Is There a VA Advantage?
Evidence from Dually Eligible Veterans

David C. Chan
David Card
Lowell Taylor*

November 2020

Abstract

Societies face a difficult choice between private and public provision of health care, and there is a shortage of credible evidence to guide this choice. The structure of health-care delivery to US veterans provides a distinctive research setting for the study of this issue. Specifically, veterans aged 65 and older are dually eligible for care in private hospitals (financed by Medicare) or in the pubic-sector Veterans Health Administration, the nation's largest integrated health-care delivery system (operated by the US Department of Veterans Affairs). We utilize the ambulance design of Doyle et al. (2015) to examine the effect of VA vs. non-VA emergency care on mortality in this high-risk population. We find a VA advantage: a 28-day mortality reduction of 46% (4.5 percentage points, with a 95% confidence interval of 1.1 to 8.0). Survival gains persist for at least a year after the initial ambulance ride, and they accrue despite *lower* spending in the VA. Evidence suggests that the VA advantage arises in part from some combination of continuity of care and health IT. These results have policy relevance—as the federal government is deciding whether to maintain the existing VA system or to expand finance of private care outside of the VA—and they shed light on sources of inefficiency in private-sector health care in the US.

JEL Codes: I10, I18

Keywords: VA, Medicare, mortality, dual eligibility

*Chan: Stanford University, Department of Veterans Affairs, and NBER, david.c.chan@stanford.edu; Card: UC Berkeley and NBER, card@berkeley.edu; Taylor: Carnegie Mellon University and NBER, lt20@andrew.cmu.edu. We thank Marika Cabral, Amy Finkelstein, Kate Ho, Peter Hull, Michael McWilliams, Maria Polyakova, David Silver, Todd Wagner, and Chris Walters for helpful comments and suggestions. We also thank Sophie Andrews, Sydney Costantini, Kaveh Danesh, Johnny Huynh, Kevin Kloiber, Uyseok Lee, Chris Lim, Matthew Merrigan, Kevin Todd, and Justine Weng for excellent research assistance. Finally, we thank Melissa Taylor for valuable insight with regard to our research design. The Stanford Institute for Economic Policy Research provided generous funding and support. Chan gratefully acknowledges support from VA HSR&D IIR 18-146.

1 Introduction

A key question in the design of health care systems across the world is whether care should be provided by government or by the private sector. In the US, the choice between public and private provision has become a top policy issue for the Department of Veterans Affairs (VA). Seeking to improve veteran access to health care, policymakers have debated whether the VA should expand the capacity of its own system—the Veterans Health Administration—or shift health-care delivery to private providers.

An extensive descriptive literature (e.g., Reid 2010; Blank et al. 2017) has compared health care outcomes in public vs. private systems. More generally, economists long have debated the appropriate size and role of the public sector in the economy, highlighting theoretical arguments about competitive pressure, ownership structure, and differences in the objectives and constraints in the public vs. private sector (Alchian 1965; Stigler 1965). Nevertheless, rigorous empirical evaluations of the performance of public vs. private health care providers have been relatively rare, in no small part because public and private providers of health care usually serve different patient populations, either by statute or by patient selection.

In this paper, we focus on "dually eligible" veterans aged 65 and older who can receive health care at both VA facilities and private hospitals that accept Medicare. We use the ambulance design proposed by Doyle et al. (2015) to study the causal effect of receiving emergency care at the VA vs. a non-VA facility. Our approach compares veterans sharing key characteristics—zip code of residence, prior VA and non-VA utilization, and location of pick-up (e.g., their home residence vs. a nursing home)—who receive the same dispatched level of ambulance service (i.e., advanced vs. basic life support) from different ambulance companies. Our main analytic sample includes 401,319 911-dispatched ambulance rides from 2001 to 2014, for veterans with prior attachment to the VA and in a zip code served by at least two ambulance companies. As in Doyle et al. (2015), we show that the leave-out share of dually eligible veterans transported to the VA by the assigned ambulance company is a strong predictor of hospital assignment. Under the plausible assumption that ambulances are quasi-randomly assigned within zip codes and in cells of key characteristics, this design allows us to study the effect of VA vs. non-VA emergency care on health outcomes.

We find that in the high-mortality population of elderly veterans with emergencies, there is a VA advantage—a 46% reduction in 28-mortality relative to baseline (4.5 p.p., with a 95% confidence interval of 1.1 to 8.0 p.p). Importantly, we show that our instrumental variables (IV) estimates of the

VA effect are robust to the inclusion of a long list of characteristics of the index patient and of other patients transported by the same ambulance company. The latter set of ambulance co-rider controls can account for unobserved patterns of selection across ambulance companies (Altonji and Mansfield 2018). The IV estimates are larger in magnitude than the corresponding OLS estimates, which center around 0.024 p.p., with tight confidence intervals. This difference suggests that VA "always takers" (patients who go to the VA even with an ambulance company with a low VA rate) have worse health than VA "never takers."

A critical question for interpreting the benefit of VA care is whether its mortality effects fade over longer horizons—as would happen if VA emergency care only temporarily displaces the mortality of fragile patients under "harvesting" (Schwartz 2000). To address this, we use an insight from Abadie (2002) to estimate the weekly potential death rates in the year after the initial ambulance ride among compliers, i.e., patients whose destination hospital is determined by the ambulance company. With this tool, we disentangle the short-term vs. long-term effects of the VA in the setting of competing risks. Despite a high long-term mortality rate (one in three veterans will be dead within one year of the ambulance ride), we find that the mortality impact of presenting at the VA is concentrated in the first week, suggesting VA survival gains from care addressing temporary emergency conditions. We find no evidence of harvesting, suggesting that the survival gains are long-lasting. Relying on intuition from Kitagawa (2015), we also use this potential outcomes framework to develop a sharper test of IV validity than the tests typical in the applied literature.[1] Finally, we use this framework to document small but systematic differences in underlying mortality hazards for VA always takers and VA never takers that account for differences between OLS and IV estimates of the VA advantage that we find to grow with longer time horizons.

The key potential threat to our research design is that veterans who are taken to the VA are healthier than who are taken to non-VA hospitals. This could arise if the choice of a specific ambulance company from among those that serve a given zip code is correlated with the risk of death. We present four main pieces of evidence to rule out this concern. First, we show balance in characteristics of patients assigned to companies with different propensities of taking patients to the VA. Second, we conduct an extensive analysis along the lines suggested by Altonji et al. (2005), evaluating the stability

[1] Specifically, we use the fact that, under IV validity, all indicators for potential outcomes must occur with positive probability among compliers (Balke and Pearl 1997; Kitagawa 2015). In the setting of survival, this implies that the incremental mortality risk must be positive for compliers in every week after the ambulance ride and in both VA and non-VA assignment. This prediction may fail if there are violations of monotonicity arising for example because ambulance companies with higher VA propensities are *less* likely to send veterans with certain potential mortality outcomes to the VA. Chan et al. (2019) show that this approach may detect violations in IV validity that remain hidden under standard "judges design" tests of monotonicity (e.g., Arnold et al. 2018).

of our estimates as we add controls to the models, including controls that measure the characteristics of *other* patients transported by the company. Third, we use our analysis of long-run survival patterns to show that after the first week, compliers who go to the VA have virtually the same mortality rates as those who go to non-VA hospitals—ruling out significant differences in underlying health between the two groups. Finally, in heterogeneity analyses, we show that the VA saves lives relative to non-VA alternatives regardless of whether the particular VA hospital or its non-VA alternatives have advanced capabilities (indicated by, e.g., stroke center status or trauma center level) that would be expected to induce selection of high-risk patients. In fact, more often than not, we find that patients transported to a hospital with advanced capabilities have *lower* mortality, the opposite of selection patterns revealed by studying predicted mortality.

In the final section of the paper we then turn to an evaluation of the mechanisms behind the VA advantage. We consider three broad classes of mechanisms. First, the VA might be better suited to treating conditions specific to veterans. Second, along the lines of Doyle et al. (2015), VA hospitals could achieve better outcomes by spending more. Third, better access to patient information and coordination of care may improve the productivity of VA-delivered care, particularly in high-uncertainty and high-stakes environments such as emergency care. The Veterans Health Administration is the nation's largest integrated delivery system with a longstanding health information technology (IT) system. In contrast, only 1.5% of non-VA hospitals in the US maintained a comprehensive electronic health record as of 2009 (Jha et al. 2009).

To evaluate the first explanation—that the VA system is uniquely suited to care for veterans—we note that, although we do not observe non-veterans being treated in the VA, we observe a detailed set of veteran and neighborhood characteristics that predict whether a veteran is more likely to be attached to the VA. We find a VA advantage for patients with very different patterns of prior utilization and comorbidities, although there is consistent evidence that medically needier patients (e.g., those with substance abuse and mental health problems) and those with greater attachment to the VA benefit more. We evaluate the second explanation by examining the cost of care in the VA and non-VA sectors, using information on actual spending by taxpayers and veterans. We find that VA emergency care actually costs less, *reducing* cumulative spending at 28 days by $\$2,548$ or about 21%. This suggests that the VA is more productive, achieving better outcomes at lower cost.

The third explanation centers on the idea that coordination and continuity of care in an integrated delivery system may improve health outcomes—an explanation consistent with the larger impacts of the VA on medically needy patients and those with greater prior attachment. To provide further

evidence on this, we draw on a secondary sample of veterans who have prior attachment to a specific *non-VA* hospital but no recent use of the VA. As noted, these veterans have little chance of receiving emergency care at a VA hospital but may return to the hospital they visited most in the prior year, where their records may be more easily accessed, or to another facility. Ambulance-based assignment to a patient's most-visited prior hospital (i.e., their "modal" hospital) indeed reduces 28-day mortality, though only modestly (by about 0.6 p.p.). We infer that the VA survival benefit arises from more than just repeated use of the same facility; instead, it may reflect better care coordination and/or more effective information retrieval in the VA health care system.

To probe these channels, we exploit two policy reforms that aimed to improve care co-ordination and information technology among US hospitals. In 2009 the Health Information Technology for Economic and Clinical Health (HITECH) Act stimulated a large increase in health IT adoption among non-VA hospitals, and in 2011 Medicare began experimenting with alternative payments to "Accountable Care Organizations" (or ACOs) (Blumenthal 2010; Greaney 2011). Consistent with technological and organizational changes to the non-VA system, we find that the modal-hospital survival benefit increases from a negligible effect prior to 2010 to about 1.9 p.p.—approximately one-half of the VA survival benefit—after 2010. We find evidence linking the increase in the modal-hospital survival benefit to health IT adoption but not to ACO participation.

Our findings contribute to three sets of related literature. First, the public vs. private provision of health care is a central question for the field of comparative health policy, which compares health care systems across the world to inform the design of health care systems (Blank et al. 2017). The literature in this field has been mainly descriptive.[2] Comparing the performance of health care systems is intrinsically difficult because populations differ across different countries, because many factors outside of health care contribute to health outcomes, and because health care systems usually differ in many dimensions. To our knowledge, our results provide the first quasi-experimental evidence on the effect of health care delivered by the government vs. private providers.

Second, an important literature has sought to measure the quality of care in the VA, which budgeted $84 billion for medical care in 2020.[3] Following a well-known reorganization and investment in health IT in the mid-1990s (Mccarthy and Blumenthal 2006), this literature has documented fa-

[2] As an example of the amount of material devoted to such comparative studies, the European Observatory on Health Systems and Policies (`www.euro.who.int`) produces policy commentary and "health system reviews" on the health care systems of individual countries.

[3] Spending continues to grow, in large part due to a growing veteran population with health care needs. The 2019 enacted budget allocated $77 billion for VA medical care, and the 2021 proposed budget requests $94 billion for medical care. For the last ten years, spending on medical care has nearly doubled. Summary budget numbers can be found at `https://www.va.gov/budget/docs/summary/fy2021VAbudgetInBrief.pdf`.

vorable VA quality, compared to care outside of the VA, in terms of process measures and health outcomes (e.g., Jha et al. 2003). The question of performance in the VA health care system has become particularly relevant in recent years, as the Department considers ways to improve access to care for veterans and as Congress has sought to increase private health-care delivery for veterans (113th Congress 2014; 115th Congress 2018). So far, however, this literature has mainly compared outcomes of veterans receiving care in the VA system to outcomes of non-veterans receiving care outside of the VA.

A third and very large literature studies why health care in the US appears to be a low-productivity outlier among developed countries, spending more as a percentage of GDP than any country but with poor outcomes relative to this spending (Garber and Skinner 2008). Experts have drawn attention to the fragmentation in health care financing and delivery in the US, as well as the lack of information on quality that consumers and policymakers can use to evaluate providers (Cebul et al. 2008; Cutler 2010).[4] Responding to these arguments, policymakers have incentivized adoption of health IT and coordination of care. Whether such policy levers can improve health outcomes, however, remains an open empirical question.[5] We show that the VA has lower mortality and spending, indicating higher productivity, and our results further suggest an important and previously undocumented complementarity between information technology and continuity of care. Finally, our results tentatively suggest that government incentives to adopt health IT may push private industry toward providing better health care, although a survival gap between private providers and the VA remains.

The remainder of this paper proceeds as follows. Section 2 describes the setting and data. Section 3 presents our main analysis of the VA survival benefit. Section 4 examines complier characteristics, heterogeneity in treatment effects, and long-term vs. short-term effects on mortality. Section 5 presents evidence on mechanisms driving the VA survival benefit. Section 6 discusses policy implications and concludes.

[4] Most of these arguments have been made conceptually. An empirical literature has estimated correlations across regions or across patients between levels of fragmentation and spending (e.g., Hussey et al. 2014). A literature in economics shows that physician behavior depends on the organization of health-care delivery (e.g., Gaynor et al. 2004; Chan 2016). Recently, Agha et al. (2019) have used a movers-based strategy to examine the causal effect of regions on health care spending. They find that regions with higher fragmentation cause higher spending.

[5] A recent empirical literature documents modest reductions in spending and improvements in patient satisfaction among provider forming ACOs (McWilliams et al. 2014b, 2016; Trombley et al. 2019). An older literature on health maintenance organizations (HMOs) documents impacts on spending and technology adoption in the 1990s (Baker 2001; Cutler et al. 2000), although policy analysts have noted that HMOs were primarily insurance products, not necessarily tied to providers, that focused more on limiting utilization than on improving quality (Luft 2010). Finally, a mixed literature on health IT adoption has shown health improvements in some cases (e.g., Miller and Tucker 2011) but null results in general (e.g., Agha 2014). To our knowledge, our paper is the first to assess the complementarity between health IT and continuity of care.

2 Setting and Data

2.1 The Veterans Health Administration and US Health Care

The Veterans Health Administration (VHA) of the US Department of Veterans Affairs (VA) provides health care for 9 million veteran enrollees every year. The VHA is the nation's largest integrated health-care delivery system, including 170 medical centers and more than 1,000 outpatient sites of care. The VHA had a budget of $84 billion in 2020 for medical care.

Two features of the VHA distinguish it from the rest of the US health-care delivery system. First, the VHA is owned and administered by the government, while the rest of US health care system is largely run by private parties. Second, health care is integrated at the VHA. The VHA directly employs all of its physicians and health care workers. In contrast, physicians outside of the VA are mostly independent of the hospitals at which they work and can affiliate at will with multiple hospitals. Health care in the VHA is organized by region and coordinated across inpatient, emergency department, and outpatient locations, as well as across different services and specialties of care. Following a well-known reorganization in the mid-1990s, the VHA implemented one of the first and most widely used electronic health record (EHR) systems in the US, and it continues to spend $5.7 billion yearly on its health information technology (IT) infrastructure. The VHA also spends around $800 million yearly on research and development in disease-specific areas (e.g., substance abuse, chronic disease, infectious disease), coordination of care, quality measurement and improvement, access to care, and veteran-specific concerns (e.g., suicide, homelessness).

Outside of the VA, the US health care system is marked by a high level of complexity involving multiple private and public (federal, state, and local) parties. The US spends more on health care per capita than any other country—50% greater than the second-highest country, Norway—but has lower life expectancy than most other high-income countries (Rice et al. 2013). Compared to other high-income countries, the private sector plays a greater role in the US health care system. Despite some recent reforms, health care financing in the US remains largely fee-for-service. Moreover, notwithstanding a large and well-trained workforce, as well as many advanced institutions of secondary and tertiary care, experts have noted poor coordination of care and strikingly low adoption of health IT (Cebul et al. 2008; Cutler 2010). Prior to the Affordable Care Act (ACAs), only 1.5% of US hospitals maintained a comprehensive EHR (Jha et al. 2009). In the wake of the ACA, federal policies have attempted to spur care coordination and health IT adoption in the private sector (Blumenthal 2010; Greaney 2011).

2.2 Comparing VA and Non-VA Care

Over the past decade, lawmakers have enacted major reforms that allow veterans to receive VA-funded care at private facilities (113th Congress 2014; 115th Congress 2018).[6] These reforms broaden the role of the VA to that of an *insurer* of care for veterans (similar to the role of Medicare for the elderly), with concomitant functions such as authorizing care, processing claims, and *ex post* monitoring of claims for waste and fraud.

Related to these initiatives, the quality of VA vs. non-VA care has been a longstanding subject of interest to policymakers and researchers. Almost uniformly, the health services literature has documented that the VA provides care that is of the same or higher quality than the private sector, as measured by a wide variety of process measures and health outcomes.[7] However, these comparisons are potentially confounded by differences, due to eligibility and self-selection, between the populations that utilize care in the VA and in non-VA facilities. Indeed, the vast majority of existing research has compared the care of veterans in the VA with the care of non-veterans in non-VA facilities.[8]

We use two key ideas to extend the literature on comparisons between VA and non-VA care. First, we focus on dually eligible veterans who are aged 65 and older. These veterans can receive care in the VHA and at non-VA hospitals using Medicare. A prior literature has shown that many dually eligible veterans use both types of care, particularly if they live close to both VA and non-VA facilities (Hynes et al. 2007). Second, we build on the ambulance design strategy of Doyle et al. (2015) to sidestep concerns about the endogenous selection of where to obtain care. Specifically, we study veterans who arrive at a hospital via a 911-dispatched ambulance, comparing veterans from the *same zip code* who could have obtained services from different ambulance companies with different propensities to transport patients to a VA hospital. Importantly, Doyle et al. (2015) document that the company dispatched to serve a given patient may be chosen independently of the patient's characteristics, due to rotational assignment, direct competition between available providers, or software that may consider the placement of available ambulance units at the time of the 911 call (Chiang et al. 2006; Ragone

[6] There have been additional well-funded efforts to shift care further into the private sector (Rein et al. 2018; Kefe 2018; Shulkin 2018; Gordon 2019). According to an official recommendation to the Congress-established Commission on Care, some have even proposed that "if veteran choice dictates it over time, the long term goal of the transformation is the total transition to community care" (Blom 2016).

[7] See Shekelle et al. (2010), Trivedi et al. (2011), and O'Hanlon et al. (2017) for systematic reviews. The literature includes dozens of studies on hundreds of quality of care process measures, as well as several studies on health outcomes.

[8] Two studies are noteworthy for having better identification. Nuti et al. (2016) compare outcomes for veterans in VA hospitals with outcomes for non-veterans in non-VA hospitals but restrict comparisons between VA and non-VA hospitals in the same metropolitan statistical areas. In an older study, Wright et al. (1999) look at 47,598 dually eligible veterans with a myocardial infarction. These studies find no difference or slightly better mortality outcomes in VA hospitals. Of note, a related literature suggests that veterans generally have poorer health than non-veterans (e.g., Agha et al. 2000).

2012). Ambulance companies also exhibit different tendencies to transport patients to different hospitals, based on their ownership, headquarter location, and other characteristics (Skura 2001). We further describe our quasi-experimental design and assess its assumptions in Section 3.

2.3 Data

We use data from two main sources—Medicare claims and VHA administrative data—for the universe of enrolled veterans in the VHA from the years 2000 to 2014. We observe all Medicare claims for any dually enrolled veteran. These claims data include the beneficiary's zip code and demographic information (age, race, and gender), as well as a record of medical services, as each defined by an encounter date, Current Procedural Terminology (CPT) code(s), diagnostic (International Classification of Diseases, Ninth Revision, or ICD-9) codes, and provider identity. On the VHA side, we have a complete record of clinical encounters in the electronic health record system that we transform into a corresponding set of encounter dates, CPT codes, ICD-9 codes, and provider identities.[9]

We begin by selecting ambulance ride events for dually eligible veterans, as recorded in the Medicare claims.[10] We restrict attention to "lights and sirens" emergency ambulance rides that originate from 911 dispatch calls.[11] As in Doyle et al. (2015), we extract the date of the ambulance ride and the identity of the ambulance company, based on its tax identification number (TIN). We use the ambulance company identity to develop our instrumental variable for the propensity of the ambulance company to deliver patients to the VA or to non-VA hospitals. We also extract information on interventions provided by the ambulance (e.g., intravenous fluids, intubation), the level of care (advanced life support or basic life support), the pick up location (e.g., private residence, nursing home, skilled nursing facility, accident), and the ambulance diagnosis (ICD-9) codes assigned by the ambulance personnel.

We then link these ambulance rides to emergency department (ED) visits at VA and non-VA hospitals. This constitutes our main treatment of interest. For each patient we collect information on medical conditions and outpatient, ED, and inpatient utilization over the prior year, as recorded in the Medicare claims and VHA records. We use the ICD-codes for past medical conditions to identify

[9] The VHA system includes patient home address information. However, we use the zip code information from the Medicare claims records as our source of home location, since this information is updated frequently and has been widely used in previous studies, including Doyle et al. (2015).

[10] VHA policy is that patients with outside insurance should have ambulance services paid for by that insurance. In our dually eligible population, therefore, ambulance rides will be recorded in the Medicare claims.

[11] We select ambulance rides with HCPCS codes A0322, A0328, A0330, A0362, A0368, A0370, A0427, A0429, A0433, or Q3019. We restrict to modifiers "RH", "SH", "NH", and "EH", corresponding to rides to a hospital from a residential location, a scene of an accident or acute event, a skilled nursing facility, and an extended care facility, respectively.

31 Elixhauser indices (Elixhauser et al. 1998) of comorbidities, noting the source of each condition (i.e., from visits to the VA, to non-VA facilities, or both). These comorbidities range from common conditions such as hypertension to rarer conditions such as lymphoma.

Our main outcome measure is mortality. We obtain information on exact date of death from three sources: records of inpatient deaths from the VA and Medicare claims; records of death from the Veterans Benefits Administration (VBA), and records of death from the Social Security Administration (SSA). The latter two sources are particularly reliable, as they determine whether the veteran will receive payments from either the VBA or the SSA, and draw on reports from family, funeral directors, post offices, financial institutions, other federal agencies, and state vital records agencies.

