsibilities. Leave out characteristics include mean age, prior visits, and predicted mortality among patients in one's hospital-time-appointment group (the same group that forms a patient's W_{-i}). Time of month are indicators for first or second half of the month. Under "Sample changes," I make various modifications to the sample selection or construction of W_{-i}. The first row amends the definition of follow up appointment from 90 days to 60 days, the second drops ED visits in December, and the third row excludes all visits on the same day as the patient themselves from the construction of W_{-i}.

Table B.3: Main results, including patients with no scheduled follow up

	2-year mortality			
	(1)	(2)	(3)	(4)
	Average wait time	Average wait time	Predicted wait time	Predicted wait time
Mean wait time	0.00070	0.00095	0.00118	0.00124
	(0.00034)	(0.00035)	(0.00035)	(0.00035)
Outcome mean	0.0956	0.0956	0.0956	0.0956
Observations	935,843	935,843	935,843	935,843
Baseline controls	Yes	Yes	Yes	Yes
Additional patient controls	No	Yes	No	Yes
Leave out controls	No	Yes	No	Yes

Note: This table shows baseline mortality results, adding patients back to the sample who had no scheduled follow up appointment. "Average wait time" refers to forming W_{-i} by grouping all appointment types together (for those without a scheduled follow up). Predicted wait time refers to setting $W_{-i} = \hat{W}_{-i}$ for these patients. All standard errors are clustered by hospital-month.

Table B.4: Effect of mental health wait times on mortality for no insurance subsample

	(1) No visit	(2) Mortality	(3) Mortality	(4) Mortality
Mean wait time	0.0218	0.00075	0.00082	0.00078
	(0.0010)	(0.00044)	(0.00043)	(0.00044)
Outcome mean	0.3624	0.0528	0.0528	0.0528
Observations	432,863	432,863	432,863	432,863
Baseline controls	Yes	Yes	Yes	Yes
Additional patient controls	No	No	Yes	Yes
Leave out controls	No	No	No	Yes

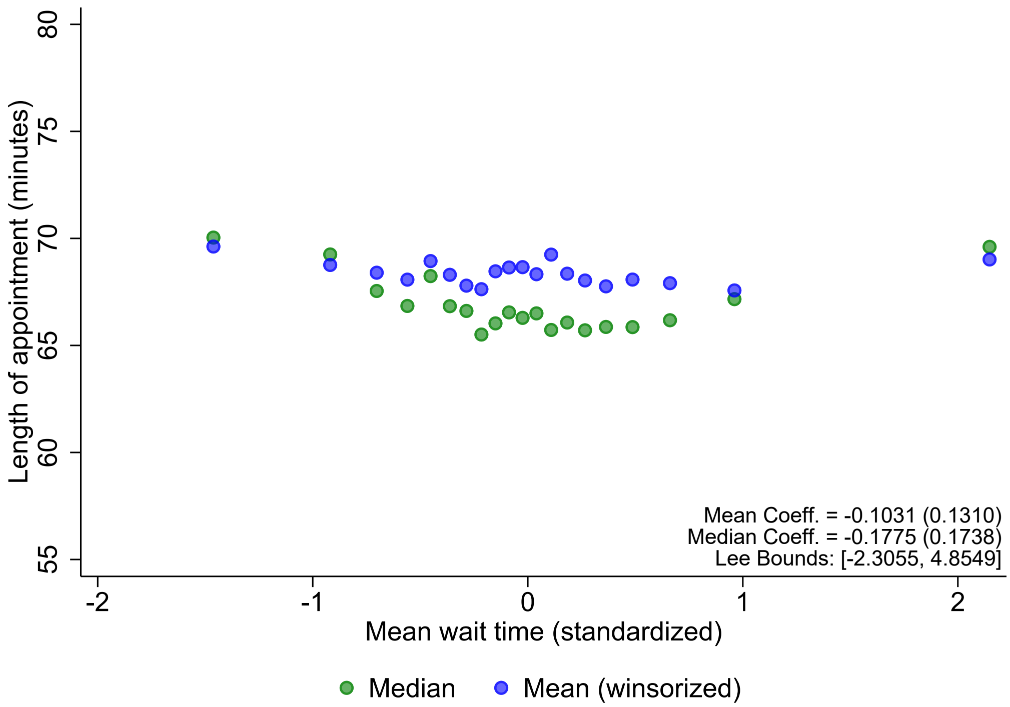
Note: This figure shows the main results for the subsample who have no other non-VA insurance. Column 1 shows the effect of mean wait time on the probability of not attending a follow up mental health visit. Column 2 shows the main mortality estimate, only using baseline controls. Columns 2 shows the estimate when I add additional patient controls, including gender, race, diagnosis, and prior utilization. Finally, column 4 shows the estimate when I control for characteristics of other patients who form a given patient's value of W_{-i}, including (leave-out) mean age, mean prior visits, and mean predicted mortality. Baseline controls include month-year × hospital, appointment type, and age buckets. Standard errors are clustered by hospital-month.

Table B.5: Additional robustness checks

	(1) Baseline	(2) Hospital-level clustering	(3) Day of week controls
Mean wait time	0.00113	0.00113	0.00114
	(0.00041)	(0.00044)	(0.00041)
Outcome mean	0.0746	0.0746	0.0746
Observations	621,289	621,289	621,289
	(4) Month × disease type controls	(5) Mean wait time excludes those with prior utilization	(6) Physical health diagnosis controls
Mean wait time	0.00121	0.00096	0.00098
	(0.00041)	(0.00040)	(0.00041)
Outcome mean	0.0746	0.0751	0.0746
Observations	621,289	575,734	621,289

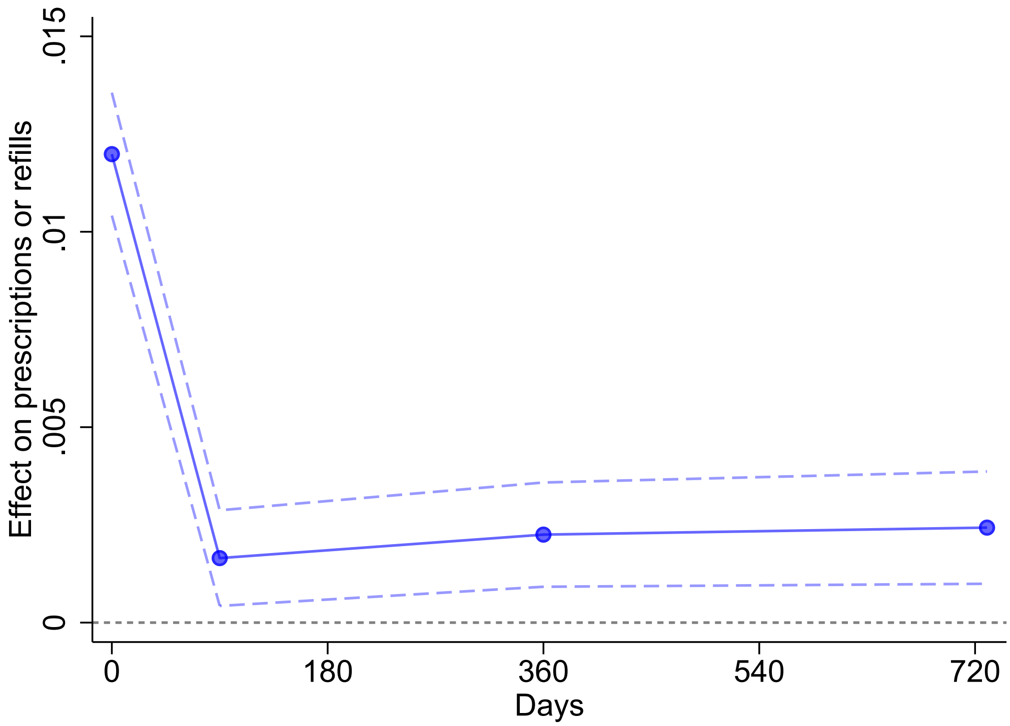
Note: This table shows additional robustness checks for the mortality estimate. The fourth column interacts month of the year with ICD-9 3-digit diagnosis group. The fifth column constructs W_{-i} only using patients with no prior mental health utilization. The sixth column controls for physical health ICD-9 codes assigned at the ED visit.

Figure B.4: Length of follow up appointment vs. mean wait time



Note: This figure shows a binned scatterplot of the length of one's follow up appointment, in minutes, versus W_{-i}, residualized by the baseline controls. The mean reported coefficient winsorizes observations at the 5th and 95th percentiles, while median reported coefficient calculates the median wait time in hospital-two-week-appointment type bins. Lee Bounds reports worst case scenario bounds using median wait times.

Figure B.5: Effect of mean wait time on prescriptions over time



Note: This figure shows the effect of mean wait time on mental health prescriptions or refills. The first dot shows the effect of mean wait on having any drug prescribed at the ED visit. The second dot shows the effect on having any drug prescribed between days 0 and 90 (including the day of the ED visit). Finally, the third and fourth dots show the effect on prescriptions or refills between days 90-360 and 360-730, respectively.

Table B.6: Cause of death

	High wait time patients	Full sample	Risk ratio
All cause mortality	0.0756	0.0745	1.015
	(0.0004)		
Physical disease	0.0660	0.0648	1.018
	(0.0004)		
Accident, suicide, or overdose	0.0096	0.0096	0.993
	(0.0002)		
Heart disease	0.0169	0.0163	1.034
	(0.0002)		
Cancer	0.0131	0.0130	1.006
	(0.0002)		
Respiratory diseases	0.0071	0.0068	1.045
	(0.0001)		
Liver disease	0.0049	0.0047	1.025
	(0.0001)		
Other cardiovascular	0.0038	0.0036	1.034
	(0.0001)		
Infections	0.0030	0.0029	1.039
	(0.0001)		
Other physical disease	0.0203	0.0203	1.001
	(0.0002)		

Note: This table shows the probability of dying from a particular cause for both the full sample and for those with a mean wait time one standard deviation above the mean. The last row shows the relative risk of dying from that cause for those with high wait times versus the full sample.