To construct our main analytical sample of 401,319 ambulance rides, we make the following restrictions, which we detail further in Appendix Table A.1. First, we remove patients who live in zip codes more than 20 miles away from the nearest VA hospital or more than 20 miles away from the nearest non-VA hospital. We also drop patients who traveled more than 50 miles from their zip code to the hospital. Second, we require that patients live in zip codes served by at least two ambulance companies with at least 20 rides, at least 5% of rides transported to a VA hospital, and at least 5% transported to a non-VA hospital. Finally, for our baseline analysis of VA vs. non-VA care, we drop veterans with no VA outpatient, ED, or inpatient care in the prior year, since fewer than 1% of these veterans are ever transported to the VA. In our secondary analysis of veterans who may be transported to modal or non-modal hospitals outside of the VA, in Section 5, we study an analogous sample of 1,414,217 ambulance rides of veterans who did not use VA care in the previous year and who live in zip codes with at least two non-VA hospitals within 20 miles. Appendix Table A.14 describes the selection process for this sample.

Table 1 describes the characteristics of the veterans and the episode associated with their ambulance ride at different steps in the creation of the main analytical sample. The average 28-day mortality rate at each step is relatively high, between 9.7 and 11.5 p.p., reflecting the illness acuity of elderly veterans who arrive by 911-dispatched ambulance. This mortality rate remains relatively stable despite the overall reduction in sample size across restrictions for our main sample. Similarly, the weekend rate, which is the proportion of ambulance rides arriving on a weekend day, is remarkably stable and close to two-sevenths, which reflects the unplanned nature of these health events (Card et al. 2009). The main impact of our sample restrictions is to increase the share of rides going to a VA hospital. In some of the steps, such as the step imposing zip-code distance restrictions, the sample becomes more concentrated in urban areas with shorter distances to nearby VA and non-VA

hospitals. Blacks also comprise a larger share of the sample. Patient characteristics otherwise remain stable across sample restriction steps.

3 Benchmark Analysis

3.1 Quasi-Experiment

Following Doyle et al. (2015), our main empirical strategy relies on the assignment of ambulances to patients in emergencies and the role of ambulance companies in determining the hospital that a patient is taken to for emergency care. As detailed in Doyle et al. (2015), several companies typically serve the same narrow geographic area. The assignment of a particular company may be quasi-experimentally determined such that the identity of the assigned ambulance company is plausibly unrelated to patient characteristics. Furthermore, ambulance companies exhibit "preferences" for delivering patients to certain hospitals, due to their ownership or the location of their operations.

We define conditioning sets within which ambulance assignment may be as good as random. First we condition on the origin zip code $z(i)$ of ambulance ride i, so that we compare patients from the same zip code who are picked up by different ambulance companies. Second, we categorize the ambulance by whether it offers advanced life support (ALS) or basic life support (BLS) based on ambulance HCPCS codes. We further categorize rides by the pickup site category (e.g., residential address, nursing home, scene of an accident), the day of the week, and month-year interactions (e.g., January 2010). Finally we condition on measures of the patient's primary care, ED, and inpatient utilization over the past year at VA and non-VA facilities.[12] For simplicity we refer to the joint set of controls for the type of ambulance, pickup site, date of pickup, and patient prior utilization as $\mathbf{X}_i^0$.

Unlike Doyle et al. (2015), we do not include patient demographics, prior medical conditions, or ambulance diagnoses in the set of baseline controls. Instead, we "hold out" these variables—many of which are highly predictive of mortality—and show that they are balanced across local ambulance companies with differing propensities to send patients to the VA, conditional on $\left(z(i), \mathbf{X}_i^0\right)$.

Our treatment of interest is delivery to a VA hospital, which we denote by the indicator $D_i \in \{0, 1\}$ for ambulance ride i. Ride i is provided by company $j(i) \in \mathcal{J}_{z(i)}$,where $\mathcal{J}_z$ is the set of companies that serve zip code z.[13] Associated with each ride and company is a potential treatment indicator $D_i(j)$;

[12]The latter set of prior utilization measures may capture ambulance service areas within large zip codes, which may in turn account for correlations between prior use of VA vs. non-VA care and the identity of ambulance companies.

[13]We define ambulance companies an "ambulance company" as the interaction between an ambulance company tax identification number (TIN) and the health referral region (HRR) of the ride. This accounts for a few large corporations with a single TIN that serve multiple regions.

thus $D_i = D_i(j(i))$. Our main outcome is 28-day mortality of the patient, denoted by $Y_i \in \{0,1\}$. The associated potential outcomes, $Y_i(d)$, depend on whether the patient was transported to a VA hospital ($d = 1$) or not ($d = 0$), with $Y_i = Y_i(D_i)$

Under the assumptions that different ambulance companies have systematically different tendencies to transport patients to the VA, and that the assignment of $j(i)$ is as good as random, conditional on $\left(z(i), \mathbf{X}_i^0\right)$, the identity of the ambulance company can be used to construct a a valid instrumental variable for D_i. More formally, we consider the following conditions for IV validity (Imbens and Angrist 1994), which we assess in Section 3.2:

Condition 1 (IV Validity). *For a random sample of ambulance rides i provided by ambulance companies j, the following conditions hold:*

(i) Relevance: $E\left[D_i(j) \middle| z(i), \mathbf{X}_i^0\right]$ *is a nontrivial function of* $j \in \mathcal{J}_{z(i)}$.

(ii) Independence and Exclusion: *The vector of potential outcomes,* $(Y_i(0), Y_i(1), D_i(j))$*, is independent of the assigned ambulance company,* $j(i) \in \mathcal{J}_{z(i)}$*, conditional on* $\left(z(i), \mathbf{X}_i^0\right)$.

(iii) Monotonicity: *Conditional on* $\left(z(i), \mathbf{X}_i^0\right)$*, for any* j *and* j'*,* $D_i(j) \geq D_i(j')$ *for all i, or* $D_i(j) \leq D_i(j')$ *for all i.*

As is standard in the judges-design literature (e.g., Kling 2006, Dahl et al. 2014), to deal with finite samples, we construct a leave-out (or jackknife) instrumental variable that reflects the propensity of the ambulance company $j(i)$ assigned to ride i to transport *other* patients to the VA. We compute this as the average fraction of other patient who were picked up by company $j(i)$ and went to the VA. Specifically, for ambulance ride i transporting patient $k(i)$ we define the leave-out probability Z_i of transport to the VA:

$$Z_i = \frac{1}{K_{j(i)} - 1} \sum_{i' \in \mathcal{I}_{j(i)}} \frac{\mathbf{1}(k(i') \neq k(i)) D_{i'}}{N_{k(i'),j(i)}}, \tag{1}$$

where K_j is the total number of patients transported by company j , $N_{k,j}$ is the total number of rides taken by patient k with company j, and $\mathcal{I}_j$ is the set of rides transported by ambulance company j. We estimate Z_i using the sample of dually eligible veteran ambulance rides described in Column 1 of Table 1.

Under Condition 1, an IV estimate based on Z_i, conditioning on $\left(z(i), \mathbf{X}_i^0\right)$, recovers a local average treatment effect (LATE) of the VA on mortality among compliers (i.e., the set of rides i such that $D_i(j) > D_i(j')$ for some $j, j' \in \mathcal{J}_{z(i)}$). For comparison, we also consider the observational "treat-

ment effect" of going to the VA on mortality of patients who arrive at hospital in a 911-dispatched ambulance, controlling for $\left(z\left(i\right), \mathbf{X}_i^0\right)$:

$$Y_i = \beta D_i + \mathbf{X}_i^0 \delta_0 + \zeta_{0,z(i)} + \varepsilon_{0,i} \tag{2}$$

where $\zeta_{0,z}$ represents an unrestricted effect for rides originating in zip code z. Estimating Equation (2) by OLS yields $\hat{\beta}_{OLS}$, while instrumenting D_i with Z_i yields $\hat{\beta}_{IV}$. The gap between $\hat{\beta}_{OLS}$ and $\hat{\beta}_{IV}$ will depend on differences in the potential outcomes between never takers (i.e., patients who go to a non-VA facility regardless of the ambulance company) and always takers (i.e., patients who go to the VA regardless of the ambulance company), as well as on differences in treatment effects between compliers and non-compliers. We explore this gap more directly in Section 4.

3.2 First Stage, Balance, and Reduced Form

We begin our empirical analysis by demonstrating instrument relevance, Condition 1(i), with the following first-stage regression:

$$D_i = \pi_1 Z_i + \mathbf{X}_i^0 \delta_1 + \zeta_{1,z(i)} + \varepsilon_{1,i}. \tag{3}$$

The coefficient π_1 reflects the impact of ambulance company preferences on the probability that the ride goes to the VA, conditional on our baseline controls for ambulance type, pickup site, zip code, date categories, and veteran prior utilization. Figure 1, Panel A, shows a binned scatterplot of residualized D_i on the y-axis with respect to residualized Z_i on the x-axis and reports $\hat{\pi}_1 = 0.882$ (s.e. 0.034). The first-stage relationship between D_i and Z_i is very predictive and close to linear.

To assess independence, Condition 1(ii), we test whether Z_i is correlated with patient characteristics that are correlated with mortality. Specifically, we construct an estimate of predicted mortality $\hat{Y}_i$ using "hold-out" patient characteristics, including patient demographics and 31 Elixhauser indices for prior medical conditions.[14] We then fit models for $\hat{Y}_i$ based on the same right-hand-side specification as in Equation (3). Panel B of Figure 1 shows (with hollow dots) that there is no relationship between $\hat{Y}_i$ and Z_i, controlling for $\left(z\left(i\right), \mathbf{X}_i^0\right)$. In contrast, the same panel shows (with solid dots) that

[14]Patient demographics include age, gender, and race and ethnicity. Age is captured two-year age bins from 65 years to 100 years. Race and ethnicity is captured with three dummies for white, Black, and Hispanic; the omitted category is Asian/other. We use the 31 Elixhauser indices as described in Elixhauser et al. (1998), interacting each index with the source of the record indicating the comorbidity. There are three possible sources: VA only, Medicare claims only, and VA and Medicare claims. This results in $3 \times 31 = 93$ dummies. Hold-out patient characteristics are described in Appendix Table A.3.

the reduced-form relationship between actual mortality, Y_i, and Z_i is significantly negative, under the same controls. Specifically, for the reduced-form relationship,

$$Y_i = \pi_2 Z_i + \mathbf{X}_i^0 \delta_2 + \zeta_{2,z(i)} + \varepsilon_{2,i}, \tag{4}$$

we find $\hat{\pi}_2 = -0.040$ (s.e. 0.016). This suggests that quasi-random assignment to an ambulance company more likely to transport to the VA results in an intention-to-treat reduction in mortality.

The exclusion condition in Condition 1(ii) asserts that ambulance companies do not affect outcomes other than through their effect on whether a patient arrives at a VA or non-VA hospital. Implicitly, however, our notation also assumes that each complier has a well-defined non-VA hospital that is stable across ambulance companies. In Appendix A.1.1, we evaluate the robustness of our results to potential violations of the strict exclusion assumption implicit. Specifically, we assess and find no evidence of any correlation between Z_i and ambulance treatments captured in summary charges or between Z_i and ambulance propensities to deliver patients to different non-VA hospitals.[15]

To assess the monotonicity assumption given by Condition 1(iii), we follow the standard practice in the judges-design literature to show that the first-stage relationship between D_i and Z_i remains positive for subgroups of patients defined by different observable characteristics (e.g., Arnold et al. 2018; Bhuller et al. 2020). We detail these analyses in Appendix A.1.2. In Section 4, we present a stronger test of monotonicity (and IV validity) based on *potential outcomes*. Following the reasoning in Kitagawa (2015), this test amounts to showing a positive density for the potential outcome of death in a given week, conditional on survival to the end of the previous week, for compliers who go to VA or non-VA facilities.

3.3 Mortality Effect

With this background, we now move to our main results on patient mortality. In Table 2, we show both OLS and IV estimation results for Equation (2). Panel A of the table shows $\hat{\beta}_{OLS}$ from Equation (2), while Panel B shows $\hat{\beta}_{IV} = \hat{\pi}_2/\hat{\pi}_1$ from the first-stage and reduced-form regressions in Equations (3) and (4). Column 1 shows our baseline specification, controlling for zip code and the variables in $\mathbf{X}_i^0$. The OLS estimate is $\hat{\beta}_{OLS} = -0.024$ (s.e. 0.001), while the IV estimate is $\hat{\beta}_{IV} = -0.045$ (s.e.

[15] Following Kolesar et al. (2015), these analyses correspond to the weaker assumption that there are no systematic correlations between our instrument and other ambulance-specific treatments that impact our outcome of interest. Specifically, under this weaker version of exclusion, we require that ambulance companies with higher values of $E\left[D_i(j)\middle|z(i), \mathbf{X}_i^0\right]$ do not also systematically apply treatments during the ambulance ride that affect mortality, or systematically deliver patients to higher- or lower-quality non-VA alternatives.

0.018).[16] Relative to the mean 28-day mortality of 9.7 p.p., both estimates imply a sizeable reduction in mortality for compliers who are taken to the VA.

The other columns in Table 2 show OLS and IV estimates as we include additional controls to the models: (i) patient demographics (age, race, gender), (ii) ambulance diagnostic (ICD-9) codes, (iii) Elixhauser comorbidity indicators, and (iv) ambulance and co-rider controls, which are all described in Appendix Table A.3. The latter controls are meant to capture any unobservable patient selection at the ambulance company level by using characteristics of *other* rides and patients under the same ambulance company, following the reasoning in Altonji and Mansfield (2018). Specifically, these controls address the potential concern that sicker patients tend to be allocated to certain ambulance companies that may be more or less likely to take patients to the VA.

Reassuringly, both $\hat{\beta}_{OLS}$ and $\hat{\beta}_{IV}$ remain stable as we add additional controls. Figure 2 illustrates this stability as controls are added in a more granular fashion, and Appendix Figure A.2 shows stability of the IV estimates as we permute the order in which the extra controls are added. The stability of both the OLS and IV estimates suggests lack of selection on observable characteristics; under the reasoning of Altonji et al. (2005), this stability suggests limited scope for selection on unobservable characteristics that predict potential 28-day mortality. However, the substantive gap between the OLS and IV estimates, with IV being larger in magnitude, suggests either that never takers are healthier than always takers (i.e., selection runs counter to treatment effects on mortality) or that the LATE among compliers is larger than the unconditional average treatment effect (ATE).[17] We investigate these possibilities in the next section and in Section A.4.

4 Survival Analysis

In this section we develop and apply a survival analysis framework to understand the dynamics of potential survival outcomes following the ambulance ride. Rather than focusing exclusively on mortality at 28 days, as in Section 3, we broaden our analysis to understand how mortality events unfold over time. We use this framework to make several insights. First, we determine the time course of VA effects on mortality. Second, we use the empirical results of this framework to provide further

[16] Appendix Figure A.1 shows the IV estimate visually, by plotting the predicted first-stage probability of treatment from Equation (3) on the x-axis and predicted reduced-form effect on mortality from Equation (4) on the y-axis. The slope of this visual IV relationship corresponds to $\hat{\beta}_{IV} = -0.045$.

[17] We note that a Hausman test for equality of the two estimates has a t-statistic of only 1.0, so based on this evidence alone, the gap between OLS and IV could be simply due to sampling error. In the next section, however, we show a dynamic pattern of IV and OLS estimates, over the year after the initial ambulance ride, that points more definitively to systematic differences.

validation of Condition 1, beyond the standard benchmark analysis in Section 3.2. Third, we investigate the implications of heterogeneity in mortality risks between compliers and non-compliers of our ambulance quasi-experiment.

4.1 Approach

Consider a set of potential survival outcomes $S_i(t;d) \in \{0,1\}$ under VA care ($d=1$) and non-VA care ($d=0$) for each week $t \in \{1,\ldots,52\}$ following the ambulance ride.[18] By definition, if $S_i(t;d) < S_i(t-1;d)$, then the patient in ambulance ride i would die in the t^{th} week following the ambulance ride if exposed to treatment d. Of course, potential survival outcomes must weakly decrease over time, i.e., $S_i(t;d) \leq S_i(t-1;d)$ for all i, d, and t.

As with mortality outcomes, for each ambulance ride i, we can only observe the set of survival outcomes corresponding to $d = D_i$: $S_i(t) = D_i S_i(t;1) + (1-D_i) S_i(t;0)$. However, appealing to Abadie (2002), we can recover the expected survival outcomes for the set of compliers C whose hospital choice depends on which ambulance company picks them up. In particular, under Condition 1, we can estimate $s_{IV}(t;1) \equiv E[S_i(t;1) | i \in C]$ by two-stage least squares using the first-stage Equation (3) and a reduced-form equation similar to Equation (4) but with dependent variable $S_i(t) D_i$. Similarly, we can estimate $s_{IV}(t;0) \equiv E[S_i(t;0) | i \in C]$ using the same first stage model but replacing the reduced-form outcome variable in Equation (4) with $S_i(t)(D_i - 1)$. Note that by construction, the IV estimand of the VA treatment effect on 28-day mortality in Section 3, satisfies

$$\beta_{IV} = s_{IV}(4;1) - s_{IV}(4;0).$$

Given the potential survival outcomes, we can then estimate potential hazard rates for mortality, under VA and non-VA assignment:

$$\begin{aligned} h_{IV}(t;d) &\equiv E[1 - S_i(t+1;d) | S_i(t;d) = 1, i \in C] \\ &= \frac{s_{IV}(t;d) - s_{IV}(t+1;d)}{s_{IV}(t;d)}, \end{aligned} \tag{5}$$

for $d \in \{0,1\}$ and $t \in \{1,\ldots,52\}$, corresponding to weekly mortality hazard rates up to one year after the initial ambulance ride. Under Condition 1, differences between $\{h_{IV}(t;1)\}_t$ and $\{h_{IV}(t;0)\}_t$ can be interpreted as the causal effect of VA assignment, among compliers, on the set of mortality

[18] We adopt the convention that a death within the first 7 days is a death in week 1. Thus a death within 28 days is a death by the end of week 4.

hazard rates.[19]

As in Section 3, we calculate conditional risk-adjusted OLS survival functions and mortality hazard rates. We estimate $s_{OLS}(t;d) \equiv E[S_i(t;d)|D_i = d] = E[S_i(t)|D_i = d]$ by OLS of Equation (2), replacing the outcome variable with $S_i(t)D_i$ for $s_{OLS}(t;1)$ and with $S_i(t)(D_i - 1)$ for $s_{OLS}(t;0)$. Our OLS estimand of the VA effect on 28-mortality, β_{OLS}, is similarly equal to $s_{OLS}(4;1) - s_{OLS}(4;0)$. Corresponding mortality hazard rates can also be calculated based on observed risk-adjusted survival:

$$\begin{aligned} h_{OLS}(t;d) &\equiv E[1 - S_i(t+1;d)|S_i(t;d), D_i = d] \\ &= \frac{s_{OLS}(t;d) - s_{OLS}(t+1;d)}{s_{OLS}(t;d)}. \end{aligned} \tag{6}$$

Compared to the potential survival functions and mortality hazards , the OLS analogues also incorporate outcomes for the always takers and never takers whose choice of hospital is unaffected by the specific ambulance company that picked them up. Specifically, $s_{OLS}(t;1)$ and $h_{OLS}(t;1)$ reflect survival outcomes for a combination of always takers and compliers, while $s_{OLS}(t;0)$ and $h_{OLS}(t;0)$ reflect survival outcomes for a combination of never takers and compliers.

4.2 Time Course of Mortality Effects

Since we examine potential survival outcomes one year after an ambulance ride, for the analysis in this section we restrict analysis to ambulance rides of patients with no prior ride within one year.[20] In Figure 3, we show our estimated potential survival curves and potential hazard rates in weeks 0 to 52 for compliers assigned to the VA and those assigned to a non-VA hospital. The potential survival curves, shown in Panel A, reveal a high risk of mortality among compliers. Mortality at 28 days among compliers assigned to a non-VA hospital is greater than the sample mean of 9.7 p.p., and cumulative mortality at one year is approximately 30 p.p. However, despite the substantial mortality risk over the subsequent year, the gap in survival between VA- and non-VA-assigned compliers (i.e., the mortality treatment effect) is fully realized at 28 days and remains stable for the rest of the year.

[19] We emphasize that any gap between $h_{IV}(t,1)$ and $h_{IV}(t,0)$ at some later time horizon (e.g., $t = 12$) could arise because treatments at the VA affected the population of compliers who survive to week $t-1$ and are therefore at risk of death in week t, or because of a treatment effect on the week t hazard, holding the population fixed.

[20] This restriction attributes survival for a given patient in a given week to the "upstream" ambulance ride, rather than attributing the survival event to both upstream and downstream ambulance rides. This changes (decreases) the sample in Appendix Table A.1 to 254,782 rides and 188,299 patients. In Appendix Figure A.8, we show that this restriction (or any other restriction on prior rides) does not lead to qualitative differences in our estimated OLS or IV treatment effects on mortality over time. Qualitatively, regardless of the number of days within which we require no prior ride, the IV estimates are larger than 4 p.p. at 28 days and remain mostly stable within the year following the ambulance ride; the OLS estimates are between 2.0 and 2.5 p.p. at 28 days and essentially disappear by one year following the ambulance ride. We evaluate the implications of the long-term difference between IV and OLS treatment effects in Section 4.4.

In Panel B, we examine the implied hazard rates and show that the differences in mortality are concentrated in the first week following the ambulance ride. Thereafter the hazard rates for both VA- and non-VA-assigned compliers remain relatively high and are indistinguishable from each other. This similarity suggests that the 4.5 p.p. VA reduction in mortality at 28 days in our benchmark analysis results entirely from events within the first week following the ambulance ride.

The potential hazard profiles in Figure 3 suggest that mortality risks for the compliers in our analysis comprise two separate risks: (i) a relatively high short-term risk component that is affected by VA vs. non-VA assignment, and (ii) a relatively stable long-term risk component that is the same between compliers who go to the VA and those who go to a non-VA hospital. If the latter risk reflects underlying patient health and is independent of the risk that led to the ambulance call, then we would expect the long-run weekly mortality rate (after, e.g., three months) to be the same for veterans who were quasi-randomly assigned to VA and non-VA hospitals; we formalize this as a test in Section 4.3.

The potential hazard rates also allow us to assess whether excess mortality at non-VA hospitals involves "harvesting," or mortality displacement, in which deaths for patients at the VA are simply delayed (Schwartz 2000; Honore and Lleras-Muney 2006). Under this hypothesis, survival gains from VA care observed at 28 days are temporary and will fade in the long-term. Such mortality displacement would imply that the hazard of dying *increases* among VA-assigned compliers after a time. We find no evidence of this in the potential hazard rates in Panel B of Figure 3. In Appendix A.2, we formally test that $h_{IV}(t;1) \leq h_{IV}(t;0)$ for all t and cannot reject this null hypothesis of no harvesting.[21] This suggests that the VA *prevents* rather than *displaces* deaths, leading to a persistent survival benefit. Visually, this is confirmed by the fact that the gap between the potential survivor functions in Panel A is very stable after 28 days.