Table B.7: Effect of mean wait time on intermediate outcomes for substance use subsample

	A: Mental health outcomes		
	(1)	(2)	(3)
	No follow up visit	Returned to mental health care	Mental health inpatient treatment
Mean wait time	0.0239	-0.0104	-0.0088
	(0.0016)	(0.0009)	(0.0015)
Outcome mean	0.3385	0.9286	0.4151
Observations	175,721	175,721	175,721
	B: Physical health outcomes		
	(4)	(5)	(6)
	Primary care visit	Inpatient visit - non-mental health	New physical comorbidity
Mean wait time	0.0027	0.0044	0.0032
	(0.0011)	(0.0013)	(0.0014)
Outcome mean	0.8317	0.2323	0.6529
Observations	175,721	175,721	175,721

Note: This table shows intermediate outcomes for the SUD subsample. Panel A shows results for engagement with mental health care. Returned to mental health care refers to attending any oupatient appointment. Mental health inpatient treatment refers to any mental health inpatient visit, including residential stays (excluding admissions within one day of the initial ED visit). Panel B presents results for non-mental health outcomes. The first column looks at whether a patient attended a primary care appointment, the second column looks at non-mental health inpatient admissions, and the third column looks at whether the patient had any new Elixhauser comorbidity (only including physical conditions) during the two-year follow up period.

Appendix C

Appendix to Chapter 3

C.1 Convex weighting of treatment effects

Recent innovations to the difference-in-difference literature note that standard designs that fit models using OLS can place negative weights on event time specific treatment effects. Thus, standard approaches can yield spurious results, particularly in cases where treatment adoption is staggered and there is treatment effect heterogeneity (Borusyak et al. 2024, Callaway and Sant'Anna 2021, Chaisemartin and D'Haultfoeuille 2020). This issue is less of a concern in my design, as I consider treatment to begin in 1936 for all Hearst counties. Nevertheless, issues with negative weights can still arise if the model is misspecified – for example, in cases where the control variables are insufficient (Blandhol et al. 2022).

Thus, to ensure my results are robust to any issues with negative weights, I employ the difference-in-differrence method proposed by Callaway and Sant'Anna (2021). Their approach breaks up the difference-in-difference coefficient into year specific treatment effects, ensuring all estimates receive a convex weight. In Table C.2, I show difference-in-difference results using Callaway-Sant'Anna weighting for both the full sample and the sample of large counties. As Callaway and Sant'Anna's approach does not allow for a continuous treatment variable, in Table C.2, I also show equivalent results using the standard difference-in-difference approach where the treatment is just an indicator for Hearst presence in a county interacted with post-1932. The Callaway-Sant'Anna estimates are very similar to those of the baseline approach.

C.2 Additional robustness checks

In Table C.3, I show estimates of $\hat{\beta}$ using various alternative specifications. Column 1 shows $\hat{\beta}$ for the full sample with all baseline and newspaper controls (as in Table 2, column 2). Column 2 adds additional county controls interacted with year, including share of population in manufacturing, share of population in agriculture, share native born white, share of population above 65, share of population that is Catholic, mean occupationational income score, and log average farm size. Column 3 adds state $\times$ year fixed effects. Column 4 excludes non-Manhattan New York City counties, as these counties may have only been partially treated (since Hearst newspapers were located in Manhattan).

Next, in the baseline specification, I use controls from 1930, before the peak of the Great Depression. Thus, one potential confounder is that Hearst counties were either 1) disproportionately hit by the Depression; 2) received a disportionate share of New Deal federal grants and loans, affecting attitudes towards FDR. Therefore, column 5 adds controls related to the Depression and the New Deal: growth in retail sales per capita (1929-1933), total New Deal grants, and total New Deal loans (Fishback et al. 2003). Finally, in column 6, I add counties that are contiguous to Hearst counties back to the sample. The point estimate is marginally smaller, reflecting the fact that those in nearby counties may have been partially treated. Overall, the results are robust across specifications.

C.3 Hearst's other media holdings

In the late 1920s, Hearst began to take an interest in radio, buying up local stations to compliment his newspapers. These stations were often directly tied to local papers, used to both advertise the Hearst brand name to non-readers and promote Hearst's politics (Nasaw 2013). To disentangle the effect of Hearst radio from Hearst newspaper circulation, in Table C.4, I show difference-in-difference results with two treatments: Hearst circulation per capita, as in the main specification, and an indicator for radio presence in the county. I run the following specification:

$$Y_{it} = \beta H_i \times \text{Post}_t + \lambda R_i \times \text{Post}_t + \mathbf{X}_{it}^0 \gamma + \eta_t + \zeta_i + \varepsilon_{it} \quad \text{(C.1)}$$

where R_i is an indicator for radio presence in a given county.[1] Estimates of λ are negative but insignificant, while estimates of β are slightly smaller in magnitude. These results provide some evidence that Hearst radio may have also affected voting outcomes, but the effect of newspaper circulation remains robust.

Additionally, during this time, Hearst owned various magazines, including *Cosmopolitan* and *Good Housekeeping*. However, these magazines both had nationwide circulation and were less politically focused, instead publishing primarily fiction or lifestyle pieces (Nasaw 2013). Therefore, these magazines likely had little effect on voting outcomes. Even if these magazines did have an influence on public opinion, since those in control counties could also buy Hearst magazines (and thus also be potentially exposed to Hearst's views), any effect of magazines would likely make $\hat{\beta}$ an underestimate of the true treatment effect.

C.4 Synthetic control

As an alternative to the event study and difference-in-difference designs, due to the small number of treated units, I instead create a counterfactual using the synthetic control method (Abadie and Gardeazabal 2003, Abadie et al. 2010). As this approach is designed for binary treated and untreated units, I define treatment here as simply Hearst presence in county. Starting with the sample of large counties and matching on pre-trends in Democratic vote share, I match each Hearst county to a weighted average of control counties (the "synthetic controls"). In Figure C.5, I plot the mean Democratic vote share across time for Hearst counties and synthetic Hearst counties, respectively.

[1] As radio listenership may spillover to nearby counties, in these specifications, I exclude contiguous border counties from the control group for both Hearst newspaper and radio counties.

I find that Democratic vote share in the synthetic control counties visually matches that of Hearst counties in the pre-period. After 1936, however, the lines diverge. Assuming Hearst counties would have followed a similarly trajectory to that of the synthetic controls under the counterfactual, these results suggest that Hearst presence in a county decreased Democratic vote shares by 2.1, 5.1, and 5.2 percentage points in 1936, 1940, and 1944, respectively.

To conduct inference on these results, I use the approach suggested by Abadie et al. (2010) and expanded to multiple treated units by Cavallo et al. (2013). In this method, Cavallo et al. (2013) compares the mean effect across treated units to a mean of placebo effects, conducting the same analysis for each control unit as if it had been treated. Let $\bar{\alpha}$ be the mean effect of Hearst presence and $\bar{\alpha_P}$ be the mean effect for a sample of placebos. The p-value is then given by $Pr(\bar{\alpha_P} < \bar{\alpha})$, or the percent of placebo averages where the effect is more extreme (in this case, negative) than the true mean effect. The corresponding p-values across 1,000,000 placebo averages for 1936, 1940, and 1944 are 0.064, 0.001, and 0.002, respectively, making them highly unlikely to be the result of noise. Overall, this alternative approach, which perfectly matches pre-trends in Democratic vote share, yields qualitatively similar results to that of the event study analysis.

Figure C.1: Additional clippings from Hearst newspapers

A

Give Us a New Deck!

It has become the strategy of some New Deal spokesmen to speak softly about a third term for President Roosevelt, but to strongly urge a "third term for the New Deal."

Two terms of the New Deal have been more than enough to demonstrate its futility.

We are no longer content with a New Deal.

GIVE US A NEW DECK!

When a card player has had unvarying bad luck for a long time, and realizes that he has been flim-flammed and deceived, he wants not a new deal but a whole new deck of cards.

What is the use of going on playing with the same old deck of marked cards, and continuing to be flim-flammed and deceived?

A new deal with a marked and stacked deck rights no wrongs.

The country has had one long run of bad luck under the New Deal, and has been fantastically deceived, and ruinously deprived of its resources and opportunities, and the people have been brazenly and contemptuously deprived of their most precious rights and institutions.

It is only natural for the country, in these circumstances and after this experience, to call for a fresh start and a complete NEW DECK.

What folly it would be to go on putting the financial fate of the country in the hands of politicians and bureaucrats who have piled a fifty billion dollar public debt on the country!

What folly it would be to leave the solution of unemployment to an administration that has steadily INCREASED unemployment

What folly it would be to leave the search and pursuit of national recovery to an administration that has squarely and deliberately turned its back to recovery through two long and disastrous terms of reckless and futile spending, borrowing and taxation!

A third term for the New Deal would be a THIRD TERM FOR POVERTY, and a THIRD TERM FOR DEPRESSION, and a THIRD TERM FOR DEBT AND TAXATION.

It would be a THIRD TERM FOR WASTE, EXTRAVAGANCE AND SQUANDERING.

It would be a THIRD TERM FOR POLITICAL CARPET-BAGGERS, and for bureaucratic autocracy, and for politics in relief, and for government by patronage, by intimidation and by capitalization of human misery.

GIVE US A NEW DECK and a straight and honest AMERICAN SQUARE DEAL.

The country has had enough of New Deal SQUANDERMANIA.

It has had enough of experimentation and flirtation with alien ideas by crackpot idealists.

B

Through its Comintern in Moscow and its puppet organizations in the United States, Soviet Russia is taking an aggressive part in the presidential campaign in America —ON THE SIDE OF THE NEW DEAL.

The challenging assertion is based upon UNCONTRADICTABLE DOCUMENTARY PROOF.

Part of the proof is available in speeches and "literature" of the Communist Party of America.