4.3 Extended IV Validity

We can also use the estimated potential survival outcomes to test the validity of our IV strategy based on ambulance assignment. Under Condition 1, the density of any characteristic, including characteristics defined by potential outcomes, must be positive among compliers of the quasi-experiment (Balke and Pearl 1997; Imbens and Rubin 1997):

$$\Pr(X_i = x, Y_i = y \,|\, i \in C) \geq 0, \tag{7}$$

[21] Our test builds on the suggestion of Wolak (1987) to form a test statistic based on a quadratic form that represents the deviations of the data from the predictions of a constrained model that imposes the inequality restrictions. We use a simple bootstrap procedure to derive critical values of the test.

for all possible characteristics $x \in \mathcal{X}$ and all possible potential outcomes $y \in \mathcal{Y}$. Kitagawa (2015) proposes a formal test of this implication, and Chan et al. (2019) show that applying this test to *potential outcomes* can provide a stronger test of the conditions for IV validity, particularly the monotonicity assumption in Condition 1(iii).[22]

In our setting, we partition survival potential outcomes into weeks of potential mortality for 52 weeks following the ambulance ride, for both VA- and non-VA-assigned compliers. Since survival can only decrease over time, the potential mortality hazard rates for any week must be positive (i.e., $h_{IV}(t;d) \geq 0$ for all $t \in \{1,\ldots,52\}$, $d \in \{0,1\}$). A complier survival rate that appears to increase (i.e., that $h_{IV}(t;d) < 0$ for some t or d) could arise if patient risk of death in some week t is correlated with the assigned ambulance's propensity to take patients to the VA (a violation of independence), or if ambulances with different overall propensities to transport patients to the VA also choose different groups of *inframarginal* patients, who systematically differ in their risk of death in some week t, to take to the VA (a violation in monotonicity). In Appendix A.2, we formally test the joint inequality constraint that $h_{IV}(t;d) \geq 0$ for all $t \in \{1,\ldots,52\}$, $d \in \{0,1\}$, and cannot reject this null hypothesis, with a bootstrap-based p-value of 1.00.

If the short-term and longer-term mortality risks facing veterans are independent (as is typically assumed in a competing risks model) and treatment at the VA only affects the short-term risk component, then Condition 1 also implies that $h_{IV}(t;1) = h_{IV}(t;0)$ for $t \geq \bar{t}$, for some $\bar{t}$ after the acute ambulance episode. Specifically, if the short-term risk component dies out after some time $\bar{t}$, and if the assignment of compliers to VA and non-VA hospitals is as good as random, then the death rates of the two groups of compliers should be the same after $\bar{t}$. Visually, it appears that the potential hazard rates of the compliers are very similar in weeks $t \in \{2,\ldots,52\}$. Consistent with this impression, in Appendix A.2, we show that we cannot reject that $h_{IV}(t;1) = h_{IV}(t;0)$ for all weeks $t \geq 2$, with a bootstrap-based p-value of 0.31.

4.4 Heterogeneity in Mortality Risks

Finally, we take a closer look at death rates during the year after the ambulance ride to better understand the differences between our main OLS and IV estimates of the VA advantage. As shown in

[22] Specifically, testing Equation (7) with respect to potential outcomes $y \in \mathcal{Y}$ may be more likely to detect violations of Condition 1 than standard tests of monotonicity, focusing on patient characteristics, that we employ in Appendix A.1.2. The intuition behind this is that testing Equation (7) with respect to potential outcomes will reveal violations in Condition 1 that relate not only to observed patient characteristics but also to unobserved patient characteristics. Violations in quasi-random assignment or monotonicity may be more likely to occur along potential outcomes if agents act according to an objective function based on potential outcomes.

Panel A of Figure 4, we find that, remarkably, OLS survival curves cross about nine to ten months after the ambulance ride. This reflects a reversal in the sign of the OLS-estimated VA treatment effect: While patients arriving at the VA experience an immediate survival benefit that peaks at 14 days after the ambulance ride, the survival benefit eventually reverses, such that patients arriving at the VA are *more* likely to die within a year.

Consistent with this observed survival pattern, Panel B of Figure 4 reveals a cross-over in the observed hazard rates of death for patients who are taken to VA and non-VA hospitals. In the first week after the ambulance ride, the death rate is lower for patients at the VA, though the gap between the VA and non-VA hazards is smaller than the corresponding potential-outcomes gap for compliers shown in Figure 3. Thereafter, the hazard rate for patients at the VA, $h_{OLS}(t;1)$, is consistently higher than the hazard rate for those who went to a non-VA hospital, $h_{OLS}(t;0)$. This gap suggests differences in baseline risk between always takers and never takers that are initially offset by the short-term VA advantage but reemerge, soon after the first week. While these differences in baseline mortality hazards may be slight on a weekly basis, they accumulate over time to generate large differences in long-term survival.

To identify differences in the baseline mortality risk between VA-assigned compliers and always takers, we compare $h_{IV}(t;1)$ and $h_{OLS}(t;1)$; to identify differences between non-VA-assigned compliers and never takers, we compare $h_{IV}(t;0)$ and $h_{OLS}(t;0)$. In Appendix A.2 we show that we cannot reject the null hypothesis that $h_{IV}(t;1) = h_{OLS}(t;1)$ for $t \geq 2$. However, we can strongly reject the null hypothesis that $h_{IV}(t;0) = h_{OLS}(t;0)$ for $t \geq 2$. This implies that never takers are healthier than compliers who are assigned to non-VA facilities. Moreover, the average value of $h_{IV}(t;0)$, for $t \geq 2$, is significantly larger than the corresponding average value of $h_{OLS}(t;0)$, for $t \geq 2$. This illuminates why the OLS estimate of the VA survival effect is downward biased in magnitude compared to the LATE, a finding first suggested in our benchmark comparison between $\hat{\beta}_{IV}$ and $\hat{\beta}_{OLS}$ in Section 3.3 albeit with less precision. Since the difference $h_{IV}(t;0) - h_{OLS}(t;0)$ is positive throughout our one-year observation window, the OLS bias increases in magnitude the longer the time horizon, eventually causing the sign of the OLS-estimated survival effect to reverse.

5 Mechanisms

This section probes further into the mechanisms behind the large VA mortality advantage. We first examine characteristics of compliers in our quasi-experiment. Second, we use a simple Olsen (1980)

control function approach to estimate the average treatment effect (ATE) and compare it with the LATE estimated in Section 3. Third, we examine treatment effect heterogeneity by hospital and patient characteristics. Fourth, we ask whether the VA produces superior health outcomes by spending more; spending less would imply mechanisms that improve productivity. Fifth, we perform an analysis of health IT and integrated care among veterans using non-VA care, as potential mechanisms that may set the VA apart from non-VA care.

5.1 Complier Characteristics

We perform a standard complier analysis examining characteristics of compliers relative to the overall sample.[23] Table 4 shows results for various characteristics. Compliers are more likely to be Black, to have lower income, to have a prior VA ED visit, and to suffer from mental illness and substance abuse. Compliers have slightly fewer recorded Elixhauser comorbidities and are less likely to receive Advanced Life Support (ALS); they are also marginally younger and have marginally lower predicted mortality. In Appendix Table A.7, we show similar patterns comparing always takers and never takers, following an approach in Dahl et al. (2014) that we describe in Appendix A.3.

Researchers and policymakers have noted greater incidence of mental health and substance abuse issues among veterans (Adamson et al. 2008). Recognizing this need, Congress allocated $152 million for increasing mental health care programming in 1999; in the following two decades, VHA stations expanded mental health services and hired thousands of mental health providers (106th Congress 1999; U.S. Government Accountability Office 2015). This capacity to treat mental health disorders contrasts with the non-VA health care sector, where mental health services have long been underfunded and underprovided (Huskamp and Iglehart 2016).

5.2 Selection Model

We consider a structural model of selection in order to assess how VA treatment effects vary with a veteran's propensity to go to the VA and to infer the ATE. Following the "marginal treatment effects" (MTE) terature (see, e.g., Heckman and Vytlacil (2007) for a review), we exploit our multivalued ambulance instrument in order to characterize the relationship between treatment effects and veterans who are induced to go to the VA.

[23] Specifically, we employ the same approach from Abadie (2002) that we introduced in Section 4.1. Under IV validity in Condition 1, we can estimate $E\left[X_i | i \in C\right]$ for some characteristic X_i by two-stage least squares, involving the first-stage Equation (3) and a reduced-form equation replacing the outcome variable in Equation (4) with $X_i D_i$.

Specifically, we allow for a flexible slope in the returns to VA care among compliers who are induced into VA care, ranging from ambulances with low propensities to deliver patients to the VA to those with high propensities. Using a control function model, we also extrapolate this relationship in treatment effects to always takers and never takers, thereby imputing the ATE from the semiparametric structure of the model. We provide further details of our approach in Appendix A.4.

We find evidence of moderate "selection on gains," in which veterans with larger mortality reductions from going to the VA are more likely to go to the VA. In Appendix Figure A.7, we show the MTE function ranging from veterans who are most likely to use the VA to those who are least likely to use the VA. Veterans induced to go to the VA by lower-propensity ambulances have higher returns to VA care than veterans who are induced by high-propensity ambulances. In Appendix Table A.8, we find a substantial ATE, only marginally smaller than the LATE, across a variety of specifications.

5.3 Heterogeneity by Hospitals and Patients

We next assess heterogeneity in the VA mortality effect by hospitals in a patient's choice set and by patient characteristics. We consider a wide range of characteristics, in three categories: (i) characteristics of non-VA hospitals serving a given zip code, weighting the hospitals by volume of patients from the zip code; (ii) characteristics of the VA hospital serving a given zip code; and (iii) patient characteristics.

For each of these characteristics x, we construct a binary indicator variable, $I_{x,i} \in \{0,1\}$. For example, for the non-VA hospital characteristic of number of staffed beds, we create a binary indicator variable for whether the volume-weighted average number of staffed beds across non-VA hospitals in a zip code is above or below median. We include a demeaned $\tilde{I}_{x,i} \equiv I_{x,i} - \hat{E}_i\left[I_{x,i}\right]$ in the following linear control function regression:

$$Y_i = \beta_x D_i + \rho_x D_i \tilde{I}_{x,i} + \pi_x \tilde{I}_{x,i} + \gamma_x \hat{\varepsilon}_{1,i} + \mathbf{X}_i^0 \delta_x + \zeta_{x,z(i)} + \epsilon_{x,i}, \tag{8}$$

where $\hat{\varepsilon}_{1,i}$ is the first-stage error from Equation (3). Controlling for the endogeneity of selection, this approach yields estimates of binary heterogeneous treatment effects along several dimensions.[24] Since $\tilde{I}_{x,i}$ has a mean of 0, we can interpret β_x as the LATE, controlling for $\tilde{I}_{x,i}$; ρ_x is the difference in the VA effect on mortality between $I_{x,i} = 1$ and $I_{x,i} = 0$. We calculate standard errors by bootstrap, drawing blocks of data by zip code.

[24] For a discussion of this general approach, see Wooldridge (2015), Section III.

We find treatment heterogeneity that is both statistically significant and intuitive. In Appendix Tables A.9 and A.10, we observe that the VA advantage is marginally smaller in zip codes featuring non-VA hospitals with advanced capabilities (e.g., advanced trauma level, teaching hospital status, or STEMI center status), a finding consistent with a causal VA survival advantage, rather than an alternative that reflects selection of healthier patients to the VA.[25] As for heterogeneity by VA hospital characteristics (Appendix Table A.11), results suggest that the VA advantage pertains across a broad spectrum of VA hospitals but is perhaps slightly greater in larger hospitals. Across patients (Appendix Table A.12), the VA advantage is likely as large for minority veterans (Black and Hispanic) as for non-minority veterans. The VA survival benefit appears marginally greater for veterans who suffer from mental illness or substance abuse, who have more prior visits at the VA, or who have higher predicted mortality. However, the differences in the VA survival benefit across all of these dimensions are small compared to the overall scale of the survival benefit. Importantly, the VA survival benefit is not limited to select medical conditions that stereotypical users of the VA might have; even patients who are less than likely to use the VA experience a VA survival benefit.

5.4 Effect on Spending and Utilization

In light of the important literature on the returns to spending in health care (e.g., Garber and Skinner 2008), we examine the causal effect of VA vs. non-VA care on spending. The motivation behind this analysis is similar to that in Doyle et al. (2015), who sought to understand whether higher-spending hospitals achieve better health outcomes. To perform this analysis, we rely on both internal VA cost data and Medicare payment data from claims. Internal VA cost accounting apportions costs by VA utilization data and scales the cost of each encounter so that total spending matches actual budgeted spending within each VHA station.[26] On the Medicare side, we include payments made both by the veteran (i.e., coinsurance and deductible) and by the government. Therefore, we measure the cost of both VA and non-VA health care in terms of dollars spent by the government and the veteran.

Using the same instrumental variables approach as in our benchmark analysis, we study the effect of VA vs. non-VA care on daily VA and non-VA spending over time since the ambulance ride.

[25] Specifically, we expect hospitals with advanced capabilities to attract patients at greater risk of death. Therefore, under a hypothesis that the VA survival advantage reflects selection of healthier patients, we would observe higher mortality at non-VA hospitals (i.e., a greater VA survival advantage) when the non-VA hospitals for a zip code have advanced capabilities, and we would observe a smaller VA survival advantage when the VA hospital for a zip code has advanced capabilities. Instead, if anything, we observe the opposite patterns.

[26] The apportioning is based on a regression prediction of non-VA spending on relative value units (RVUs) associated with CPT codes, diagnosis-related group (DRG) weights, patient characteristics, and admission lengths of stay. This methodology is detailed in Wagner et al. (2003).

Specifically, we combine VA and Medicare spending in various intervals of days since the ambulance ride, and we divide this combined spending by the number of days in the interval to arrive at a daily spending flow during the interval, which we use as an outcome variable in Equation (4). Figure 5 shows the effect of the VA on the combined daily spending flow. It shows that, if anything, the VA *reduces* spending by more than $120 per day in the first ten days following the ambulance ride. Table 3 further shows that the VA reduces 28-day combined spending by $2,548, or by 21% of the mean 28-day spending. The reduction in spending reflects a lower probability of inpatient admission and fewer hospital days associated with VA care, although VA care results in slightly more outpatient visits in the following 28 days.[27]

The result that the VA saves lives while reducing spending is significant for two reasons. First, the result speaks directly to the policy question of whether the VA should privatize its care in a Medicare-type arrangement. The potential role of the VA as an insurer of private care has featured heavily in recent policy proposals and laws (113th Congress 2014; 115th Congress 2018). We show that, at least for the patients in our design, this privatization arrangement would be dominated by the status quo, as it would lead to both higher spending and worse health outcomes. Second, this joint finding suggests that the general mechanism behind the VA survival benefit is not higher spending but higher productivity. Our evidence points to productive inefficiency, rather than "flat of the curve" spending that underlies the relatively low-returns to US health care. This implication complements a growing literature on productivity differences across personnel (Chan et al. 2019; Silver 2020) and hospitals (Chandra and Staiger 2007, 2020) by showing an important productivity difference between health care *systems*.

5.5 Health IT and Integrated Care

Our final analysis investigates the role of health IT and integrated care in explaining the VA survival advantage. A substantial literature has reported the qualitative importance of these mechanisms in the VA's "transformation" into a high-quality health care organization in the mid-1990s (e.g., Jha et al. 2003). The lack of information flow and the high degree of fragmentation across providers in the US private health care sector have long been highlighted as potential roots of inefficiency (Cebul et al. 2008; Jha et al. 2009; Cutler 2010). These information-based mechanisms would be consistent with

[27] Although the average Medicare outpatient visit costs less than the average VA outpatient visit, the average Medicare inpatient day costs more than the average VA inpatient day. In our main analytic sample described in Appendix Table A.1, the average VA outpatient visit costs $181.39, and the average VA inpatient day costs $1,580.27. In the same sample, the average Medicare outpatient visit costs $108.32 (lower than the average VA cost), but the average Medicare inpatient day costs $1,816.54 (higher than the average VA cost).

greater health care productivity of the VA, particularly for regular users and patients whose conditions are more responsive to informed management (e.g., substance abuse).

Ideally, we would study these mechanisms by observing the VA's implementation of health IT and its reorganization into more integrated care in the mid-1990s. However, these reforms predate our data.[28] Similarly, it is not possible to examine the VA's effect on mortality among veterans who have no prior utilization at the VA, since it is exceedingly rare for these veterans to be transported to the VA by ambulance, as shown in Appendix Figure A.8. Nevertheless, this figure also shows that veterans may utilize more than one non-VA hospital system and that veterans may or may not be transported to their *modal* non-VA hospital, defined as the hospital system at which they had the largest number of utilization days in the prior year.

Our analytic strategy thus centers on a sample of veterans with no prior VA utilization but with some prior non-VA utilization.[29] While these veterans will almost certainly be transported to a non-VA hospital, we assess mortality outcomes as a function of whether they are quasi-randomly assigned—via a similar ambulance instrument as the one we use in our benchmark analysis—to their modal non-VA hospital. This modal-hospital effect on mortality arguably captures at least some of the potential effect of continuity of care in the private sector. In order to more explicitly investigate the role of health IT and integrated care, we further exploit two changes induced by incentives in federal laws and payment policies during our study period. First, the HITECH Act of 2009 dramatically increased the share of hospitals using health IT (Blumenthal 2010).[30] Second, in 2011, Medicare began to incentivize care integration via alternative payment arrangements to "Accountable Care Organizations" (ACOs) (Greaney 2011).

As an analog to our benchmark VA instrument in Equation (1), we construct an instrument that reflects a given ambulance company's leave-out propensity to deliver patients to the index patient's modal non-VA hospital. Let $h(i)$ denote the hospital that ambulance ride i is transported to, and let

[28] Indeed, the VHA's adoption of a common health IT platform (VistA) in the mid-1990s paved the way for research on health services within the VHA system, including this study.

[29] We detail the sample selection process for this analysis in Appendix Table A.13 and present patient and ride characteristics in this sample Appendix Table A.14. Since we require that veterans in this sample have no prior VA utilization, while the sample in our benchmark analysis only includes veterans with prior VA utilization, this sample is completely disjoint from the sample in our benchmark analysis. We only include zip codes with at least two non-VA hospitals within 20 miles, but we make not requirement on proximity to a VA hospital. Compared to our benchmark sample, this sample features a negligible probability of transport to a VA hospital (0% as opposed to 33%) yet remarkably similar rates of weekend transport of 28-day mortality.

[30] Jha et al. (2009) document in 2009 that 1.5% of US non-federal hospitals have an electronic health record (EHR) system present in all clinical units and an additional 7.6% have an EHR system present in at least one clinical unit. According to the Office of the National Coordinator on Health Information Technology (ONC), by 2014, 97% of such hospitals had possessed an EHR technology meeting requirements of the Department of Health and Human Services, and 76% of hospitals had implemented the EHR system in at least one clinical unit (Charles et al. 2015).

$h^m(i)$ represent the modal non-VA hospital used by patient $k(i)$ in ride i. Our treatment of interest is $D_i^m \equiv \mathbf{1}(h(i) = h^m(i))$, which indicates whether ambulance ride i transports its patient $k(i)$ to his modal hospital. Our instrumental variable for this treatment is:

$$Z_i^m = \frac{1}{K_{j(i),z(i)} - 1} \sum_{i' \in \mathcal{I}_{j(i),z(i)}} \frac{\mathbf{1}(k(i') \neq k(i)) \, 1(h(i') = h^m(i))}{N_{k(i'),z(i'),j(i')}}. \tag{9}$$

where $K_{j,z}$ is the number of patients transported by company j from zip code z, $N_{k,z,j}$ is the number of rides taken by patient k originating in zip code z with company j, and $\mathcal{I}_{j,z}$ is the set of rides transported by ambulance company j from zip code z. This is the leave-out probability that ambulance company $j(i)$ transports other patients from the same zip code to the modal hospital $h^m(i)$ of patient $k(i)$.[31] We use the following first-stage and reduced-form equations, similar to Equations (4) and (3):

$$D_i^m = \pi_1^m Z_i^m + \gamma_1^m \overline{Z}_i^m + \mathbf{X}_i^0 \delta_1^m + \zeta_{1,z(i)}^m + \varepsilon_{1,i}^m; \tag{10}$$

$$Y_i = \pi_2^m Z_i^m + \gamma_2^m \overline{Z}_i^m + \mathbf{X}_i^0 \delta_2^m + \zeta_{2,z(i)}^m + \varepsilon_{2,i}^m, \tag{11}$$

where we include an additional control variable:

$$\overline{Z}_i^m = \frac{1}{K_{z(i)} - 1} \sum_{i' \in \mathcal{I}_{z(i)}} \frac{\mathbf{1}(k(i') \neq k(i)) \, 1(h(i') = h^m(i))}{N_{k(i'),z(i')}},$$

where K_z is the number of patients from zip code z, $N_{k,z}$ is the number of rides taken by patient k originating in zip code z, and $\mathcal{I}_z$ is the set of rides originating in zip code z This is the leave-out probability that patients from the same zip code $z(i)$ are transported to hospital $h^m(i)$, *unconditional* on the ambulance company. The modal-hospital effect may also capture hospital quality or hospital-patient match effects. We further assess the modal hospital effect both (i) by including hospital fixed effects in Equations (10) and (11) and (ii) by splitting rides i into samples based on whether the ride was before or after the hospital $h(i)$ adopted health IT or joined an ACO.

In the sample of veterans with only non-VA prior utilization (Panel B of Appendix Table A.13), we demonstrate in Appendix Figure A.9 a well-behaved first-stage relationship between D_i^m and Z_i^m and balance between predicted mortality, $\hat{Y}_i$, and Z_i^m, conditional on $\left(\overline{Z}_i^m, \mathbf{X}_i^0, z(i)\right)$.[32] The IV estimate

[31] As with the benchmark instrument, we construct this instrument from data in the overall sample of ambulance rides with dually eligible veterans (Column 1, Table 1 and A.14). For patients with multiple hospitals that tie for highest utilization in the prior year, we designate the set of these highest-use hospitals as the "modal hospital."

[32] Analogously to Figure 1, this figure presents binned scatterplots of the first-stage regression in Equation (10), the reduced-form regression in Equation (11), and a balance regression with predicted mortality as the outcome variable and

of the modal-hospital effect on mortality is −0.006 (s.e. 0.004), which is less than 20% of the VA effect on mortality. The visual IV graph in Appendix Figure A.10 shows that the overall relationship between the reduced form and first stage is not particularly striking.[33] However, computing the same IV estimate separately by years, we show in Figure 6 a stronger modal-hospital effect emerges after the passage of the HITECH Act of 2009, which led to a rapid rise in electronic medical record systems. The modal-hospital effect is close to 0 and stable prior to 2009; following 2009, the modal-hospital effect grows to about half the size of the VA effect on mortality.

In Table 6, we examine how the modal-hospital effect relates to dates of hospital health IT or ACO adoption.[34] We estimate the modal-hospital effect in four subsamples defined by whether or not each veteran's modal hospital had adopted health IT, at the time of his ride, and similarly by whether or not each veteran's modal hospital had joined an ACO. We also use a control function approach to estimate separate modal-hospital effects, depending on health IT or ACO adoption, in the overall sample and with hospital fixed effects (see Appendix A.6 for details). The results in the table provide suggestive evidence that the growth in the modal-hospital effect is associated with health IT adoption, holding hospitals fixed; the relationship with ACO adoption appears similar but is imprecise. While most estimates in the table control for hospital fixed effects, we find that results are essentially unchanged regardless of their inclusion.