C

Through his record and his character, Mr. Dewey appeals to Republicans and Democrats WHO BELIEVE IN PERPETUATION OF THE AMERICAN CONSTITUTIONAL SYSTEM.

On the other hand, the New Deal crowd has unified the malcontents and revolutionaries WHO WANT TO TERMINATE THE AMERICAN SYSTEM, and subtly put some totalitarian substitute in its place.

No New Deal stump speaker will delineate the issue so clearly.

To do so would be bad tactics in this year of competition for votes.

Politicians in an election year want to be all things to all men.

But it is impossible for Candidate Roosevelt to draw a flattering portrait of himself, for he will find it impossible to offset the impression created by the disreputable company he keeps.

Instead of pussyfooting on this issue, Mr. Dewey, as a progressive American and believer in balanced rights between the states and the central government, should boldly take pride in the fact that subversive and un-American elements refuse to cluster around him.

D

THE tragic injustice of the social security program imposed upon the American nation by the Roosevelt New Deal has been completely exposed in the presidential campaign.

No greater service has ever been done the people of the country than has been accomplished in this exposure.

It has been done in the face of threatened reprisals, of harassment by Federal investigators, of persecution by New Deal agencies afraid of having the people know the truth.

President Roosevelt had the astounding audacity to suggest in a campaign speech that those who disagree with his Social Security Act should get out of the country.

The President said:

"It might be well for them to move to some other nation in which they may have greater faith."

* * *

MR. ROOSEVELT is willing that Communists should stay in America, that plotters against the American form of government should stay, but he invites those who have the temerity to oppose him TO LEAVE AMERICA.

But America has been given the truth, in spite of all threats.

The Federal Government, under this scheme, is confiscating a substantial portion of the wages of twenty-six million workers.

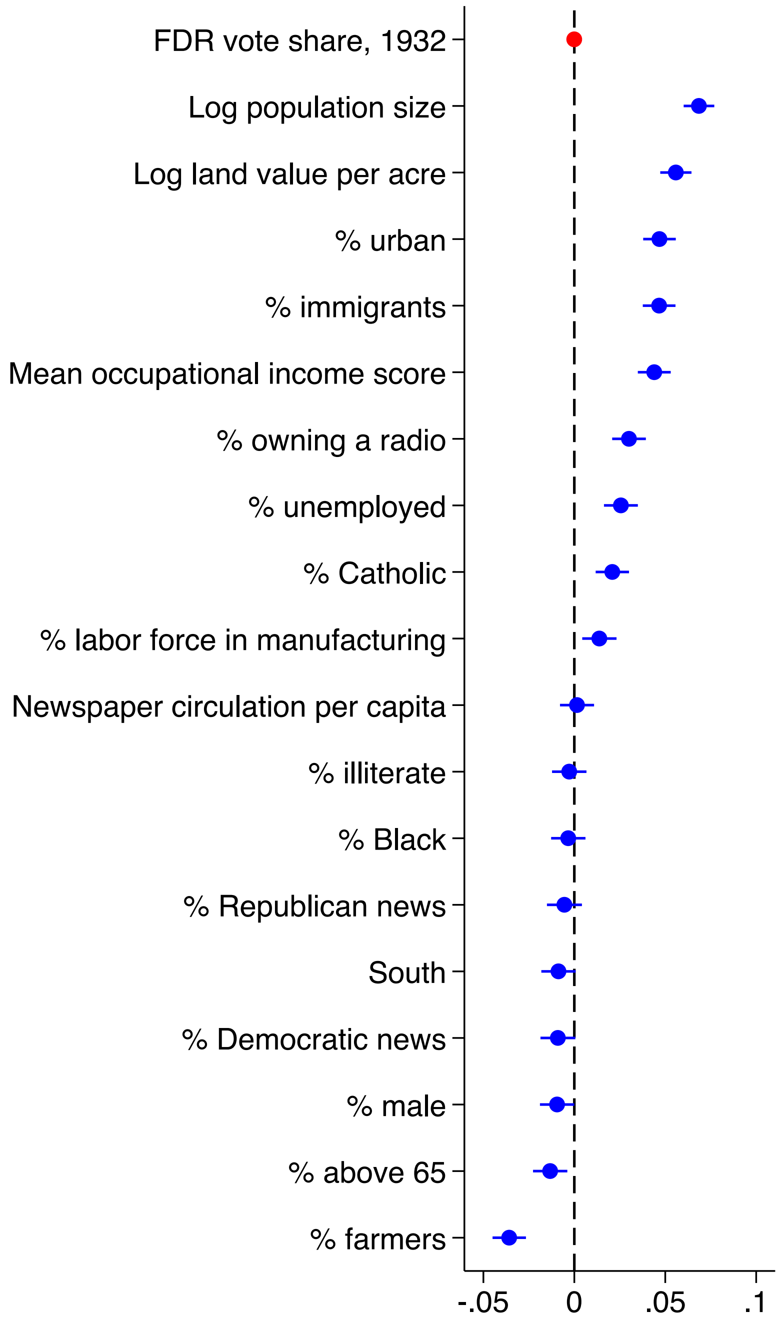
It is laying the groundwork for future unemployment, for lower wage levels, for higher prices to consumers—in short, for a depreciated standard of American living.

The cruelty goes far beyond the present needs and comforts of the people.

It is building up tragedy for them in the future, causing them to expect security in old age which the plan cannot possibly deliver.

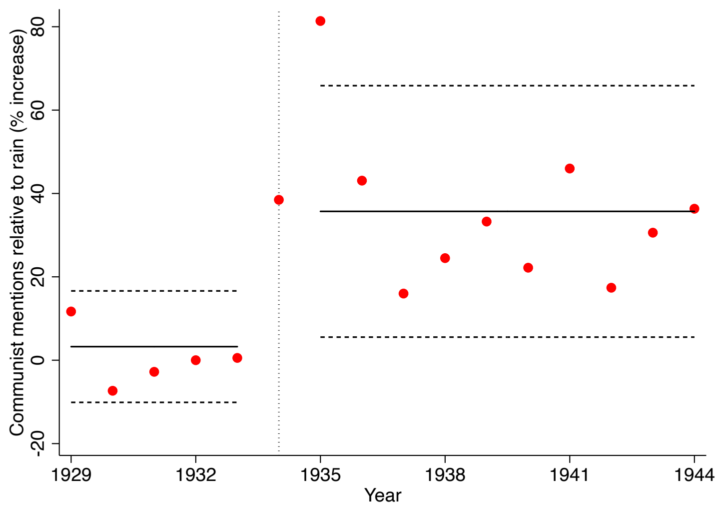
Note: This figure shows additional clippings from Hearst newspapers, courtesy of Newspapers.com. A: *Pittsburgh Sun-Telegraph,* August 23, 1939. B: *Washington Times*, September 21, 1936. C: *Pittsburgh Sun-Telegraph*, September 22, 1944. D: *San Francisco Examiner*, November 2, 1936.

Figure C.2: Characteristics that predict Hearst presence in a county



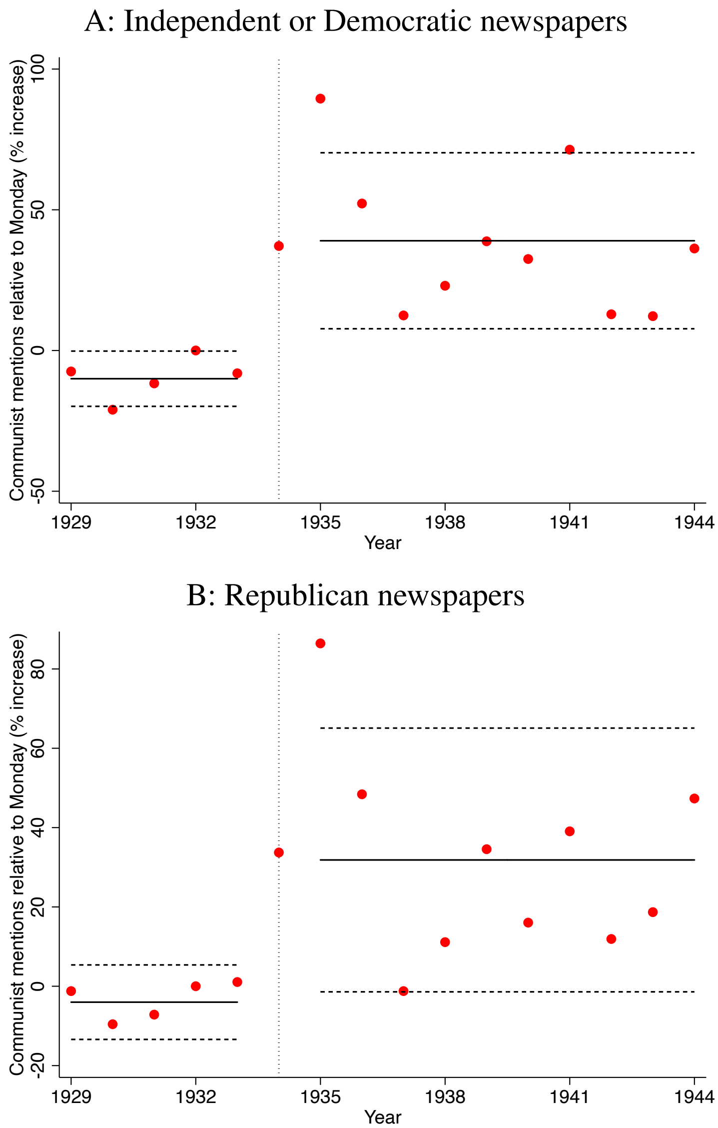
Note: This figure shows coefficients from bivariate regressions of $\mathbf{1}[\text{Hearst}_i]$, indicating that Hearst owned a newspaper in that county, on standardized county characteristics. All characteristic means come from the 1930 decennial census. Percent Democratic (Republican) news refers to the share of non-Hearst circulation with a Democratic (Republican) political affiliation. Dots shows coefficients, and lines show 95% confidence intervals.