6 Conclusion

The structure of health-care delivery to US veterans provides a distinctive research opportunity, allowing us to study fundamentally different systems of health care that coexist for a large patient population. Specifically, millions of older veterans (those at least aged 65) are dually eligible for care in a public system, operated by the Veterans Health Administration, or in private-sector hospitals, financed by Medicare. The ambulance setting provides plausible quasi-experimental assignment of

the same design matrix.

[33] Analogously to Figure 2 and Appendix Figure A.2 in the benchmark analysis, Appendix Figure A.11 shows stability in OLS and two-stage least squares estimates with increasing controls, and Appendix Figure A.12 shows robustness of two-stage least squares estimates under an exhaustive set of control combinations.

[34] We measure health IT adoption from a dataset from the Office of the National Coordinator of Health Information Technology (ONC). This dataset merges hospital attestation data from the Medicare EHR Incentive Program with certified EHR product information from ONC's Certified Health IT Product List (CHPL), and we code the use of any certified product as health IT adoption. We find that greater than 95% of non-VA hospitals in the last year of our sample. We measure ACO participation from the Medicare Shared Savings Program (MSSP) Accountable Care Organizations (ACO) dataset. Consistent with other research, we find that only 11% of non-VA hospitals participated in ACOs by the last year of our sample (Colla et al. 2016). The lack of effect of ACO participation that we find is also consistent with a recent literature showing limited changes in utilization patterns and largely null effects on outcomes under ACOs (McWilliams et al. 2014a,b, 2016; Trombley et al. 2019).

veterans to these health care systems. Our work has current policy relevance, as the Department of Veterans Affairs is now considering whether to bolster its existing public delivery system or to replace it, either partially or fully, with a system of financing private care. Our work has implications more broadly for understanding the impact of public vs. private health care on mortality and spending.

We find a significant VA advantage: our preferred instrumental variables estimate, based on veterans who are induced by their ambulance company to use the Veterans Health Administration (VHA), shows a 4.5 p.p. survival gain at 28 days (confidence interval 1.1 to 8.0 p.p.), implying about a 46% reduction in mortality relative to the overall average. In a novel survival analysis of this quasi-experiment, we show that these survival gains occur in the first week following the ambulance ride and appear to be long-lasting. We further use this survival analysis framework to validate our IV quasi-experiment and to demonstrate differences in long-term mortality hazards between VA and non-VA users who are non-compliers. Our analysis of long-term hazards provides a compelling explanation for the difference in magnitude between IV and OLS estimates of the VA effect on 28-day mortality. Although we find some intuitive margins of heterogeneity in the VA advantage, the VA outperforms the non-VA alternative in a wide variety of locations with different types of non-VA hospitals and for all types of patients we consider, not only for patients with stereotypical medical conditions.

Importantly, we also find that the VA reduces total spending, including government costs and patient out-of-pocket expenses, by 21% relative to non-VA providers, which points to higher productivity in the VHA than in the private sector. Using our quasi-experiment, we shed light on mechanisms, many of which have been raised more descriptively or qualitatively (e.g., Jha et al. 2003). We interpret our findings as consistent with the idea that the VA advantage arises from continuity of care, health IT, and organization. For example, we find that compliers are more likely to have prior VA care and have larger survival gains from VA assignment than average; in a selection model that rationalizes this finding, we show veterans who are more likely to use the VA also have larger survival gains. Interestingly, we show that a similar effect occurs in the private sector for veterans who primarily use a private hospital system. These veterans also experience reduced mortality when quasi-experimentally assigned to their modal private hospital, but only in a period following adoption of health IT due to government incentives, and even then on a smaller scale (approximately half the size) than the overall VA advantage.

Our results contribute more broadly to two streams of literature on the efficiency of production. First, we contribute to the descriptive analysis that compares the performance of the US health care system to systems in other developed countries (Blank et al. 2017). By almost all accounts, compar-

isons of US health outcomes and health care spending are unfavorable with those of other developed countries (Garber and Skinner 2008; Rice et al. 2013). Our analysis points to a potentially large source of inefficiency in the US context: its private provision of health care. Of note, several developed countries that outperform the US also feature private provision, although the US system arguably has the most complex configuration of financing and delivery, with the highest levels of uninsurance and administrative costs. It is also important to note that, since we study care financed generally by the government, our analysis is silent on the contribution of health care financing to overall system performance, though analogous and interacting inefficiencies may arise from fragmented and private health care financing in the US (Cebul et al. 2008).

Second, we provide empirical support in the context of health care for the general idea of production complementarities among three innovations in production: information technology (IT), workplace reorganization, and products and services (Bresnahan et al. 2002). The VHA adopted a comprehensive health IT system almost two decades before the vast majority of private hospitals in the US. This reform was accompanied by an integration of care involving both reorganizing the delivery system and redefining services involved in patient care. For private hospitals, redefining health care products and services is limited by fee-for-service payment systems and the difficulty of measuring quality (Cutler 2010). Hospitals without a broad network of clinics and an overarching mandate for a population's health may find it difficult to reorganize and redefine its services. Our result that health IT in private hospitals may improve survival—but only for patients that the hospitals have previously treated—is consistent with production complementarities. Complementarities in health care production pose barriers for replicating the VA advantage in the fragmented private landscape of US health care.

References

106TH CONGRESS (1999): "Veterans Millennium Health Care and Benefits Act," Public Law 106-117.

113TH CONGRESS (2014): "Veterans Access, Choice, and Accountability Act of 2014," Public Law 113-146.

115TH CONGRESS (2018): "Maintaining Internal Systems and Strengthening Integrated Outside Networks (MISSION) Act of 2018," Public Law 115-182.

ABADIE, A. (2002): "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, 97, 284–292.

ADAMSON, D. M., M. A. BURNAM, R. M. BURNS, L. B. CALDARONE, R. A. COX, E. J. D'AMICO, C. DIAZ, C. EIBNER, G. FISHER, T. HELMUS, B. KARNEY, B. KILMER, G. N. MARSHALL, L. T. MARTIN, L. S. MEREDITH, K. N. METSCHER, K. C. OSILLA, R. L. PACULA, R. RAMCHAND, J. RINGEL, T. SCHELL, J. SOLLINGER, M. E. VAIANA, K. M. WILLIAMS, AND M. R. YOCHELSON (2008): *Invisible Wounds of War: Psychological and Cognitive Injuries, Their Consequences, and Services to Assist Recovery*, Santa Monica, CA: RAND Corporation.

AGHA, L. (2014): "The effects of health information technology on costs and quality of medical care," *Journal of Health Economics*, 19–30.

AGHA, L., B. FRANDSEN, AND J. B. REBITZER (2019): "Fragmented division of labor and healthcare costs: Evidence from moves across regions," *Journal of Public Economics*, 169, 144–159.

AGHA, Z., R. P. LOFGREN, J. V. VANRUISWYK, AND P. M. LAYDE (2000): "Are patients at Veterans Affairs medical centers sicker? A comparative analysis of health status and medical resource use," *Archives of Internal Medicine*, 160, 3252–3257.

ALCHIAN, A. A. (1965): "Some Economics of Property Rights," *Il Politico*, 30, 816–829.

ALTONJI, J. G., T. E. ELDER, AND C. R. TABER (2005): "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 113, 151–184.

ALTONJI, J. G. AND R. K. MANSFIELD (2018): "Estimating Group Effects Using Averages of Observables to Control for Sorting on Unobservables: School and Neighborhood Effects," *American Economic Review*, 108, 2902–2946.

ALTONJI, J. G. AND L. M. SEGAL (1996): "Small-Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business & Economic Statistics*, 14, 353–366.

ARNOLD, D., W. DOBBIE, AND C. S. YANG (2018): "Racial Bias in Bail Decisions," *Quarterly Journal of Economics*, 133, 1885–1932.

BAKER, L. C. (2001): "Managed care and technology adoption in health care: evidence from magnetic resonance imaging," *Journal of Health Economics*, 20, 395–421.

BALKE, A. AND J. PEARL (1997): "Bounds on Treatment Effects from Studies with Imperfect Compliance," *Journal of the American Statistical Association*, 92, 1171–1176.

BHULLER, M., G. B. DAHL, K. V. LOKEN, AND M. MOGSTAD (2020): "Incarceration, Recidivism, and Employment," *Journal of Political Economy*, 128, 1269–1324.

BLANK, R., V. BURAU, AND E. KUHLMANN (2017): *Comparative Health Policy*, Macmillan International Higher Education.

BLOM, D. (2016): "Considerations and Recommendations," Tech. rep., Commission on Care, Washington, D.C.

BLUMENTHAL, D. (2010): "Launching HITECH," *New England Journal of Medicine*, 362, 382–385.

BLUNDELL, R. AND R. L. MATZKIN (2014): "Control functions in nonseparable simultaneous equations models," *Quantitative Economics*, 5, 271–295.

BRESNAHAN, T. F., E. BRYNJOLFSSON, AND L. M. HITT (2002): "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence," *The Quarterly Journal of Economics*, 117, 339–376.

CARD, D., C. DOBKIN, AND N. MAESTAS (2009): "Does Medicare Save Lives?" *Quarterly Journal of Economics*, 124, 597–636.

CEBUL, R. D., J. B. REBITZER, L. J. TAYLOR, AND M. E. VOTRUBA (2008): "Organizational fragmentation and care quality in the U.S. healthcare system," *Journal of Economic Perspectives*, 22, 93–113.

CENTERS FOR MEDICARE & MEDICAID SERVICES (2002): "Medicare Program; Fee Schedule for Payment of Ambulance Services and Revisions to the Physician Certification Requirements for Coverage of Nonemergency Ambulance Services," in *Federal Register*, 67, 9099–9135.

CHAN, D. (2016): "Teamwork and Moral Hazard: Evidence from the Emergency Department," *Journal of Political Economy*, 124, 734–770.

CHAN, D. C., M. GENTZKOW, AND C. YU (2019): "Selection with Variation in Diagnostic Skill: Evidence from Radiologists," Working Paper 26467, National Bureau of Economic Research.

CHANDRA, A., A. FINKELSTEIN, A. SACARNY, AND C. SYVERSON (2016): "Healthcare Exceptionalism? Productivity and Allocation in the U.S. Healthcare Sector," *American Economic Review*, 106, 2110–2144.

CHANDRA, A. AND D. O. STAIGER (2007): "Productivity spillovers in healthcare: evidence from the treatment of heart attacks," *The Journal of Political Economy*, 115, 103–140.

——— (2020): "Identifying Sources of Inefficiency in Healthcare," *The Quarterly Journal of Economics*, 135, 785–843.

CHARLES, D., M. GABRIEL, AND T. SEARCY (2015): "Adoption of Electronic Health Record Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2014," ONC Data Brief 23, The Office of the National Coordinator for Health Information Technology.

CHIANG, A. J., G. DAVID, AND M. G. HOUSMAN (2006): "The Determinants of Urban Emergency Medical Services Privatization," *Critical Planning*, 13.

COLLA, C. H., V. A. LEWIS, E. TIERNEY, AND D. B. MUHLESTEIN (2016): "Hospitals Participating In ACOs Tend To Be Large And Urban, Allowing Access To Capital And Data," *Health Affairs*, 35, 431–439.

CUTLER, D. (2010): "Where Are the Health Care Entrepreneurs?" *Issues in Science and Technology*, 27, 49–56.

CUTLER, D. M., M. MCCLELLAN, AND J. P. NEWHOUSE (2000): "How Does Managed Care Do It?" *The RAND Journal of Economics*, 31, 526–548.

DAHL, G. B., A. R. KOSTOL, AND M. MOGSTAD (2014): "Family Welfare Cultures," *The Quarterly Journal of Economics*, 129, 1711–1752.

DOYLE, J. J., J. A. GRAVES, J. GRUBER, AND S. KLEINER (2015): "Measuring Returns to Hospital Care: Evidence from Ambulance Referral Patterns," *Journal of Political Economy*, 123, 170–214.

ELIXHAUSER, A., C. STEINER, D. R. HARRIS, AND R. M. COFFEY (1998): "Comorbidity Measures for Use with Administrative Data," *Medical Care*, 36, 8–27.

GARBER, A. M. AND J. SKINNER (2008): "Is American Health Care Uniquely Inefficient?" *The Journal of Economic Perspectives*, 22, 27–50.

GAYNOR, M., J. REBITZER, AND L. TAYLOR (2004): "Physician incentives in Health Maintenance Organizations," *The Journal of Political Economy*, 112, 915–931.

GORDON, S. (2019): "The VA Is Privatizing Veterans' Health Care While Launching a Campaign to Deny It," *The American Prospect*.

GREANEY, T. L. (2011): "Accountable Care Organizations: The Fork in the Road," *New England Journal of Medicine*, 364, e1.

HECKMAN, J. J. AND E. VYTLACIL (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73, 669–738.

HECKMAN, J. J. AND E. J. VYTLACIL (2007): "Chapter 71 Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments," in *Handbook of Econometrics*, ed. by J. J. Heckman and E. E. Leamer, Elsevier, vol. 6, 4875–5143.

HONORE, B. E. AND A. LLERAS-MUNEY (2006): "Bounds in Competing Risks Models and the War on Cancer," *Econometrica*, 74, 1675–1698.

HUSKAMP, H. A. AND J. K. IGLEHART (2016): "Mental Health and Substance-Use Reforms: Milestones Reached, Challenges Ahead," *New England Journal of Medicine*, 375, 688–695.

HUSSEY, P. S., E. C. SCHNEIDER, R. S. RUDIN, D. S. FOX, J. LAI, AND C. E. POLLACK (2014): "Continuity and the Costs of Care for Chronic Disease," *JAMA Internal Medicine*, 174, 742–748.

HYNES, D. M., K. KOELLING, K. STROUPE, N. ARNOLD, K. MALLIN, M.-W. SOHN, F. M. WEAVER, L. MANHEIM, AND L. KOK (2007): "Veterans' Access to and Use of Medicare and Veterans Affairs Health Care," *Medical Care*, 45, 214–223.

IMBENS, G. W. AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475.

IMBENS, G. W. AND D. B. RUBIN (1997): "Estimating Outcome Distributions for Compliers in Instrumental Variables Models," *Review of Economic Studies*, 64, 555–574.

JHA, A. K., C. M. DESROCHES, E. G. CAMPBELL, K. DONELAN, S. R. RAO, T. G. FERRIS, A. SHIELDS, S. ROSENBAUM, AND D. BLUMENTHAL (2009): "Use of Electronic Health Records in U.S. Hospitals," *New England Journal of Medicine*, 360, 1628–1638.

JHA, A. K., J. B. PERLIN, K. W. KIZER, AND R. A. DUDLEY (2003): "Effect of the Transformation of the Veterans Affairs Health Care System on the Quality of Care," *New England Journal of Medicine*, 348, 2218–2227.

KEFE, J. (2018): "Is the VA being privatized? Critics say this Koch-backed group is fighting to do just that," *Newsweek*.

KITAGAWA, T. (2015): "A Test for Instrument Validity," *Econometrica*, 83, 2043–2063.

KLINE, P. M. AND C. R. WALTERS (2019): "On Heckits, LATE, and Numerical Equivalence," *Econometrica*, 87, 677–696.

KLING, J. R. (2006): "Incarceration Length, Employment, and Earnings," *The American Economic Review*, 96, 863–876.

KOLESAR, M., R. CHETTY, J. FRIEDMAN, E. GLAESER, AND G. W. IMBENS (2015): "Identification and Inference With Many Invalid Instruments," *Journal of Business and Economic Statistics*, 33, 474–484.

LUFT, H. S. (2010): "Becoming Accountable: Opportunities and Obstacles for ACOs," *New England Journal of Medicine*, 363, 1389–1391.

MCCARTHY, D. AND D. BLUMENTHAL (2006): "Stories from the Sharp End: Case Studies in Safety Improvement," *The Milbank Quarterly*, 84, 165–200.

MCWILLIAMS, J. M., M. E. CHERNEW, J. B. DALTON, AND B. E. LANDON (2014a): "Outpatient Care Patterns and Organizational Accountability in Medicare," *JAMA Internal Medicine*, 174, 938.

MCWILLIAMS, J. M., L. A. HATFIELD, M. E. CHERNEW, B. E. LANDON, AND A. L. SCHWARTZ (2016): "Early Performance of Accountable Care Organizations in Medicare," *New England Journal of Medicine*, 374, 2357–2366.

MCWILLIAMS, J. M., B. E. LANDON, M. E. CHERNEW, AND A. M. ZASLAVSKY (2014b): "Changes in Patients' Experiences in Medicare Accountable Care Organizations," *New England Journal of Medicine*, 371, 1715–1724.

MILLER, A. R. AND C. E. TUCKER (2011): "Can Health Care Information Technology Save Babies?" *Journal of Political Economy*, 119, 289–324.

NUTI, S. V., L. QIN, J. S. RUMSFELD, J. S. ROSS, F. A. MASOUDI, S.-L. T. NORMAND, K. MURUGIAH, S. M. BERNHEIM, L. G. SUTER, AND H. M. KRUMHOLZ (2016): "Association of Admission to Veterans Affairs Hospitals vs Non-Veterans Affairs Hospitals With Mortality and Readmission Rates Among Older Men Hospitalized With Acute Myocardial Infarction, Heart Failure, or Pneumonia," *JAMA*, 315, 582.

O'HANLON, C., C. HUANG, E. SLOSS, R. A. PRICE, P. HUSSEY, C. FARMER, AND C. GIDENGIL (2017): "Comparing VA and Non-VA Quality of Care: A Systematic Review," *Journal of General Internal Medicine*, 32, 105–121.

OLSEN, R. J. (1980): "A Least Squares Correction for Selectivity Bias," *Econometrica*, 48, 1815–1820.

PERLMAN, M. D. (1969): "One-Sided Testing Problems in Multivariate Analysis," *The Annals of Mathematical Statistics*, 40, 549–567.

RAGONE, M. (2012): "Evolution or revolution: EMS industry faces difficult changes," *Journal of Emergency Medical Services*, 37, 34–39.

REID, T. R. (2010): *The Healing of America: A Global Quest for Better, Cheaper, and Fairer Health Care*, Penguin.

REIN, L., P. RUCKER, E. WAX-THIBODEAUX, AND J. DAWSEY (2018): "Trump taps his doctor to replace Shulkin at VA, choosing personal chemistry over traditional qualifications," *Washington Post*.

RICE, T., P. ROSENAU, L. Y. UNRUH, AND A. J. BARNES (2013): "United States of America: Health system review," *Health Systems in Transition*, 15, 1–431.

SCHWARTZ, J. (2000): "Harvesting and Long Term Exposure Effects in the Relation between Air Pollution and Mortality," *American Journal of Epidemiology*, 151, 440–448.

SHEKELLE, P. G., S. ASCH, P. GLASSMAN, S. MATULA, A. TRIVEDI, AND I. MIAKE-LYE (2010): "Comparison of Quality of Care in VA and Non-VA Settings: A Systematic Review," Tech. rep., Department of Veterans Affairs (US), Washington (DC).

SHULKIN, D. J. (2018): "Privatizing the V.A. Will Hurt Veterans," *The New York Times*.

SILVER, D. (2020): "Haste or Waste? Peer Pressure and Productivity in the Emergency Department," Tech. rep., Princeton University.

SKURA, B. (2001): "Where Do 911 System Ambulances Take Their Patients? Differences between Voluntary Hospital Ambulances and Fire Department Ambulances." Tech. rep., Office Comptroller, City of New York, New York, NY.

STIGLER, G. J. (1965): "The Economist and the State," *The American Economic Review*, 55, 1–18.

TELSER, L. G. (1964): "Advertising and Competition," *Journal of Political Economy*, 72, 537–562.

TRIVEDI, A. N., S. MATULA, I. MIAKE-LYE, P. A. GLASSMAN, P. SHEKELLE, AND S. ASCH (2011): "Systematic Review: Comparison of the Quality of Medical Care in Veterans Affairs and Non-Veterans Affairs Settings," *Medical Care*, 49, 76–88.

TROMBLEY, M. J., B. FOUT, S. BRODSKY, J. M. MCWILLIAMS, D. J. NYWEIDE, AND B. MOREFIELD (2019): "Early Effects of an Accountable Care Organization Model for Underserved Areas," *New England Journal of Medicine*, 381, 543–551.

U.S. GOVERNMENT ACCOUNTABILITY OFFICE (2015): "VA Mental Health: Clearer Guidance on Access Policies and Wait-Time Data Needed," Tech. Rep. GAO-16-24.

WAGNER, T. H., S. CHEN, AND P. G. BARNETT (2003): "Using Average Cost Methods to Estimate Encounter-Level Costs for Medical-Surgical Stays in the VA," *Medical Care Research and Review*, 60, 15S–36S.

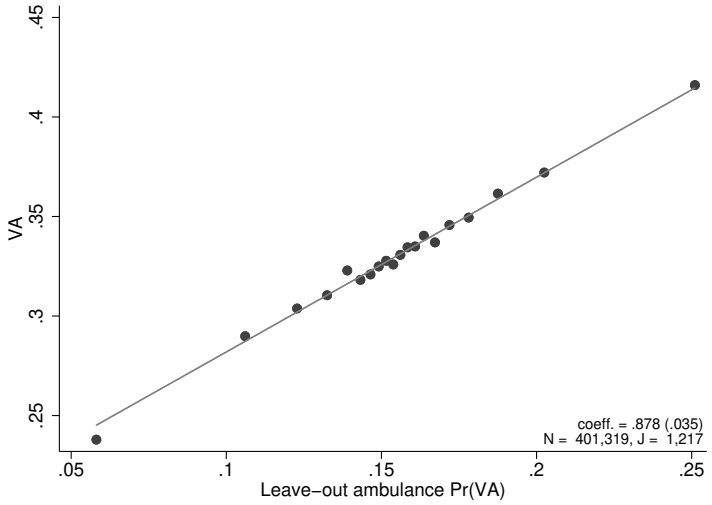
WOLAK, F. A. (1987): "An Exact Test for Multiple Inequality and Equality Constraints in the Linear Regression Model," *Journal of the American Statistical Association*, 82, 782–793.

WOOLDRIDGE, J. M. (2015): "Control Function Methods in Applied Econometrics," *Journal of Human Resources*, 50, 420–445.

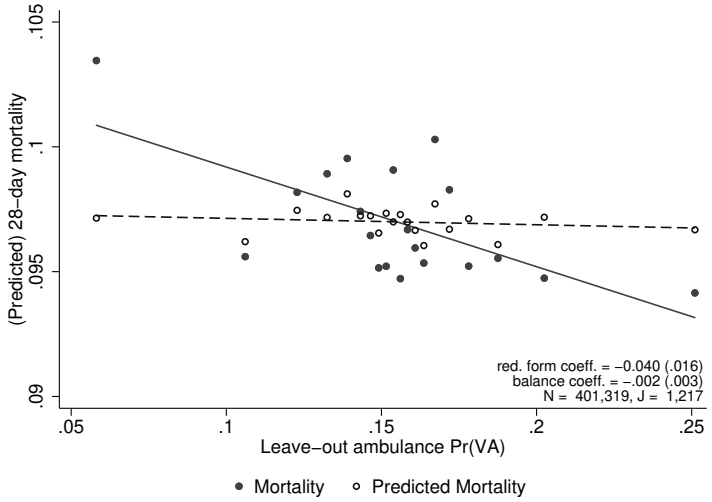
WRIGHT, S. M., L. A. PETERSEN, R. P. LAMKIN, AND J. DALEY (1999): "Increasing Use of Medicare Services by Veterans With Acute Myocardial Infarction," *Medical Care*, 37, 529–537.