Figure C.3: "Communist" mentions relative to "rain"



Note: This figure shows the effect of Hearst ownership on newspaper mentions of the word "communist" relative to another neutral word, rain. I divide the dependent variable by the sample mean so that coefficients represent the percent increase in mentions. Red dots show yearly point estimates, and the black lines show the average effect between 1929 and 1933 (the pre-period) and 1935 and 1944 (the post-period). The reference year is 1932. Dotted lines show 95% confidence intervals for the mean pre- and post-period effects.

Figure C.4: Comparing “communist” mentions to Republican vs. non-Republican newspapers



Note: This figure shows the effect of Hearst ownership on newspaper mentions of the word “communist,” relative to either Democratic or Independent newspapers (Panel A) or other Republican newspapers (Panel B) in propensity-score matched counties. Seven of the control newspapers in propensity-score matched counties had a Republican affiliation; eleven had an Independent or Democratic affiliation.

Table C.1: Difference-in-difference estimates with and without controls

	FDR vote share post-1932			
	(1)	(2)	(3)	(4)
Hearst circulation per capita × post-1932	0.0176	-0.214	-0.175	-0.177
	(0.066)	(0.077)	(0.084)	(0.080)
Outcome mean	0.594	0.594	0.584	0.584
Observations	3,757	3,757	940	940
Controls				
County fixed effects	Yes	Yes	Yes	Yes
County char. × year	No	Yes	No	Yes
Newspaper char × year	No	Yes	No	Yes
Sample	Full	Full	Large	Large

Note: This table shows difference-in-difference results with and without demographic and newspaper controls interacted with election year. Columns 1 and 2 show the estimate for the full sample with and without controls, respectively, and columns 3 and 4 show the estimate for the sample of large counties.

Table C.2: Convex weighted treatment effects

	FDR vote share post-1932			
	(1)	(2)	(3)	(4)
Difference-in-difference coefficient	-0.091	-0.086	-0.072	-0.078
	(0.024)	(0.041)	(0.024)	(0.042)
Outcome mean	0.594	0.594	0.584	0.584
Observations	3,757	3,757	940	940
Controls				
County fixed effects	Yes	Yes	Yes	Yes
County char. × year	Yes	Yes	Yes	Yes
Newspaper char × year	Yes	Yes	Yes	Yes
Estimation approach	Baseline	Callaway-Sant'Anna (2021)	Baseline	Callaway-Sant'Anna (2021)
Sample	Full	Full	Large	Large

Note: This table shows difference-in-difference estimates using Callaway-Sant'Anna's proposed weighting, which corrects for any potential issues with negative weights on year-specific treatment effects. In columns 2 and 4, I show results using the Callaway-Sant'Anna estimation procedure (for the full sample and the sample of large counties, respectively). As Callaway-Sant'Anna does not allow for a continuous treatment variable, in columns 1 and 3, I show equivalent results using the baseline difference-in-difference approach where the treatment is an indicator for Hearst presence in the county interacted with post-1932.

Table C.3: Additional robustness checks

	FDR vote share post-1932		
	(1) Baseline	(2) Additional controls	(3) State time trends
Hearst circulation per capita × post-1932	-0.214	-0.176	-0.143
	(0.077)	(0.074)	(0.052)
Outcome mean	0.594	0.594	0.594
Observations	3,757	3,753	3,753
	(4) Drop NYC outer counties	(5) Depression and New Deal controls	(6) Add contiguous border counties
Hearst circulation per capita × post-1932	-0.216	-0.243	-0.186
	(0.065)	(0.072)	(0.072)
Outcome mean	0.594	0.594	0.588