Figure 1: First Stage, Balance, and Reduced Form

A: First Stage

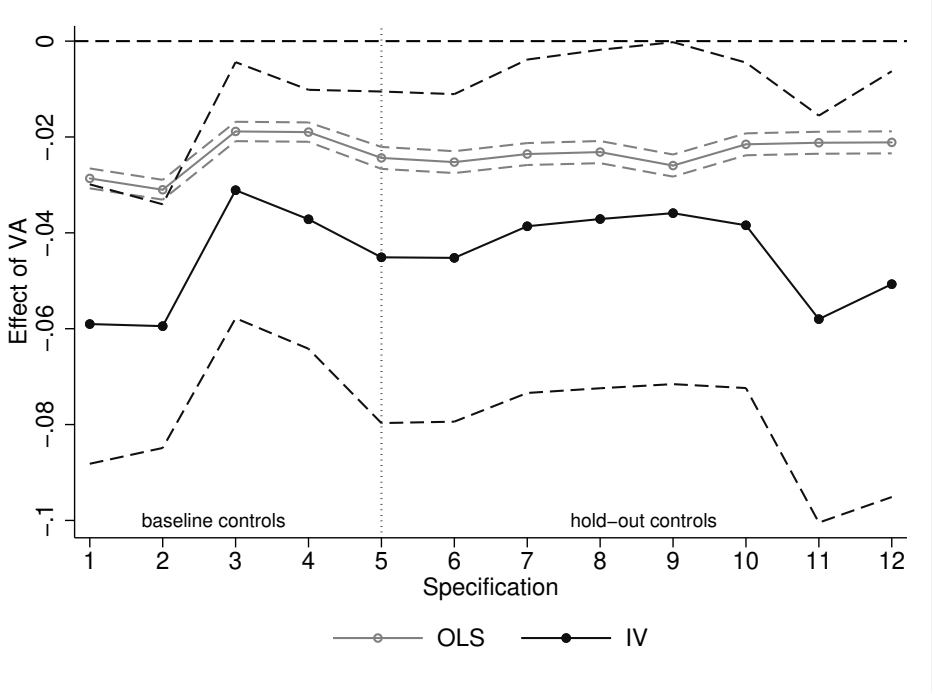


B: Balance and Reduced Form



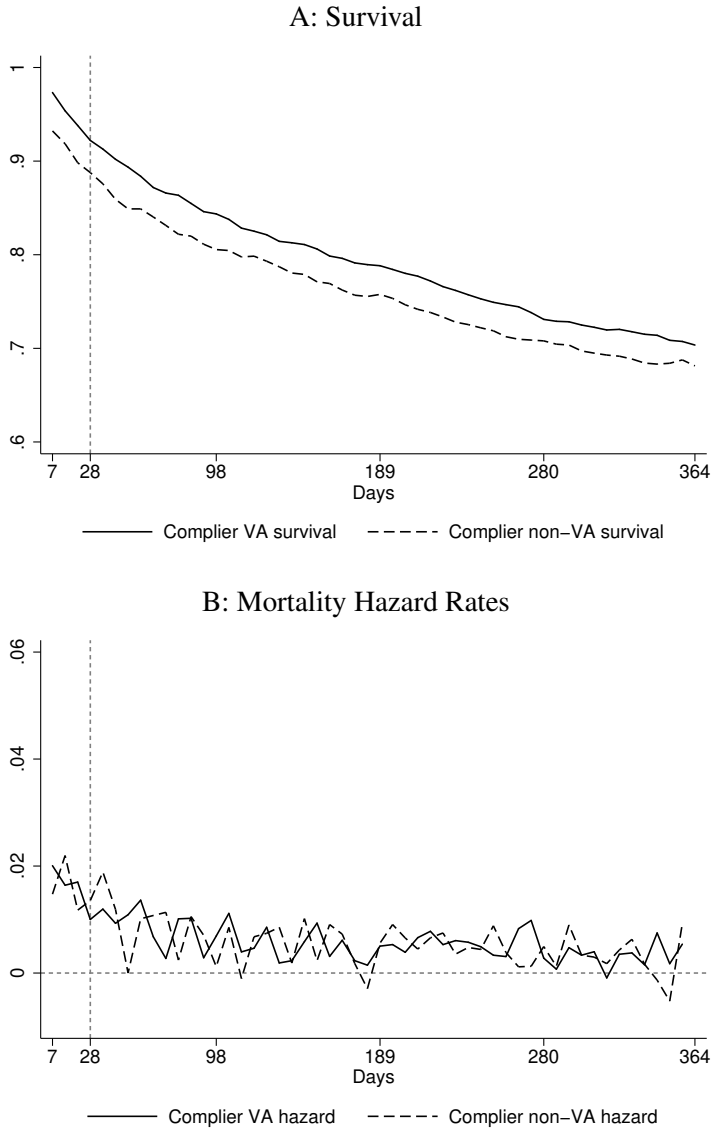
Note: Panel A shows a binned scatterplot of arrival at a VA hospital on the *y*-axis against the ambulance leave-out propensity to arrive at a VA hospital on the *x*-axis. The figure is a graphical representation of the first-stage regression in Equation (3). Panel B shows binned scatterplots of 28-day mortality and predicted 28-day mortality on the *y*-axis against the ambulance leave-out propensity to arrive at a VA hospital on the *x*-axis. Mortality bin means are shown in solid circles, while predicted mortality bin means are shown in hollow circles. The figure represents the reduced-form regression in Equation (4) and the corresponding balance regression replacing mortality with predicted mortality. The sample includes 400,769 ambulance rides and 1,267 combinations of ambulance company identifiers and Dartmouth Atlas Hospital Referral Regions (HRRs). The sample selection is given in Appendix Table A.1. Baseline controls are detailed in Appendix Table A.2 and include patient zip code dummies, ALS/BLS dummies, source of the ambulance ride, time categories, and patient prior utilization.

Figure 2: OLS and IV Specifications



Note: This figure shows OLS and IV estimates of the effect of the VA on 28-day mortality, represented in Equation (2) as β, with progressive sets of controls. Numbered incremental controls correspond to categories or subcategories of variables that are presented in order in Appendix Tables A.2 and A.3. Estimates are shown along solid lines, while 95% confidence intervals are shown in dashed lines. All specifications use the baseline sample, given in Appendix Table A.1.

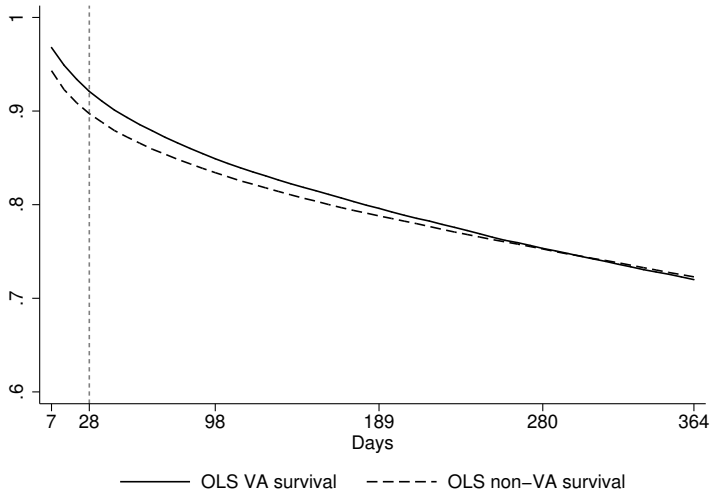
Figure 3: Complier Potential Outcomes



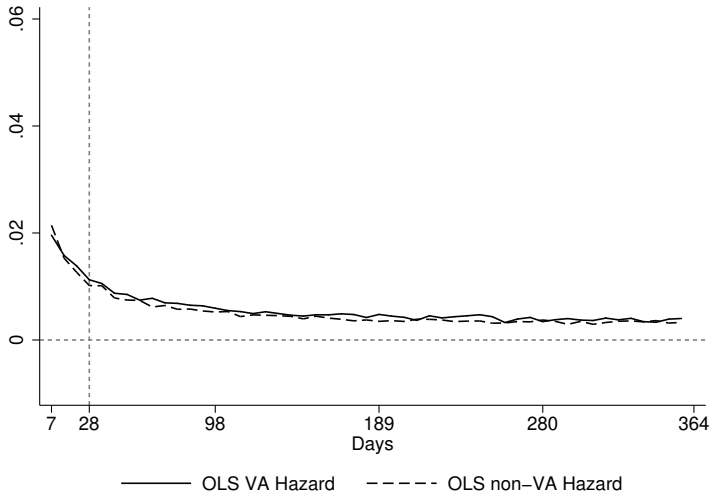
Note: This figure shows potential outcomes for ambulance compliers who arrive at a VA hospital and those who arrive at a non-VA hospital. Panel A shows survival outcomes as a function of days from the ambulance ride. "Day 0" indicates survival on the day of the ambulance ride; subsequent days indicate survival at one-week intervals from the ambulance ride. Denote $S_i(t;d) \in \{0,1\}$ as an indicator for whether patient i survives up to time t after the ambulance ride, depending on whether the patient arrives at the VA ($d = 1$) or a non-VA hospital ($d = 0$). Observed survival is $S_i(t) = D_i S_i(t;1) + (1 - D_i) S_i(t;0)$. We estimate complier VA survival, or $E[S_i(t;1) | i \in C]$, by a IV regression with a dependent variable of $S_i(t) D_i$, the endogenous VA treatment D_i, and the same first-stage and reduced-form design matrix implied by Equations (3) and (4). We estimate complier non-VA survival, or $E[S_i(t;0) | i \in C]$, by a similar IV regression with a dependent variable of $S_i(t)(D_i - 1)$. All regressions use a sample of ambulance rides with no prior ride in the last year and the same baseline controls as described in Figure 1. Panel B presents implied weekly mortality hazard rates, as given by Equation (5).

Figure 4: Observed Risk-Adjusted Outcomes

A: Survival

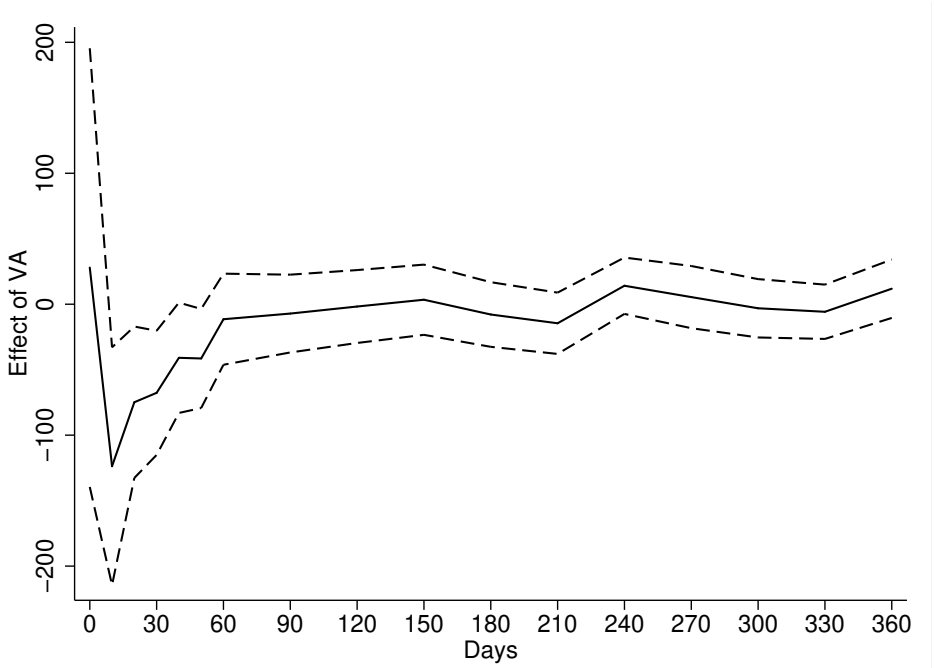


B: Mortality Hazard Rates



Note: This figure shows observed risk-adjusted outcomes for patients who arrive at a VA hospital and those who arrive at a non-VA hospital. Panel A shows survival outcomes as a function of days from the ambulance ride. "Day 0" indicates survival on the day of the ambulance ride; subsequent days indicate survival at one-week intervals from the ambulance ride. Denote $S_i(t;d) \in \{0,1\}$ as an indicator for whether patient i survives up to time t after the ambulance ride, depending on whether the patient arrives at the VA ($d = 1$) or a non-VA hospital ($d = 0$). Observed survival is $S_i(t) = D_i S_i(t;1) + (1 - D_i) S_i(t;0)$. We estimate VA survival, or $E[S_i(t) | D_i = 1]$, by an OLS regression with a dependent variable of $S_i(t) D_i$ and the same design matrix implied by Equation (2); we estimate non-VA survival, or $E[S_i(t) | D_i = 0]$, by a similar OLS regression with a dependent variable of $S_i(t)(D_i - 1)$. All regressions use a sample of ambulance rides with no prior ride in the last year and the same baseline controls as described in Figure 1. Panel B presents implied weekly mortality hazard rates, as given by Equation (6).

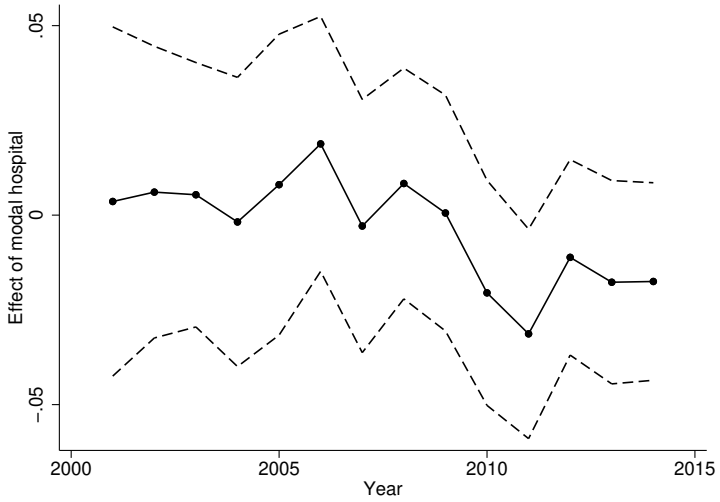
Figure 5: Spending



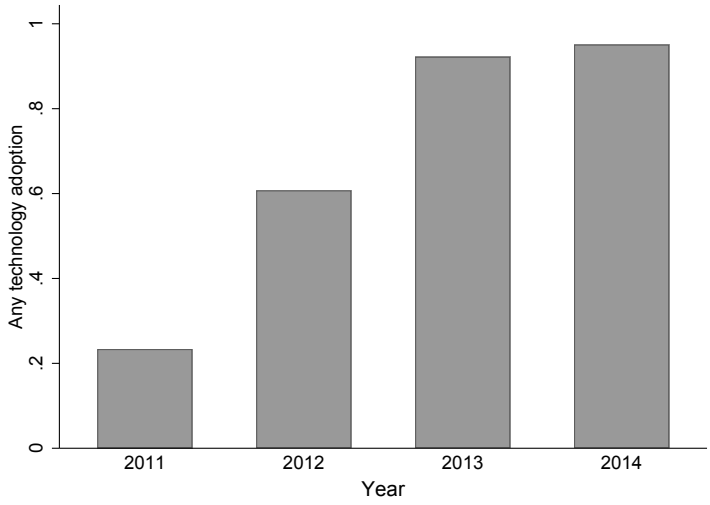
Note: This figure shows the effect of the VA on daily spending flow from both VA and non-VA sources by days after the ambulance ride. The first stage and reduced form of the IV regression are given in Equations (3) and (4), where the outcome variable in Equation (4) is daily spending flows during each of interval period. The sample includes 400,769 ambulance rides and 1,267 combinations of ambulance company identifiers and Dartmouth Atlas Hospital Referral Regions (HRRs). The sample selection is given in Appendix Table A.1. Baseline controls are detailed in Appendix Table A.2 and include patient zip code dummies, ALS/BLS dummies, source of the ambulance ride, time categories, and patient prior utilization. Estimates are shown in the solid line, while 95% confidence intervals are shown in dashed lines.

Figure 6: Modal Hospital Effect and Health IT Adoption

A: Modal Hospital Effect by Year



B: Health IT Adoption by Year



Note: Panel A of this figure shows the IV estimate of the modal non-VA hospital effect on 28-day mortality by calendar year. The first-stage and reduced-form equations are given in Equations (10) and (11). The overall sample is the same alternative sample designed to study choice among non-VA hospitals for patients with only non-VA utilization in the prior year. Results for the overall IV estimates are shown in Appendix Figure A.9. Details of the sample selection are given in Appendix Table A.13. Estimates are shown in connected dots, while 95% confidence intervals are shown in dashed lines. Panel B of the figure shows the percent of rides going to hospitals after health IT adoption in our analytic sample. Health IT adoption is defined from a dataset from the Office of the National Coordinator of Health Information Technology (ONC). This dataset merges hospital attestation data from the Medicare EHR Incentive Program with certified EHR product information from ONC's Certified Health IT Product List (CHPL), and we code the use of any certified product as health IT adoption.

Table 1: Characteristics of Baseline Sample

	Sample characteristics				
Restrictions	Dually eligible	Add zip × hospital	Add zip × ambulance	Add VA prior utilization	Add no ride in prior month
Male	0.899	0.883	0.863	0.962	0.963
Age	77.04	76.89	76.13	75.62	76.03
Black	0.111	0.163	0.187	0.200	0.194
Income	$22,222	$21,819	$21,339	$20,905	$20,905
Rural zip code	0.255	0.043	0.045	0.050	0.051
Residential Source	0.610	0.600	0.652	0.685	0.705
Comorbidity count	6.53	6.69	6.44	6.54	6.14
Prior VA ED visit	0.136	0.197	0.264	0.565	0.529
Prior Medicare ED visit	0.695	0.675	0.626	0.539	0.482
Ambulance rides in prior year	2.77	3.05	3.25	3.12	2.16
Advanced Life Support	0.696	0.655	0.655	0.674	0.684
Weekend rate	0.272	0.269	0.270	0.270	0.269
28-day mortality	0.115	0.109	0.104	0.100	0.097
Present at VA	0.044	0.088	0.166	0.336	0.330
Number of patients	2,862,557	1,118,302	365,163	188,299	188,299
Number of ambulance rides	8,828,997	3,465,588	1,051,093	491,193	401,319

Note: This table presents characteristics of observations remaining at each step of creating the baseline sample, detailed in Appendix Table A.1.

Table 2: Effect of VA Hospitals on Mortality

	Dependent variable: 28-day mortality				
	(1)	(2)	(3)	(4)	(5)
	A: OLS				
VA hospital	-0.024	-0.023	-0.026	-0.022	-0.021
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Outcome mean	0.097	0.097	0.097	0.097	0.097
Observations	401,319	401,319	401,319	401,319	401,319
	B: IV				
First stage	0.878	0.853	0.839	0.837	0.860
	(0.035)	(0.034)	(0.034)	(0.034)	(0.043)
IV estimate	-0.045	-0.037	-0.036	-0.038	-0.049
	(0.018)	(0.018)	(0.018)	(0.017)	(0.023)
Outcome mean	0.097	0.097	0.097	0.097	0.097
Observations	401,319	401,319	401,319	401,319	401,319
Demographic controls	No	Yes	Yes	Yes	Yes
Comorbidity controls	No	No	Yes	Yes	Yes
Ambulance diagnosis controls	No	No	No	Yes	Yes
Ambulance and co-rider controls	No	No	No	No	Yes

Note: This table shows OLS and IV estimates of the effect of VA hospitals on 28-day mortality. Panel A gives OLS estimates, $\hat{\beta}_{OLS}$, for β in Equation (2). Panel B gives IV estimates, $\hat{\beta}_{IV}$, as well as the first stage coefficient, $\hat{\pi}_1$ in Equation (3), with respect to the leave-out probability of the assigned ambulance company to transport patients to the VA. Baseline controls in all specifications are described in Appendix Table A.2 and include patient zip code dummies, ALS/BLS dummies, source of the ambulance ride, time categories, and patient prior utilization. Additional controls are described in Appendix Table A.3. The estimation sample is described in Appendix Table A.1.

Table 3: Effect of VA Hospitals on Other Outcomes

	Dependent variable				
	Admission	Hospital days	ED revisits	Outpatient visits	Spending
	(1)	(2)	(3)	(4)	(5)
			A: OLS		
VA hospital	-0.004	0.514	-0.036	0.200	932
	(0.003)	(0.045)	(0.007)	(0.017)	(87)
Outcome mean	0.589	4.380	0.318	1.443	12,173
Observations	401,319	401,319	401,319	401,319	401,319
			B: IV		
IV estimate	-0.090	-0.468	0.029	0.379	-2,548
	(0.032)	(0.434)	(0.044)	(0.174)	(822)
Outcome mean	0.589	4.380	0.318	1.443	12,173
Observations	401,319	401,319	401,319	401,319	401,319

Note: This table shows OLS and IV estimates of the effect of VA hospitals on various outcomes. Spending is defined as total spending over the 28 days following the ambulance ride. Panel A gives OLS estimates, $\hat{\beta}_{OLS}$, for β in Equation (2). Panel B gives IV estimates, $\hat{\beta}_{IV}$, as well as the first stage coefficient, $\hat{\pi}_1$ in Equation (3), with respect to the leave-out probability of the assigned ambulance company to transport patients to the VA. Baseline controls in all specifications are described in Appendix Table A.2 and include patient zip code dummies, ALS/BLS dummies, source of the ambulance ride, time categories, and patient prior utilization. Additional controls are described in Appendix Table A.3. The estimation sample is described in Appendix Table A.1.

Table 4: Complier Characteristics

	Overall	Compliers	Ratio
Male	0.963	0.952	0.99
		(0.006)	[0.98 - 1.00]
Age	76.0	74.9	0.99
		(0.433)	[0.97 - 1.00]
Black	0.194	0.257	1.33
		(0.028)	[1.05 - 1.61]
Income	$20,905	$16,972	0.81
		($611)	[0.75 - 0.87]
Rural zip code	0.051	0.091	1.78
		(0.025)	[0.82 - 2.75]
Residential source	0.705	0.647	0.92
		(0.033)	[0.83 - 1.01]
Comorbidity count	6.143	5.447	0.89
		(0.113)	[0.85 - 0.92]
Mental illness	0.43	0.44	1.04
		(0.015)	[0.97 - 1.11]
Substance abuse	0.144	0.163	1.13
		(0.011)	[0.97 - 1.28]
Prior VA ED visit	0.529	0.712	1.35
		(0.012)	[1.30 - 1.39]
Prior Medicare ED visit	0.482	0.336	0.70
		(0.014)	[0.64 - 0.75]
Ambulance rides in prior year	2.156	2.178	1.01
		(0.084)	[0.93 - 1.09]
Advanced Life Support	0.684	0.600	0.88
		(0.024)	[0.81 - 0.95]
Predicted VA user	0.847	0.939	1.11
		(0.004)	[1.10 - 1.12]
Predicted mortality	0.097	0.092	0.94
		(0.004)	[0.87 - 1.02]

Note: This table presents average complier characteristics and the ratio between this average and the average among all veterans in the sample. Average complier characteristics and standard errors are calculated by performing two-stage least squares using the first stage Equation (3) and a reduced-form equation replacing the outcome variable in Equation (4) with $X_i D_i$, where X_i is the characteristic corresponding to ride i. Regressions use baseline controls described in Appendix Table A.2; the regression sample is the baseline sample described in Appendix Table A.1. Standard errors for each average are presented in parentheses. The corresponding 95% confidence intervals for each ratio are presented in brackets.

Table 5: Non-VA Hospital Characteristics

	Hospital Sample		
	National average	Baseline sample	Complier weighted
Basic Characteristics			
Admissions	18,368	18,442	12,736
Average daily census	271	261	180
Total staffed beds	379	371	257
Teaching hospital	0.52	0.50	0.36
Urban location	0.89	0.92	0.86
Payment and Organization			
Capitated lives covered	11,400	8,197	7,111
Network participant	0.46	0.49	0.49
Hospital system	0.61	0.65	0.65
HMO	0.20	0.17	0.15
PPO	0.19	0.20	0.14
ACO	0.09	0.04	0.05
Health IT Adoption			
Adoption by 2011	0.02	0.02	0.02
Adoption by 2012	0.07	0.07	0.07
Adoption by 2013	0.17	0.19	0.17
Adoption by 2014	0.33	0.36	0.34
Spending and Outcomes			
Relative spending	1.00	1.01	1.02
Mortality rate	12.23	12.21	11.97
Readmission rate	18.14	18.16	18.49

Note: This table presents average characteristics of non-VA hospitals in different samples. The national average weights hospital characteristics by their yearly admissions in the American Hospital Association (AHA) Annual Survey. The average in the baseline sample weights hospital characteristics by rides in that sample, described in Appendix Table A.1. The complier-weighted average is calculated by performing two-stage least squares using the first stage Equation (3) and a reduced-form equation replacing the outcome variable in Equation (4) with $X_i D_i$, where X_i is the hospital characteristic corresponding to ride i. Hospital characteristics are described in further detail in Appendix A.5.

Table 6: Modal Hospital Mechanisms

	Dependent variable: 28-day mortality					
	(1)	(2)	(3)	(4)	(5)	(6)
	A: OLS					
Modal hospital	-0.005	-0.006		-0.012	-0.006	
	(0.001)	(0.001)		(0.005)	(0.001)	
× Adoption			-0.006			-0.008
			(0.001)			(0.003)
× No adoption			-0.006			-0.006
			(0.001)			(0.001)
	B: IV					
First stage	0.745	0.689		0.506	0.703	
	(0.011)	(0.008)		(0.026)	(0.007)	
Modal hospital	-0.015	-0.004		-0.011	-0.006	
	(0.009)	(0.006)		(0.034)	(0.005)	
× Adoption			-0.015			-0.015
			(0.006)			(0.019)
× No adoption			-0.005			-0.006
			(0.005)			(0.005)
Outcome mean	0.106	0.113	0.112	0.107	0.112	0.112
Observations	338,313	1,075,528	1,414,197	58,968	1,354,196	1,413,573
Fixed effects						
Hospital identities	Yes	Yes	No	Yes	Yes	No
Hospital ever adopted	N/A	N/A	Yes	N/A	N/A	Yes
Sample	IT adoption	No IT adoption	Full	ACO adoption	No ACO adoption	Full

Note: This table shows OLS and IV estimates of the effect of presenting to a veteran's modal hospital on 28-day mortality, depending on whether the modal hospital has adopted health IT or whether the modal hospital has joined an Accountable Care Organization (ACO). Columns 1 and 2 show results estimated in subsamples defined by whether the modal hospital has adopted health IT or not. Columns 5 and 6 show results estimated in subsamples defined by whether the modal hospital has joined an ACO or not. The first-stage and reduced-form equations for the IV estimation (Panel B) are given in Equations (10) and (11); while this table presents results with hospital fixed effects, results do not qualitatively depend on the inclusion of hospital fixed effects. Columns 3 and 6 present results estimated on the overall sample with interactions for adoption status; these specifications are described in detail in Appendix A.6. We include baseline controls defined in Appendix Table A.2. The overall sample is the same alternative sample designed to study choice among non-VA hospitals for patients with only non-VA utilization in the prior year. Details of the sample selection are given in Appendix Table A.13, Panel B.

Appendix

A.1 IV Validity

A.1.1 Exclusion Restriction

Under the standard assumptions for IV validity in Imbens and Angrist (1994), ambulance companies would be subject to the exclusion restriction, in Condition 1(ii), that they only affect outcomes by whether they transport patients to the VA, and not by other treatments that they may administer during the ambulance ride or by their choice of non-VA hospitals. Following Kolesar et al. (2015), we relax this assumption to allow for differences in potential treatments and non-VA hospital choices across ambulance companies but require that such differences that may affect outcomes are not systematically related to ambulance propensity to transport to the VA.

Specifically, we include controls $\mathbf{C}_i$ that are related to actions by the ambulance after pickup in the first-stage and reduced-form relationships:

$$\begin{aligned} D_i &= \pi_1 Z_i + \mathbf{X}_i^0 \delta_1 + \mathbf{C}_i \eta_1 + \zeta_{1,z(i)} + \varepsilon_{1,i}; \\ Y_i &= \pi_2 Z_i + \mathbf{X}_i^0 \delta_2 + \mathbf{C}_i \eta_2 + \zeta_{2,z(i)} + \varepsilon_{2,i}. \end{aligned}$$

Under each set of ambulance-related controls, we examine the stability of $\hat{\beta}_{IV} = \hat{\pi}_2 / \hat{\pi}_1$.

We consider four sets of controls in $\mathbf{C}_i$. First, we control for splines of ambulance charges reflected in their Medicare claims. Consistent with a health economics literature on productivity and the returns to spending (Doyle et al. 2015; Chandra et al. 2016), we consider charges incurred by the ambulance company as a sufficient statistic for the intensity of treatment during the ride.[35] Second, we control for splines of the mileage of the ride. Third, we control for indicators of the number of non-VA hospitals to which the ambulance company transports patients from a zip code.

Fourth, we control for average measures of non-VA hospitals that the ambulance company delivers its patients to. For each non-VA hospital h, we measure average mortality and spending outcomes $\overline{Y}_h$, among veterans outside of our benchmark analytic sample who *only* have non-VA prior utilization (Panel B of Appendix Table A.13). We also measure the share, w_{jh}, that each ambulance company j delivers patients to each non-VA hospital h, also among veterans with non-VA-only prior utilization. For each ride i, we then control for average non-VA hospital measures of mortality and spending, calculated as $\sum_h w_{j(i),h} \overline{Y}_h$, weighted by the hospital-specific shares of the assigned ambulance $j(i)$. As in Section 5.4, we use information on Medicare claims to infer non-VA hospital spending.

Appendix Table A.4 shows estimates of the VA effect on mortality and on spending, using the same baseline controls as in our benchmark analyses in Section 3 with the addition of various ambu-

[35] In principle, we also observe detailed CPT procedure codes for services rendered during the ambulance ride (e.g., supplemental oxygen, medications, or intravenous fluids). However, in 2002, Medicare changed to a simple payment arrangement that depended only on a few characteristics of the ride, such as ALS vs. BLS level, mileage, and the use of lights and sirens (Centers for Medicare & Medicaid Services 2002). Consistent with this payment policy, detailed CPT codes for extra services are usually missing in the claims data.

lance related controls. We find that results are highly robust to the addition of these controls.

A.1.2 Monotonicity

We test the monotonicity condition in Condition 1(iii) by tests standard in the judges-design literature that demonstrate a positive first-stage relationship across subgroups of observations (Arnold et al. 2018; Bhuller et al. 2020). We define eight pairs of subsamples based on several important patient characteristics: (i) age $\leq$ 80 years vs. age $>$ 80 years; (ii) white vs. non-white race; (iii) comorbidity count above vs. below median; (iv) either vs. neither mental illness or substance abuse present; (v) VA visits in the prior year above vs. below median; (vi) Advanced Life Support vs. Basic Life Support; (vii) prediction of VA user above vs. below median; and (viii) prediction of mortality above vs. below median.

Under monotonicity, we expect that an ambulance that has a higher propensity to transport veterans to the VA should weakly increase the probability of transport to the VA for any set of veterans. Specifically, using the set of observations $\mathcal{I}_m$ for each subsample m, we estimate a first-stage regression with respect to our baseline instrument, Z_i, from Equation (1):

$$D_i = \pi_1^m Z_i + \mathbf{X}_i^0 \delta_1^m + \zeta_{1,z(i)}^m + \varepsilon_{1,i}^m, \tag{A.1}$$

and we assess whether $\hat{\pi}_1^m \geq 0$.

We further assess monotonicity in each subsample m by constructing a "reverse-sample" instrument that only uses observations in the analytical sample (Step 6 in Appendix Table A.1) that are not in $\mathcal{I}_m$:

$$\tilde{Z}_i^{-m} = \frac{1}{\tilde{K}_{j(i)}^{-m}} \sum_{i' \in \tilde{\mathcal{I}}_{j(i)} \setminus \mathcal{I}_m} \frac{\mathbf{1}\left(k\left(i'\right) \neq k\left(i\right)\right) D_{i'}}{\tilde{N}_{k(i'),j(i)}}. \tag{A.2}$$

Within the *analytical* sample, $\tilde{\mathcal{I}}_j$ denotes the set of rides assigned to j, $\tilde{K}_j^{-m}$ is the number of patients assigned to ambulance j without characteristic m, and $\tilde{N}_{k,j}$ is the number of rides by patient k with ambulance j.[36] In each subsample m, we also perform first-stage regressions of the form in Equation (A.1) that use $\tilde{Z}_i^{-m}$ instead of Z_i as the instrument.

Recall that the baseline instrument, Z_i, is computed in the much larger sample of dually eligible veterans (Step 1 in Appendix Table A.1). Since the reverse-sample instruments are based on much smaller patient populations, they may be weaker predictors of underlying ambulance propensities to transport to the VA.

In Appendix Table A.5, we demonstrate a positive and statistically significant first-stage coefficient in every subsample and for both the baseline instrument and the reverse-sample instrument. Coefficient sizes are generally smaller for the reverse-sample instruments. In Appendix Table A.6, we show first-stage relationships using two other instruments that are both based on the smaller ana-

[36] We use the analytical sample construct the reverse-sample instruments, so that the samples used to construct instruments are roughly the same between pairs of characteristics (e.g., subsamples for comorbidity count above vs. below median).

lytical sample. Specifically, we construct a "baseline" instrument, $\tilde{Z}_i$, and an "in-sample" instrument, $\tilde{Z}_i^m$, from the analytical sample:

$$\tilde{Z}_i = \frac{1}{\tilde{K}_{j(i)} - 1} \sum_{i' \in \tilde{\mathcal{I}}_{j(i)}} \frac{\mathbf{1}\left(k\left(i'\right) \neq k\left(i\right)\right) D_{i'}}{\tilde{N}_{k(i'),j(i)}}, \text{ and} \tag{A.3}$$

$$\tilde{Z}_i^m = \frac{1}{\tilde{K}_{j(i)}^m - 1} \sum_{i' \in \tilde{\mathcal{I}}_{j(i)} \cap \mathcal{I}_m} \frac{\mathbf{1}\left(k\left(i'\right) \neq k\left(i\right)\right) D_{i'}}{\tilde{N}_{k(i'),j(i)}}. \tag{A.4}$$

First-stage coefficients for these instruments are also all positive and statistically significant. They are similar in magnitude to the coefficients for the reverse-sample instruments, which suggests that lower signal-to-noise ratios due to smaller sample sizes explain much of decrease in coefficient magnitude for the reverse-sample instruments, compared to the baseline (overall-sample) instrument.

A.2 Statistical Tests of Hazard Functions

A.2.1 Potential Survival Rates and Hazard Rates

Following the notation in Section 4, let $s_{IV}\left(t;d\right) \equiv E\left[S_i\left(t;d\right) | i \in C\right]$ denote the IV estimands of the potential survival rates among compliers, where $d \in \{0,1\}$ indicates outcomes under VA care ($d = 1$) or non-VA care ($d = 0$), for each week $t \in \{0,1,\ldots,52\}$. We then define the corresponding estimands of the potential mortality *hazards* as follows:

$$h_{IV}\left(t;d\right) \equiv \frac{s_{IV}\left(t-1;d\right) - s_{IV}\left(t;d\right)}{s_{IV}\left(t-1;d\right)}.$$

We use two-stage least squares to construct estimates of the potential survivor fractions at each time horizon, $\hat{s}_{IV}\left(t;d\right)$ and then construct the corresponding potential hazard functions, $\hat{h}_{IV}\left(t;d\right)$. We also construct a set of 250 block bootstrap samples (selecting samples by zip code, with replacement), and for replication sample $r \in \{1,\ldots,R\}$, we construct $\hat{s}_{IV}^r\left(t;d\right)$ and $\hat{h}_{IV}^r\left(t;d\right)$. Using these samples we construct the mean estimated potential hazard for each week across the replications:

$$\overline{h}_{IV}^B\left(t;d\right) = \frac{1}{R}\sum_r \hat{h}_{IV}^r\left(t;d\right). \tag{A.5}$$

We also construct the standard deviation of the bootstrap-estimated potential hazard for each week:

$$\hat{\sigma}_{IV}^B\left(t;d\right) = \sqrt{\frac{1}{R-1}\sum_r \left[\hat{h}_{IV}^r\left(t;d\right) - \hat{h}_{IV}^B\left(t;d\right)\right]^2}. \tag{A.6}$$

We construct similar objects for potential survival and hazard rates under OLS: $\hat{s}_{OLS}\left(t;d\right)$ and $\hat{h}_{OLS}\left(t;d\right)$, respectively. Using the same set of block bootstrap samples, we compute $\hat{s}_{OLS}^r\left(t;d\right)$ and $\hat{h}_{OLS}^r\left(t;d\right)$ in each bootstrap replication sample r. We similarly construct the mean estimated potential OLS hazard for each week across replications, $\overline{h}_{OLS}^B\left(t;d\right)$, and the standard deviation of

bootstrap-estimated potential hazards, $\hat{\sigma}^{B}_{OLS}(t;d)$, in which we use OLS hazards $\hat{h}^{r}_{OLS}(t;d)$ in formulas otherwise the same as Equations (A.5) and (A.6).

A.2.2 Test of Mortality Displacement

To detect "mortality displacement" (Schwartz 2000), in which deaths of VA patients are simply delayed, we test the joint null hypothesis that $h_{IV}(t;1) \leq h_{IV}(t;0)$ for all $t \geq 1$. This null hypothesis states that the mortality hazard under the VA never overtakes the mortality hazard under non-VA hospitals, even in later periods, and it is consistent with no mortality displacement.

Restating the null hypothesis as

$$H_{0,1} : h_{IV}(t;0) - h_{IV}(t;1) \geq 0, \text{ for all } t \geq 1, \tag{A.7}$$

we use estimates $\hat{h}_{IV}(t;0) - \hat{h}_{IV}(t;1)$ and consider the following test statistic of the null, based on Wolak (1987):

$$Q_1 \equiv \sum_{t=1}^{52} w_{1,t} \mathbf{1}\left(\hat{h}_{IV}(t;0) - \hat{h}_{IV}(t;1) < 0\right)\left(\hat{h}_{IV}(t;0) - \hat{h}_{IV}(t;1)(t)\right)^2, \tag{A.8}$$

where $w_{1,t}$ is a strictly positive weight. This test statistic penalizes only negative differences $\hat{h}_{IV}(t;0) - \hat{h}_{IV}(t;1) < 0$ that can be consistent with the null hypothesis that $\hat{h}_{IV}(t;0) - \hat{h}_{IV}(t;1) \geq 0$, for all $t \geq 1$, only by statistical noise.

To derive a critical value for Q_1, we use our bootstrap sample to form a set of recentered bootstrap estimates of the potential hazards at each week:

$$\begin{aligned} \tilde{h}^{r}_{IV}(t;0) &= \hat{h}^{r}_{IV}(t;0) - \overline{h}^{B}_{IV}(t;0); \\ \tilde{h}^{r}_{IV}(t;1) &= \hat{h}^{r}_{IV}(t;1) - \overline{h}^{B}_{IV}(t;1). \end{aligned}$$

We then construct the empirical distribution of the test statistic, in Equation (A.8), under the recentered bootstrap deviations:

$$Q^{r}_1 \equiv \sum_{t=1}^{52} w_{1,t} \mathbf{1}\left(\tilde{h}^{r}_{IV}(t;0) - \tilde{h}^{r}_{IV}(t;1) < 0\right)\left(\tilde{h}^{r}_{IV}(t;0) - \tilde{h}^{r}_{IV}(t;1)\right)^2. \tag{A.9}$$

We take the 95th percentile of this distribution as the critical value above which our test statistic Q_1 can reject the null hypothesis $H_{0,1}$, in Equation (A.7).

Following Wolak (1987), this distribution is formed under the data generating process implied by the "least favorable null" for testing joint inequality constraints (Perlman 1969). Specifically, we consider the least favorable data generating process that satisfies the null hypothesis H_0, in Equation (A.7), which is

$$\underline{H}_{0,1} : h_{IV}(t;0) - h_{IV}(t;1) = 0, \text{ for all } t \geq 1. \tag{A.10}$$

If we obtain a test statistic Q_1 with improbable negative deviations that reject the least favorable null hypothesis $\underline{H}_{0,1}$ in Equation (A.10), then we can also reject the null hypothesis $H_{0,1}$ in Equation (A.7).

We use the same weights $w_{1,t}$ in Equations (A.8) and (A.9) and set them as the inverse of the estimated sampling variance of the recentered deviations:

$$w_{1,t}^{-1} = \frac{1}{R-1} \sum_r \left(\tilde{h}_{IV}^r (t;0) - \tilde{h}_{IV}^r (t;1) \right)^2 . \tag{A.11}$$

These weights standardize the statistical distribution of $\hat{h}_{IV} (t;0) - \hat{h}_{IV} (t;1)$, so that the test statistic distribution can be considered as chi-squared. Although we use critical values derived from the bootstrap distribution, we find the scale of our test statistic to be more intuitive with this normalization.[37]

We show results in Panel A of Appendix Figure A.4. We find that Q_1 is within the distribution of bootstrapped values of Q_1^r. We therefore cannot reject the null of no mortality displacement.

A.2.3 Extended Test of IV Validity

In addition to standard tests of IV validity that are based on observable characteristics—including tests of balance in Section 3.2 and monotonicity in Appendix A.1.2—we develop a tractable extended test of IV validity using the insights in Balke and Pearl (1997) and Heckman and Vytlacil (2005, Proposition A.5) that are based on *potential outcomes*.

Kitagawa (2015) summarizes these insights as follows for a binary instrument $Z \in \{0,1\}$, a binary treatment $D \in \{0,1\}$ (increasing in probability with Z), and an outcome $Y \in \mathcal{Y}$. For any Borel set B in $\mathcal{Y}$, IV validity in Condition 1 implies that

$$\Pr(Y \in B, D = 1 | Z = 1) - \Pr(Y \in B, D = 1 | Z = 0) \geq 0; \tag{A.12}$$

$$\Pr(Y \in B, D = 0 | Z = 0) - \Pr(Y \in B, D = 0 | Z = 1) \geq 0. \tag{A.13}$$

Kitagawa (2015, Proposition 1.1) further states that tests of Equations (A.12) and (A.13) constitute the strongest possible tests of IV validity in the sense that no other feature of the data can contribute further to screening out invalid instruments.[38]

We note that, given the approach in Abadie (2002), testing Equations (A.12) and (A.13) is alge-

[37] Wolak (1987) proposes to use an optimal minimum distance test statistic that would use the full covariance matrix of $\delta(t)$. We avoid this formulation due to finite-sample issues that would cause this covariance matrix to be poorly estimated by the full covariance matrix of $\tilde{\delta}^r(t)$, noted by Altonji and Segal (1996). Results are qualitatively similar when we choose a weight of $w_t = 1$ for all t, but we find that using w_t from Equation (A.11)—i.e., normalizing each $\delta(t)$ by its bootstrap standard error—affords greater power in rejecting the null. This approach is equivalent to our best estimate of a diagonal covariance matrix in place of the full covariance matrix.

[38] Chan et al. (2019) provides an applied example, in the setting of radiologists, in which standard monotonicity tests in Appendix A.1.2 are satisfied but a simple version of this extended test of validity is not satisfied. They find that radiologists who diagnose more cases with pneumonia do so in a wide range of subgroups of patients defined by observable characteristics (i.e., standard tests of monotonicity) but that the same radiologists who diagnose more cases with pneumonia are more likely to miss cases of pneumonia (i.e., $\Pr(|Y \in B, D = 0\ Z = 0) - \Pr(|Y \in B, D = 0\ Z = 1) < 0$).

braically equivalent to testing, for all $B \subset \mathcal{Y}$,

$$\Pr\left(Y_i\left(0\right) \in B | i \in C\right) \geq 0; \tag{A.14}$$

$$\Pr\left(Y_i\left(1\right) \in B | i \in C\right) \geq 0. \tag{A.15}$$

Thus we use the Abadie (2002) approach to define a partition of mortality outcomes $\mathcal{Y}$ in terms of weekly hazard rates, by the date of death (if any) following the ambulance ride. Such a partition implies that potential hazard rates among compliers, $h_{IV}\left(t;d\right)$, are non-negative in every week $t \in \{1,\ldots,52\}$ under both VA assignment ($d=1$) and non-VA assignment ($d=0$).

That is, our extended test of IV validity amounts to testing the following joint null hypothesis of inequality constraints:

$$H_{0,2} : h_{IV}\left(t;d\right) \geq 0, \text{ for all } t \geq 1, d \in \{0,1\}. \tag{A.16}$$

Following a similar approach as for mortality displacement in Appendix A.2.2, our test statistic is

$$Q_2 \equiv \sum_{d=0}^{1}\sum_{t=1}^{52} w_{2,t}\mathbf{1}\left(\hat{h}_{IV}\left(t;d\right) < 0\right)\left(\hat{h}_{IV}\left(t;d\right)\right)^2,$$

where $w_{2,t}^{-1} = \left(\hat{\sigma}_{IV}^{B}\left(t;d\right)\right)^2$. We obtain the critical value for our test statistic by the distribution of recentered bootstrapped estimates, defined above. For the rth bootstrap replication, the test statistic is

$$Q_2^r \equiv \sum_{d=0}^{1}\sum_{t=1}^{52} w_{2,t}\mathbf{1}\left(\tilde{h}_{IV}^{r}\left(t;d\right) < 0\right)\left(\tilde{h}_{IV}^{r}\left(t;d\right)\right)^2.$$

We take the 95th percentile of the distribution of Q_2^r across replications $r \in \{1,\ldots,R\}$ as the critical value for Q_2. As above, this test of inequality constraints is based upon a least favorable null hypothesis. In this case, the least favorable null hypothesis is

$$\underline{H}_{0,2} : h_{IV}\left(t;d\right) = 0, \text{ for all } t \geq 1, d \in \{0,1\}. \tag{A.17}$$

We show results in Panel B of Appendix Figure A.4. We find that Q_2 is lower than any bootstrapped value of Q_2^r. This suggests not only that we cannot reject the null hypothesis $H_{0,2}$ in Equation (A.16), but also that the realized data are significantly more favorable than the least favorable null hypothesis $\underline{H}_{0,2}$ in Equation (A.17). In other words we can strongly reject the null that $h_{IV}\left(t;d\right) = 0$, for all $t \geq 1, d \in \{0,1\}$, which means that $h_{IV}\left(t;d\right) > 0$ for some $t \geq 1, d \in \{0,1\}$.