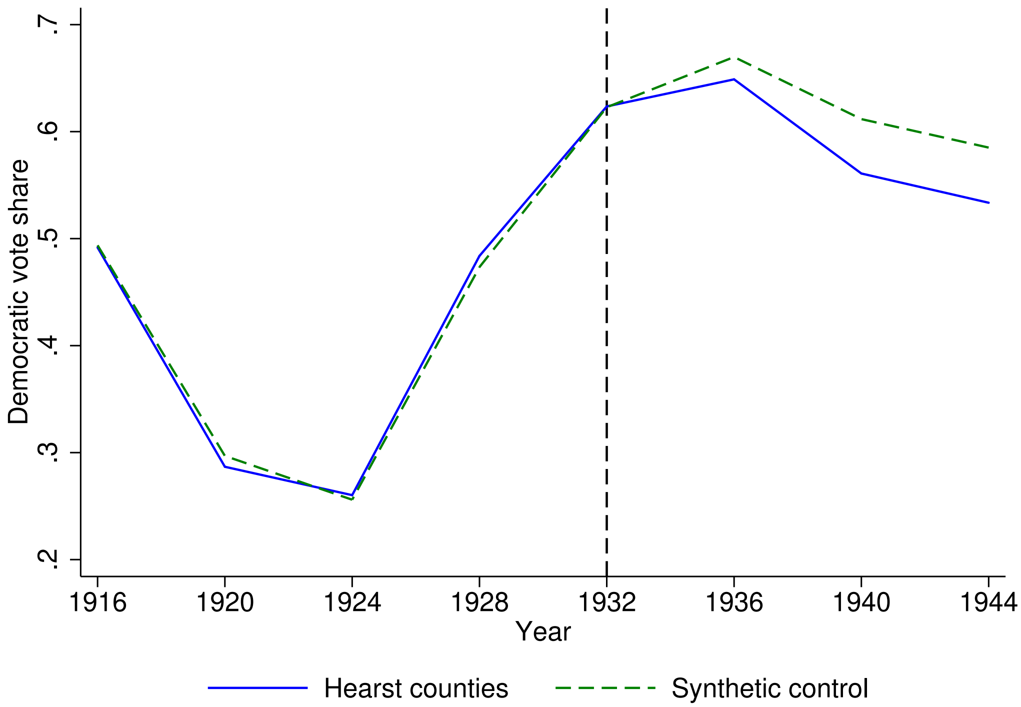
Note: This table shows difference-in-difference results when modifications are made to the sample or controls. Column 1 shows the baseline coefficient for the full sample with all baseline and newspaper controls. Column 2 adds additional county controls (interacted with year), including share of population in manufacturing, share of population in agriculture, share native born white, share of population above 65, share of population that is Catholic, mean occupationational income score, and log average farm size. Column 3 adds state × year fixed effects. Column 4 excludes non-Manhattan New York City counties, as these counties may have only been partially treated. Column 5 adds controls related to the Depression and the New Deal: growth in retail sales per capita (1929-1933), total New Deal grants, and total New Deal loans. Column 6 adds back counties that are contiguous to Hearst counties to the sample.

Table C.4: Effect of Hearst newspaper and radio presence on support for FDR

	FDR vote share post-1932			
	(1)	(2)	(3)	(4)
Hearst circulation per capita × post-1932	-0.155	-0.160	-0.157	-0.166
	(0.076)	(0.075)	(0.076)	(0.083)
Hearst radio × post-1932	-0.071	-0.068	-0.035	-0.091
	(0.036)	(0.035)	(0.027)	(0.053)
Outcome mean	0.594	0.592	0.584	0.577
Observations	3,826	3,722	936	168
Controls				
County fixed effects	Yes	Yes	Yes	Yes
County char. × year	Yes	Yes	Yes	Yes
Newspaper char. × year	No	Yes	Yes	Yes
Sample	Full	Full	Large	Matched

Note: This table shows difference-in-difference regression results with Hearst radio presence as an additional treatment variable. Hearst radio is an indicator for Hearst having a local radio station in that county. Column 1 is a regression on the full sample using baseline controls, column 2 is a regression on the full sample using additional newspaper controls, column 3 is a regression on the sample of large counties, and column 4 is a regression on the propensity-score matched sample. Standard errors, clustered by county, are shown in parentheses. I exclude contiguous border counties from the control group for both newspaper and radio counties in these specifications.

Figure C.5: Synthetic control



Note: This figure shows the results of the synthetic control analysis. In blue, I plot Democratic vote share over time for counties with at least one Hearst newspaper. In green (dotted line), I plot the Democratic vote share over time for "synthetic" counties, which are a weighted average of control counties. The algorithm matches controls to Hearst counties based on the trend in Democratic vote share prior to 1936.

Table C.5: Effect of Hearst exposure on long run attitudes

	(1) Support public healthcare	(2) Disapprove of welfare
A: Full Sample		
Exposed	-0.044	0.051
	(0.013)	(0.017)
Outcome mean	0.459	0.381
Number exposed	604	451
Observations	17,563	16,877
B: Respondents above 30		
Exposed	-0.042	0.041
	(0.021)	(0.017)
Outcome mean	0.451	0.378
Number exposed	604	451
Observations	13,103	12,665
Survey years	1970-1994	1976-1994
County fixed effects	Yes	Yes
Respondent characteristics	Yes	Yes
Survey year	Yes	Yes

Note: This table shows the effect of exposure to Hearst as an adult on attitudes towards government spending and redistribution. "Disapprove of welfare" is a combined measure of either wanting to reduce spending on welfare or feeling negatively towards those on welfare. Panel A shows results for the full sample, and Panel B shows results when restricting the sample to those above 30. Standard errors, clustered by county, are shown in parentheses.

Table C.6: Long run robustness: Red Scare ends in 1960

	(1) Support public healthcare	(2) Disapprove of welfare
Exposed	-0.047	0.049
	(0.018)	(0.015)
Outcome mean	0.459	0.381
Number exposed	688	514
Observations	17,563	16,877
Survey years	1970-1996	1978-1996
County fixed effects	Yes	Yes
Respondent characteristics	Yes	Yes
Survey year	Yes	Yes

Note: This table shows the effects of Hearst exposure on long run public opinion using a modified definition of exposure. In the baseline specification, I consider the last year of potential exposure to Hearst as the minimum between the year the newspaper was sold and 1957 (often considered the end of the Red Scare). In this specification, I more conservatively date the end of the Red Scare to 1960. Standard errors, clustered by county, are shown in parentheses.