A.2.4 Tests of Hazard Rate Equality

We finally perform tests of the equality of hazard rates after the first week after the ambulance ride. Comparing hazard rates across different groups of veterans, we aim to shed light on heterogeneity in longer-term mortality risk across these groups. To define these tests generally, consider two sets of hazard rates, $h_1\left(t\right)$ and $h_2\left(t\right)$, for $t \geq 2$. We consider two types of null hypothesis.

First, we assess mean differences in hazard rates between $\{h_1(t)\}_t$ and $\{h_2(t)\}_t$, for $t \geq 2$, under the null hypothesis that the mean hazard rate is the same between the two sets:

$$H_{0,3} : \frac{1}{51} \sum_{t=2}^{52} (h_1(t) - h_2(t)) = 0. \tag{A.18}$$

We test this null hypothesis by comparing $\frac{1}{51} \sum_{t=2}^{52} \left(\hat{h}_1(t) - \hat{h}_2(t) \right)$ against the bootstrapped distribution of recentered differences. Specifically, for replication $r \in \{1, \ldots, R\}$, denote the bootstrap-estimate hazard rates of $(h_1(t), h_2(t))$ as $\left(\hat{h}_1^r(t), \hat{h}_2^r(t) \right)$. Define the recentered bootstrap hazard rate as

$$\begin{aligned} \tilde{h}_1^r(t) &\equiv \hat{h}_1^r(t) - \overline{h}_1^B(t) \text{ and} \\ \tilde{h}_2^r(t) &\equiv \hat{h}_2^r(t) - \overline{h}_2^B(t), \end{aligned}$$

where $\overline{h}_1^B(t) \equiv \frac{1}{R} \sum_r h_1(t)$ and $\overline{h}_2^B(t) \equiv \frac{1}{R} \sum_r h_2(t)$. The distribution of $\left\{ \frac{1}{51} \sum_{t=2}^{52} \left(\tilde{h}_1^r(t) - \tilde{h}_2^r(t) \right) \right\}_r$ determines the two-sided critical values for the mean hazard difference. By construction, this distribution will have mean 0.

Second, we consider the joint null hypothesis that the difference between each pair of hazards is equal to 0:

$$H_{0,4} : h_1(t) - h_2(t) = 0, \text{ for all } t \geq 2. \tag{A.19}$$

Using estimates $\hat{h}_1(t) - \hat{h}_2(t)$, we construct the following test statistic:

$$Q_4(h_1(\cdot), h_2(\cdot)) \equiv \sum_{t=2}^{52} w_{4,t} \left(\hat{h}_1(t) - \hat{h}_2(t) \right)^2.$$

We compute the empirical distribution of Q_4 under the null hypothesis by using recentered differences $\tilde{h}_1^r(t) - \tilde{h}_2^r(t)$. Each bootstrap replication r yields

$$Q_4^r(h_1(\cdot), h_2(\cdot)) \equiv \sum_{t=2}^{52} w_{4,t} \left(\tilde{h}_1^r(t) - \tilde{h}_2^r(t) \right)^2.$$

We take the 95th percentile of the distribution of Q_4^r across replications $r \in \{1, \ldots, R\}$ as the critical value for Q_4. We set $w_{4,t}^{-1} = \frac{1}{R-1} \sum_r \left(\widetilde{h}_1^r(t) - \widetilde{h}_2^r(t) \right)^2$ to standardize the distribution of $\hat{h}_1(t) - \hat{h}_2(t)$.

In Appendix Figures A.5 and A.6, we consider five comparisons of hazard rates, for $t \geq 2$, under the null hypotheses of Equations (A.18) and (A.19), respectively. First, we test the null hypothesis that $h_{IV}(t;1) - h_{IV}(t;0) = 0$, for all $t \geq 2$. Under quasi-experimental assignment of compliers (Condition 1), we expect not to reject this null if longer-term hazard rates reflect underlying health. Second, we test the null hypothesis that $h_{OLS}(t;1) - h_{OLS}(t;0) = 0$, for all $t \geq 2$. While we show stability of OLS results in Figure 2, this test may reveal differences in underlying health between veterans assigned to the VA and those assigned to a non-VA hospital that are not captured by observable

patient characteristics.

Third, we test the null hypothesis that $h_{IV}(t;1) - h_{OLS}(t;1) = 0$, for all $t \geq 2$. This reveals differences in underlying health between compliers and VA-assigned veterans, which includes compliers and always takers. Fourth, we similarly test the null hypothesis that $h_{IV}(t;0) - h_{OLS}(t;0) = 0$, for all $t \geq 2$. This reveals differences in underlying health between compliers and non-VA-assigned veterans, which includes compliers and never takers.

A.3 Non-Complier Characteristics

In this appendix section, we describe a simple approach to calculate characteristics of non-compliers, following Dahl et al. (2014), and we discuss results. In our approach, we first residualize the leave-out ambulance propensity to transport to the VA, Z_i, by our key controls, $\left(z(i), \mathbf{X}_i^0\right)$. Denote this residual as Z_i^*. We categorize always takers as rides with Z_i^* below the 20th percentile that still went to the VA ($D_i = 1$). We categorize never takers as rides with Z_i^* above the 80th percentile that still did not go to the VA ($D_i = 0$).

Among each group of always takers and never takers, we compute characteristics along the same dimensions as those in our compliers analysis, in Table 4. Specifically, for each characteristic, we compute mean values among the group of always takers and among the group of never takers, and we compare these means with the overall mean by a ratio. We compute standard errors of these means by drawing bootstrapped samples, blocked by zip code, and repeating this procedure with each bootstrapped sample.

As shown in Appendix Table A.7, we mostly find results that are consistent with our earlier results of complier characteristics and the fact that the majority of non-compliers are never takers: Characteristics that are more common among compliers tend to be more common among always takers and less common among never takers. For expositional brevity, when we omit mention of a never-taker characteristic, we mean that they are the relative opposite of the corresponding always-taker characteristic. Compared to the overall population, always takers are more likely to be Black and have lower income. Always takers are more likely to have mental illness, and they have a slightly higher rate of substance abuse, though the latter is not statistically significant. Always takers are more likely to have prior VA ED visits and less likely to have prior non-VA ED visits.

A.4 Marginal and Average Treatment Effects

Consider the probability of going to the VA as a function of our instrument Z_i and key controls $\left(z(i), \mathbf{X}_i^0\right)$: $P(Z_i)$, where we have omitted the key controls for brevity. Following Heckman and Vytlacil (2005), we can state the treatment rule as

$$D_i = \mathbf{1}\left(P(Z_i) \geq U_i\right), \tag{A.20}$$

where U_i is uniformly distributed in the interval $(0,1)$. Individuals with low U_i relative to $\underline{p} \equiv \arg\min_i P(Z_i)$ are always takers, while individuals with high U_i relative to $\overline{p} \equiv \arg\max_i P(Z_i)$ are never takers.

In this appendix, we estimate two objects relative to selection, as defined by $U_i \sim U(0,1)$. The marginal treatment effect (MTE) for rides with $U_i = u$ is

$$MTE(u) \equiv E[Y_i(1) - Y_i(0) | U_i = u].$$

The average treatment effect (ATE) is

$$ATE = \int_0^1 MTE(u)\, du.$$

We estimate $MTE(u)$, for $u \in \left[\underline{p}, \overline{p}\right]$, using variation in the propensity of ambulances to transport to the VA. We estimate the ATE by extrapolating $MTE(u)$ to $u \in [0,1]$ with a control function approach.

A.4.1 Marginal Treatment Effects

We first estimate marginal treatment effects using a local instrumental variables approach that exploits outcomes along the distribution of ambulance propensity to transport to the VA. The intuition for this approach is that $MTE(u)$ can be stated as

$$MTE(u) = \frac{\partial}{\partial p} E[Y_i | P(Z_i) = u].$$

That is, if mortality decreases linearly with ambulance propensity to transport to the VA, then the data would be consistent with constant treatment effects. On the other hand, if mortality decreases at a faster rate for lower $P(Z_i)$, then the data would suggest "selection on gains," in which veterans who are more likely to benefit from VA care are also more likely to be transported to the VA given a set of ambulances. The visual IV relationship in Appendix Figure A.1 suggests a slightly convex shape in the relationship between mortality and $P(Z_i)$, which implies selection on gains.

We proceed with estimating a flexible relationship between Y_i and $P(Z_i)$ as follows. We compute $P(Z_i) = \hat{D}_i$ from the first-stage Equation (3). We then residualize $\hat{D}_i$ by baseline controls, defined in Appendix Table A.2, and denote the residual as $\hat{D}_i^*$. We similarly residualize Y_i by baseline controls and denote the residual as Y_i^*. For interpretation, we set Y_i^* and $\hat{D}_i^*$ to have the same respective means as Y_i and D_i. A regression of Y_i^* on $\hat{D}_i^*$ yields a point estimate that is numerically identical to the IV estimate $\hat{\beta}_{IV}$.[39]

Rather than fitting a straight line through points $\left(\hat{D}_i^*, Y_i^*\right)$, we fit a flexible function with Gaussian basis splines with four knots (k_1, k_2, k_3, k_4) corresponding to the 5th, 35th, 65th, and 95th percentiles

[39] This regression corresponds to the indirect least squares version of IV and is also numerically identical to the visual IV coefficient that corresponds to the two-stage least squares version of IV.

of $\hat{D}_i^*$. Specifically, for each ride i, we form five basis functions

$$f_n(p) = \exp\left(-(k_n - k_{n-1})(p - c_n)^2\right),$$

where $c_n = \frac{1}{2}(k_{n-1} + k_n)$, $k_0 = \min \hat{D}_i^*$, and $k_5 = \max \hat{D}_i^*$. We regress

$$Y_i^* = \sum_{n=1}^{5} \gamma_n f_n\left(\hat{D}_i^*\right) + \varepsilon_i$$

and form a flexible prediction $\hat{Y}^*(p) = \sum_{n=1}^{5} \hat{\gamma}_n f_n(p)$.

This prediction yields a convenient analytical derivative for the MTE

$$\widehat{MTE}(u) = \sum_{n=1}^{5} \hat{\gamma}_n f_n'(u) = -\sum_{n=1}^{5} 2(k_n - k_{n-1})^2 (u - c_n) \hat{\gamma}_n f_n(u).$$

For each $p \in [0.05, 0.20]$, corresponding to the range of $\hat{D}_i^*$, we compute 95% confidence intervals of $\hat{Y}^*(p)$ by taking the standard deviations of $\hat{Y}^*(p)$ across 50 bootstrapped iterations (with samples drawn by zip code, with replacement). Similarly, for each $u \in [0.05, 0.20]$, we compute 95% confidence intervals of $\widehat{MTE}(u)$ by taking the standard deviations of $\widehat{MTE}(u)$ across these same bootstrapped iterations. We display both $\hat{Y}^*(p)$ and $\widehat{MTE}(u)$ in Appendix Figure A.7.

A.4.2 Average Treatment Effect

In order to estimate the ATE, we adopt a control function model in order to extrapolate treatment effects to non-compliers. Specifically, we model potential outcomes as

$$E[Y_i(d) | U_i = u] = \alpha_d + \gamma_d (J(u) - \mu_J) + \mathbf{X}_i^0 \delta + \zeta_{z(i)}, \tag{A.21}$$

where $d \in \{0, 1\}$ and $u \in (0, 1)$. $J(u)$ is a strictly increasing, continuous function that maps selection to potential outcomes, and $\mu_J \equiv E[J(U_i)]$. Since $E[J(u) - \mu_J] = 0$, we can interpret $\alpha_1 - \alpha_0$ as the ATE. Kline and Walters (2019) show that the control function model in Equations (A.20) and (A.21) can also rationalize the Imbens and Angrist (1994) LATE that we estimate in Section 3, regardless of the choice of $J(u)$.[40]

For our baseline specification, we adopt the linear selection function of $J(u) = u$ from Olsen (1980), which we use with Equation (A.21) to state the following expectation, conditional on the

[40] Kline and Walters (2019) show algebraic equivalence between the control function LATE implied by Equation (A.21), $\underline{p}$, and $\overline{p}$, when the instrument is binary and there are no controls. They also generalize their result for multivalued instruments. With controls, the equivalence may not hold in the standard regression approach in which controls are treated as additively separable but will hold under a propensity score approach.

first-stage error $\varepsilon_{1,i}$ from Equation (3):[41]

$$\begin{aligned} E\left[Y_i | D_i = d, \varepsilon_{1,i} = \varepsilon\right] &= \alpha_d + \gamma_d E\left[J(u) - \mu_J | D_i = d, \varepsilon_{1,i} = \varepsilon\right] + \mathbf{X}_i^0 \delta + \zeta_{z(i)} \\ &= \alpha_d - \gamma_d \frac{\varepsilon}{2} + \mathbf{X}_i^* \delta + \zeta_{z(i)}. \end{aligned} \tag{A.22}$$

This expectation corresponds to the following regression:

$$Y_i = \alpha_\Delta D_i + \gamma_0 \left(-\frac{\hat{\varepsilon}_{1,i}}{2}\right) + \gamma_\Delta \left(-\frac{\hat{\varepsilon}_{1,i}}{2}\right) D_i + \mathbf{X}_i^0 \delta + \zeta_{z(i)} + v_i, \tag{A.23}$$

plugging in the estimated first-stage residual $\hat{\varepsilon}_{1,i}$ from Equation (3). We can compute the ATE from this equation as $\alpha_\Delta = \alpha_1 - \alpha_0$. We estimate Equation (A.23) by OLS to yield $\hat{\alpha}_\Delta = -0.037$, slightly smaller in magnitude than the LATE estimate of -0.041 from Section 3. For inference on the difference between the ATE and the LATE, we recover a numerically equivalent LATE with the following control function regression:[42]

$$Y_i = \beta_{CF} D_i + \gamma \hat{\varepsilon}_{1,i} + \mathbf{X}_i^0 \delta_0 + \zeta_{0,z(i)} + v_i, \tag{A.24}$$

where $\hat{\beta}_{CF}$ is estimated by OLS and is numerically equivalent to $\hat{\beta}_{IV}$ estimated by two-stage least squares. For each bootstrapped replication, we estimate both the ATE, $\hat{\alpha}_1 - \hat{\alpha}_0$, and its difference with the LATE, $\hat{\beta}_{CF}$, in order to obtain standard errors on both the ATE and the difference.

We also examine semiparametric specifications that allow for flexible relationships between the first-stage residual and the structural error term. These alternative specifications allow nonlinear relationships of $g_d(\varepsilon) \equiv E\left[\varepsilon_{0,i} \middle| D_i = d, \varepsilon_{1,i} = \varepsilon\right]$, where $\varepsilon_{0,i}$ is the structural error term in Equation (2). Specifically, we estimate regressions of the following form:

$$Y_i = \alpha_\Delta D_i + g_0\left(\hat{\varepsilon}_{1,i}\right)(1 - D_i) + g_1\left(\hat{\varepsilon}_{1,i}\right) D_i + \mathbf{X}_i^0 \delta + \zeta_{z(i)} + v_i, \tag{A.25}$$

where $g_d\left(\hat{\varepsilon}_{1,i}\right)$, $d \in \{0,1\}$, are flexible functions of the first-stage residual that are non-zero when $D_i = 0$ and $D_i = 1$, respectively. To estimate $g_d\left(\hat{\varepsilon}_{1,i}\right)$, $d \in \{0,1\}$, we use a vector of restricted cubic spline functions or Gaussian basis functions, with three or five knots. Ensuring that $E\left[g_d\left(\hat{\varepsilon}_{1,i}\right)\right] = 0$ by demeaning each spline or basis function, we can interpret α_Δ as the ATE.

In Appendix Table A.8, we show estimates of the ATE and the ATE-LATE difference. ATE estimates are all smaller in magnitude than the LATE estimate from Section 3. We compute standard errors on this difference with 50 bootstrapped iterations (selecting samples by zip code, with replacement). The ATE-LATE difference is statistically significant in our baseline specification in Equation

[41] To see this, assume that the first stage regression in Equation (3) estimates a well-behaved $P(Z_i) \in (0,1)$ such that $D_i = P(Z_i) + \varepsilon_{1,i}$. Define $\lambda_d(p) \equiv E[J(U_i) - \mu_J | D_i = d, P(Z_i) = p]$. We have $\lambda_1(p) = \frac{p}{2} - \frac{1}{2} = \frac{p-1}{2}$, and $\lambda_0(p) = \frac{p+1}{2} - \frac{1}{2} = \frac{p}{2}$. Note that $\lambda_d(p) = \frac{p-d}{2} = \frac{-\varepsilon}{2}$, where $\varepsilon \equiv d - p$. This implies that $\varepsilon_{1,i} = D_i - P(Z_i)$ is a sufficient statistic for $(D_i, P(Z_i))$, and we can state the expectation $J(U_i) - \mu_J$ conditional on $\varepsilon_{1,i}$: $E\left[J(U_i) - \mu_J | \varepsilon_{1,i} = \varepsilon\right] = -\frac{\varepsilon}{2}$.

[42] Blundell and Matzkin (2014) attribute the first proof of this equivalence between control function and two-stage least squares approaches to estimating the LATE to Telser (1964).

(A.23), though they are not statistically significant in the semiparametric specifications.

A.5 Hospital Characteristics

In this appendix, we provide further details on hospital characteristics that we use in our heterogeneity analyses in Section 5.3. For each zip code and year, we use characteristics of the closest VA hospital and a weighted average of the characteristics of associated non-VA hospitals. Weights for each non-VA hospital are proportional to the number of ambulance rides originating from a given zip code to the hospital in that year. Unless otherwise noted, characteristics are observed at the hospital-year level.

We use the American Hospital Association (AHA) Annual Survey to collect the following VA and non-VA hospital characteristics at the hospital-year level: (i) number of ED visits; (ii) total number of facility admissions; (iii) number of available hospital beds; (iv) teaching hospital status; (v) trauma center status; (vi) ED staff full-time equivalents (FTEs), which we use to construct ED staff per 100 ED visits given (i); (vii) nurse FTEs, which we use to construct nurses per 100 admissions given (ii); (viii) hospitalist FTEs, which we use to construct hospitalists per 100 admissions; and (ix) intensivist FTEs, which we use to construct intensivists per 100 admissions given (ii).

We construct a measure of advanced cardiac care, which we define as either the capability to perform interventional cardiac catheterization or cardiac surgery as measured by the AHA Annual Survey (at the hospital-year level) or listing as an ST-Elevation Myocardial Infarction (STEMI) center by the American Heart Association (at the hospital level). We record whether each hospital is certified as a Primary Stroke Center according to the Joint Commission, the American Heart Association, and the American Stroke Association (at the hospital level).

For VA hospitals, we form measures of relative spending from the average cost of an inpatient-day, available from the VA Health Economics Resource Center (HERC). For non-VA hospitals, we use data from `Data.Medicare.gov` on Medicare spending per beneficiary at the hospital level. Similarly, we obtain mortality and readmission rates from `Data.Medicare.gov` for non-VA hospitals and from the VA's Strategic Analytics for Improvement and Learning (SAIL). For each hospital's mortality rate, we take the mean of all available 30-day mortality rates, including disease-specific rates such as heart attack and pneumonia; we form similar means for each hospital's readmission rate based on available 30-day readmission rates, including disease-specific rates. Because some years are missing mortality or readmission rates, for each hospital and rate, we first form averages across years at the hospital level.

For measures of non-VA hospital organization, we use AHA Annual Survey measures of network status, hospital system status, and health maintenance organization (HMO) affiliation. We also obtain whether the hospital participates in an Affordable Care Organization (ACO) from the Medicare Shared Savings Program (MSSP) ACO provider-level dataset. We measure health IT adoption for each hospital and year from electronic health record certified products measured in `healthIT.gov`. Additional characteristics in Table 5 are also obtained from the AHA Annual Survey: (i) average daily

census, (ii) urban location (i.e., hospital is not classified as either "micro" or rural), (iii) capitated lives covered, and (iv) Preferred Provider Organization (PPO) affiliation.

A.6 Modal-Hospital Mechanisms

In Section 5.5, we investigate the role of health IT and integrated care in improving outcomes for patients with prior care at a given non-VA hospital. As described in Appendix A.5, we measures dates of hospital health IT adoption or ACO participation. During our sample period a sizable proportion of hospitals adopted health IT and, to a much lesser extent, participated in an ACO. In Figure 6, we show that the survival effect of a veteran being transported to his modal hospital emerges after the passage the HITECH Act of 2009. This law led to a rapid rise in electronic medical record systems in US hospitals, which had previously been close to absent among non-VA hospitals.

To investigate this further, we focused on four subsamples defined by whether or not each veteran's modal hospital had adopted health IT, at the time of his ambulance ride, and similarly by whether or not each veteran's modal hospital had joined an ACO. In each of these subsamples, we performed the same IV regression of the effect of transport to a veteran's modal hospital. Results are shown in Table 6, Columns 1, 2, 4, and 5. We obtain all of these results after adding hospital fixed effects in the first-stage and reduced-form regressions in Equations (10) and (11), respectively. Results are qualitatively unchanged regardless of their inclusion.

In Columns 3 and 6 of Table 6, we also perform regressions in the overall sample (described in Panel B of Appendix Table A.13). We maintain all of the interactions implicit in our subsample results except that we allow hospital fixed effects to remain constant before and after adoption of health IT or an ACO. We do so with the following control function approach. First, we estimate a first-stage regression that interacts everything with adoption status, except for hospital fixed effects:

$$D_i^m = \sum_{a\in\{0,1\}} \mathbf{1}\left(\text{Adopted}_i = a\right)\left(\pi_{1,a}^m Z_i^m + \gamma_{1,a}^m \overline{Z}_i^m + \mathbf{X}_i^0 \delta_{1,a}^m + \zeta_{1,z(i),a}^m\right) + \xi_{1,h(i)}^m + \varepsilon_{1,i}^m. \tag{A.26}$$

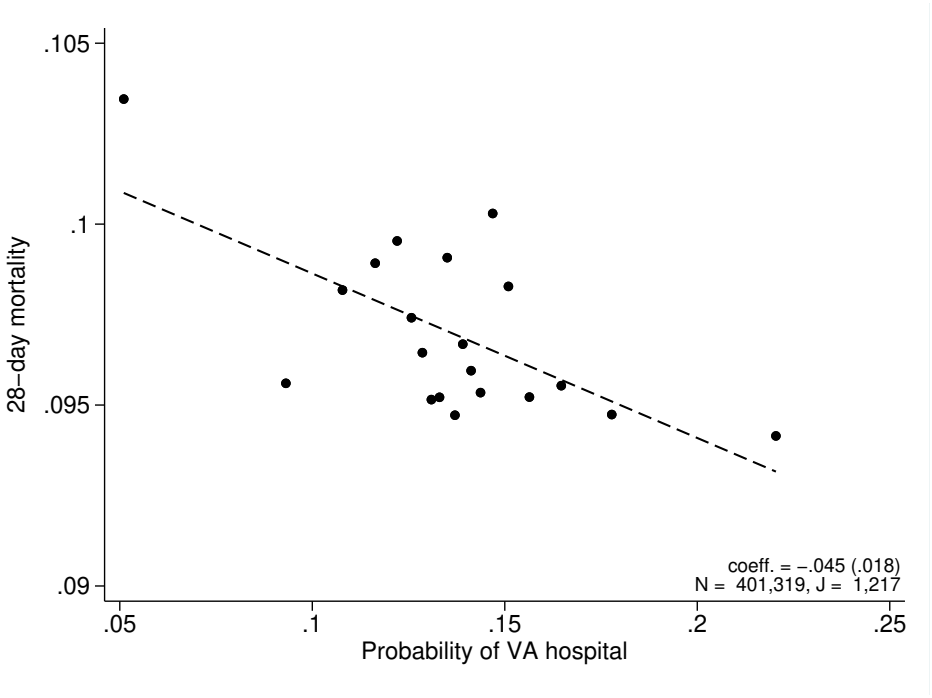
We then take estimated first-stage residuals $\hat{\varepsilon}_{1,i}^m$ and include them in an interacted control-function model:

$$Y_i = \sum_{a\in\{0,1\}} \mathbf{1}\left(\text{Adopted}_i = a\right)\left(\beta_a D_i^m + \gamma_a \hat{\varepsilon}_{1,i}^m + \mathbf{X}_i^0 \delta_a + \zeta_{z(i),a}\right) + \xi_{h(i)} + \epsilon_i. \tag{A.27}$$

As with our other control-function regressions, we compute standard errors by 50 bootstrapped iterations, drawing samples by zip code blocks, with replacement.

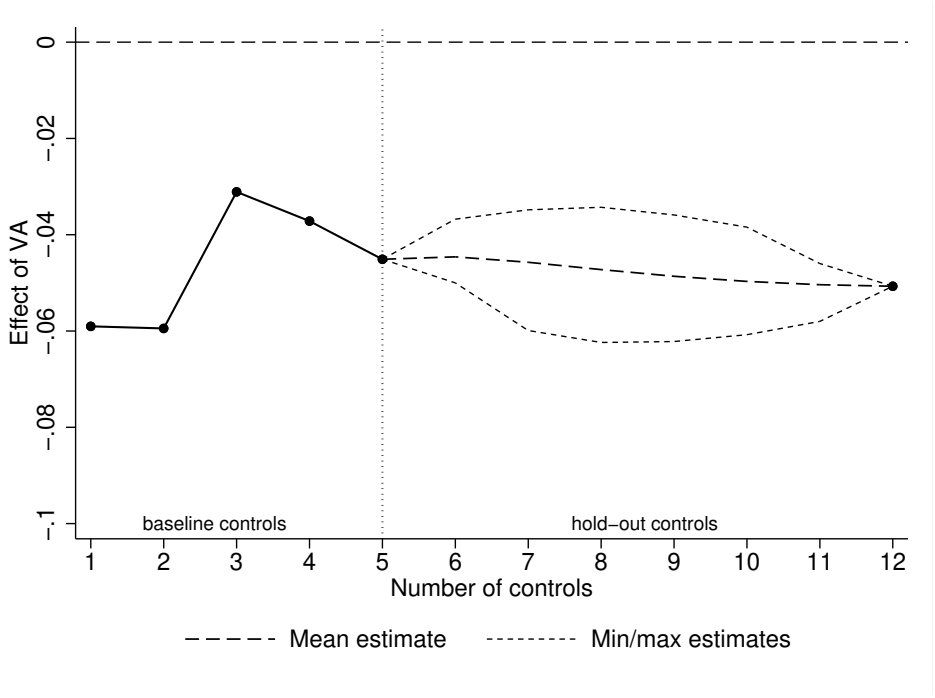
Note that, if we omit hospital fixed effects from Equations (A.26) and (A.27), the coefficients β_a from the control-function regression will numerically equal coefficients from IV regressions, omitting hospital fixed effects, in subsamples either with observations from the "adoption sample" or with only observations from the "no-adoption sample." The inclusion of uninteracted hospital fixed effects in Equations (A.26) and (A.27) allows us to model hospital fixed effects that are the same before and after adoption, whereas fixed effects in each subsample regressions are not linked across subsamples.

Figure A.1: Visual IV



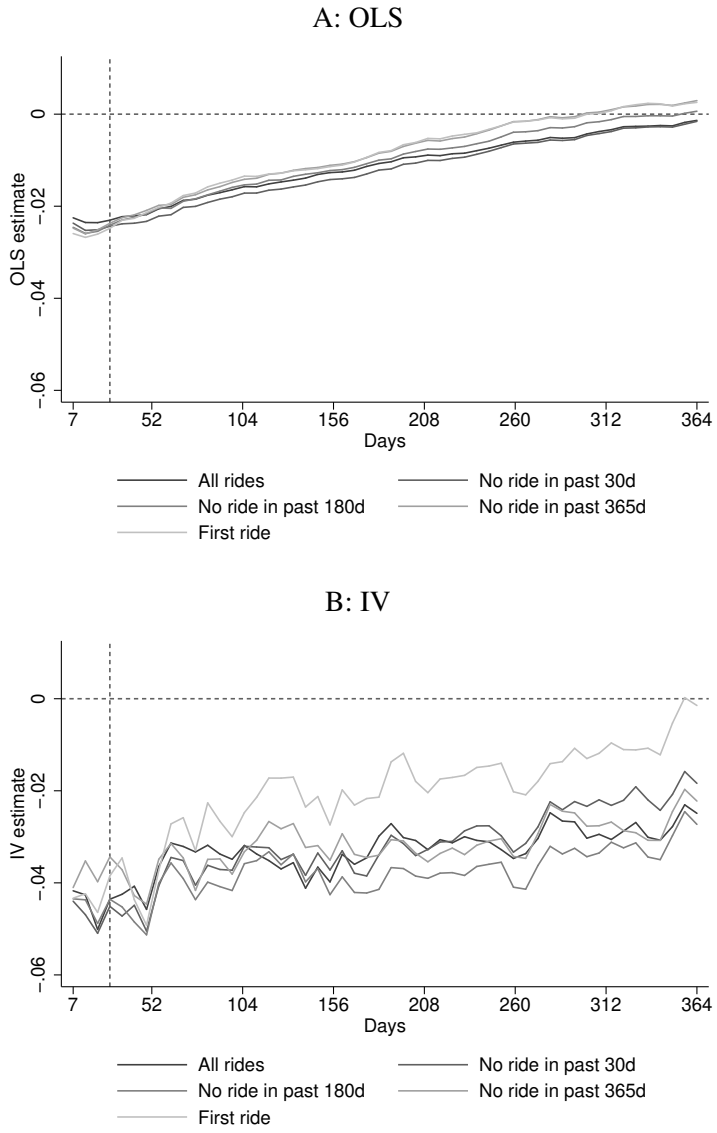
Note: This figure shows the visual IV plot corresponding to our baseline IV regression of the effect of the VA on 28-day mortality. For each bin of the instrument, which is the ambulance leave-out propensity to arrive at a VA hospital, we plot the mean 28-day mortality on the *y*-axis and the probability that the index patient arrives at a VA hospital on the *x*-axis. VA arrival predictions correspond to a first-stage regression in Equation (3), and mortality predictions correspond to a reduced-form regression in Equation (4). The best-fit line in the visual IV plot replicates the IV estimate of the effect of the VA on 28-day mortality, which we perform to obtain the standard error (in parentheses). This IV regression uses 400,769 observations and 1,267 combinations of ambulance company identifiers and Dartmouth Atlas Hospital Referral Regions (HRRs). The baseline sample selection is given in Appendix Table A.1. Controls include patient zip code dummies, ALS/BLS dummies, source of the ambulance ride, time categories, and patient prior utilization, which are detailed in Appendix Table A.2.

Figure A.2: Combinations of Controls



Note: This figure shows IV estimates of the VA effect on 28-day mortality on the y-axis, from Equation (2), varying the number of controls included in the IV regression. Numbered incremental controls correspond to categories or subcategories of variables that are presented in order in Appendix Tables A.2 and A.3. All specifications include the five baseline controls. Specifications with fewer than 10 controls do not include any leave-out (co-rider) controls. Specifications with ten or more controls include the five baseline controls and index patient controls. Therefore, the figure represents $5 + \left(2^5 - 1\right) + \left(2^6 - 1\right) = 99$ specifications. For each number of controls n for $n \in [5, 10]$, we consider "5 choose $n-5$" specifications. For each $n \geq 10$, we consider "6 choose $n-10$" specifications. The mean IV estimate is shown with a dashed line; the minimum and maximum IV estimates are shown with a short dashed line. We use our baseline sample, described in Appendix Table A.1.

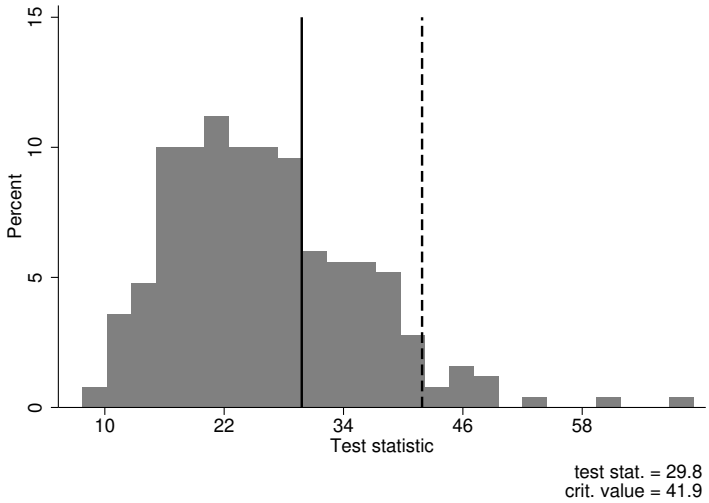
Figure A.3: Treatment Effects by Time and Sample



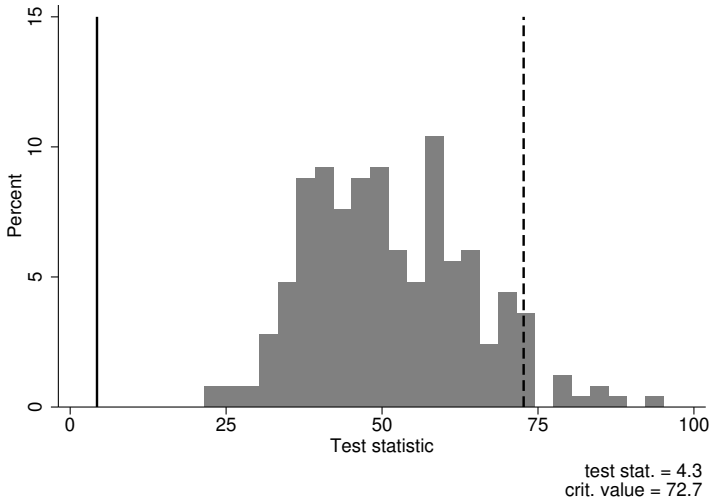
Note: This figure shows mortality treatment effects over varying days since the ambulance ride and in varying samples restricting by prior rides. "Day 0" considers mortality on the day of the ambulance ride; subsequent days indicate survival at one-week intervals from the ambulance ride. Panel A shows OLS results corresponding to Equation (6). Panel B shows IV results corresponding to Equation (5). The vertical dashed line indicates treatment effects on 28-day mortality, our baseline outcome. All regressions use a sample of ambulance rides with no prior ride in the last year.

Figure A.4: Joint Inequality Constraints

A: No Mortality Displacement

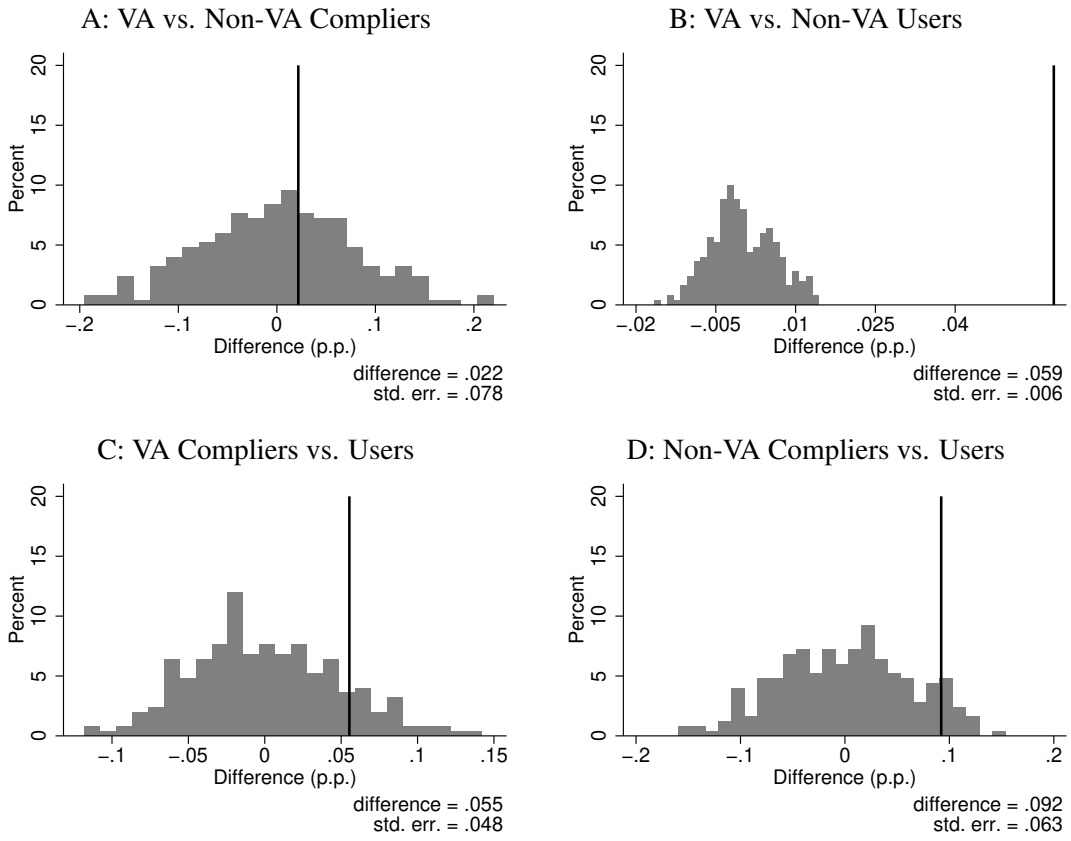


B: No Negative Hazard Rate



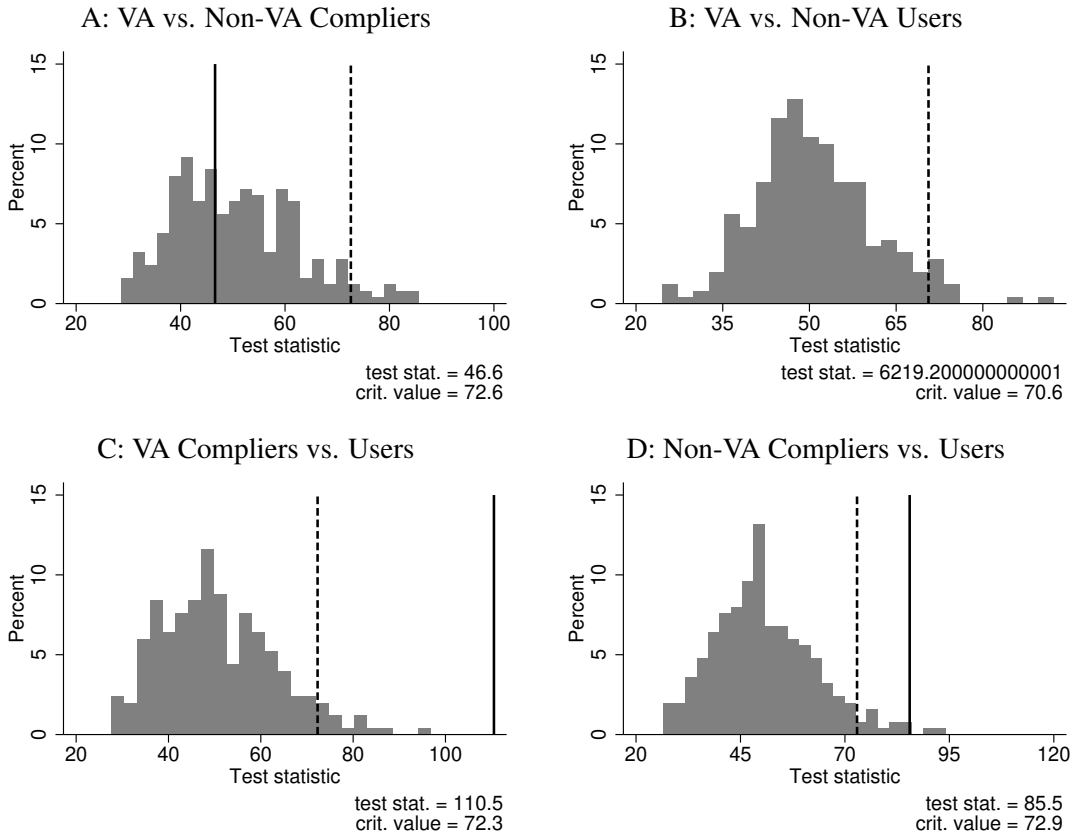
Note: This figure shows the test statistic for joint inequality constraints and bootstrapped-generated distributions of the test statistic under the least favorable version of the null hypothesis. Panel A shows the joint inequality test of no mortality displacement, as defined by null hypothesis in Equation (A.7). Panel B shows the joint inequality test of no negative hazard rates, as defined by the null hypothesis in Equation (A.16). The test statistic for both tests is shown as a solid vertical line. The one-sided critical value, or 95th percentile of the bootstrapped distribution of the test statistic under the least favorable version of the null hypothesis, is shown as a dashed vertical line. Details of the test statistic and the bootstrap procedure for Panels A and B are given in Appendices A.2.2 and A.2.3, respectively.

Figure A.5: Mean Hazard Differences



Note: This figure shows tests of equality of mean hazard rates for different sets of hazard rates, as defined by the null hypothesis in Equation (A.18). Each panel corresponds to a comparison between sets of hazards corresponding to VA or non-VA compliers or users. Details of the statistical procedure are given in Appendix A.2.4. Hazard rates for compliers are estimated by two-stage least squares and denoted in the appendix by $\hat{h}_{IV}(t;d)$, where $d = 1$ for compliers assigned to the VA and $d = 0$ for compliers assigned to a non-VA hospital. Hazard rates for users are estimated by OLS and denoted by $\hat{h}_{OLS}(t;d)$, where d similarly denotes VA users ($d = 1$) vs. non-VA users ($d = 0$). The solid black line shows the test statistic, and the histogram shows the distribution of bootstrapped test statistics under the null hypothesis. Bootstrapped standard errors are given in the caption.

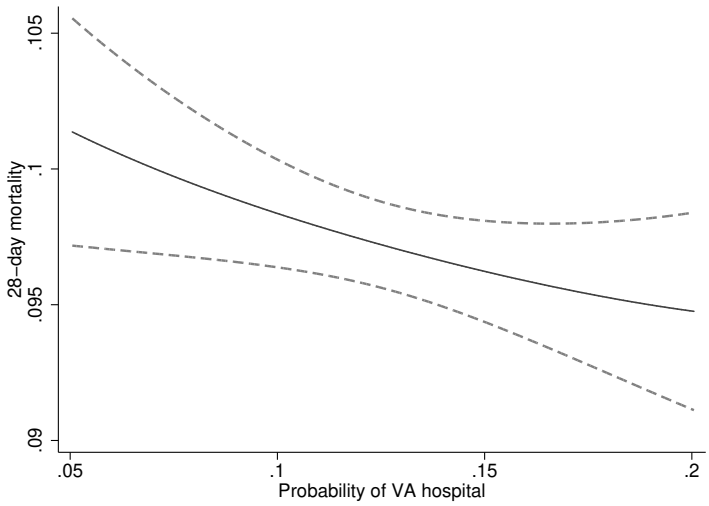
Figure A.6: Joint Equality Constraints



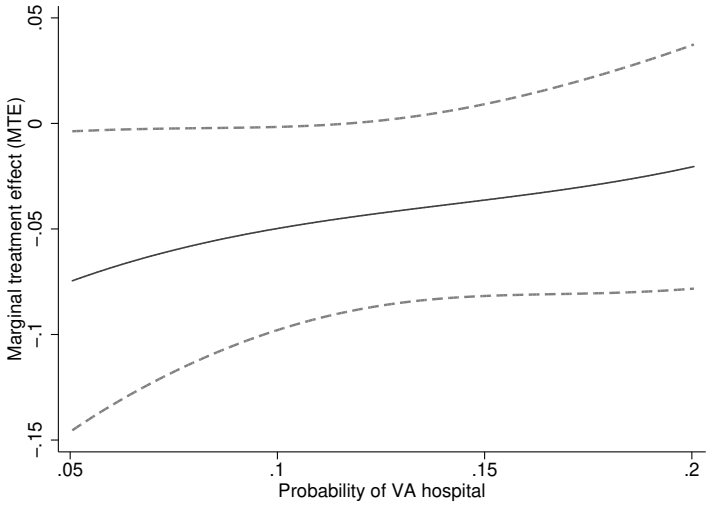
Note: This figure shows tests of joint equality of hazard rates for different sets of hazard rates, as defined by the null hypothesis in Equation (A.19). Each panel corresponds to a comparison between sets of hazards corresponding to VA or non-VA compliers or users. Details of the statistical procedure are given in Appendix A.2.4. Hazard rates for compliers are estimated by two-stage least squares and denoted in the appendix by $\hat{h}_{IV}(t;d)$, where $d = 1$ for compliers assigned to the VA and $d = 0$ for compliers assigned to a non-VA hospital. Hazard rates for users are estimated by OLS and denoted by $\hat{h}_{OLS}(t;d)$, where d similarly denotes VA users ($d = 1$) vs. non-VA users ($d = 0$). The solid line shows the test statistic, the histogram shows the distribution of bootstrapped test statistics under the null hypothesis, and the dashed line shows the one-sided 95th percentile critical value.

Figure A.7: Marginal Treatment Effects

A: Flexible Visual IV

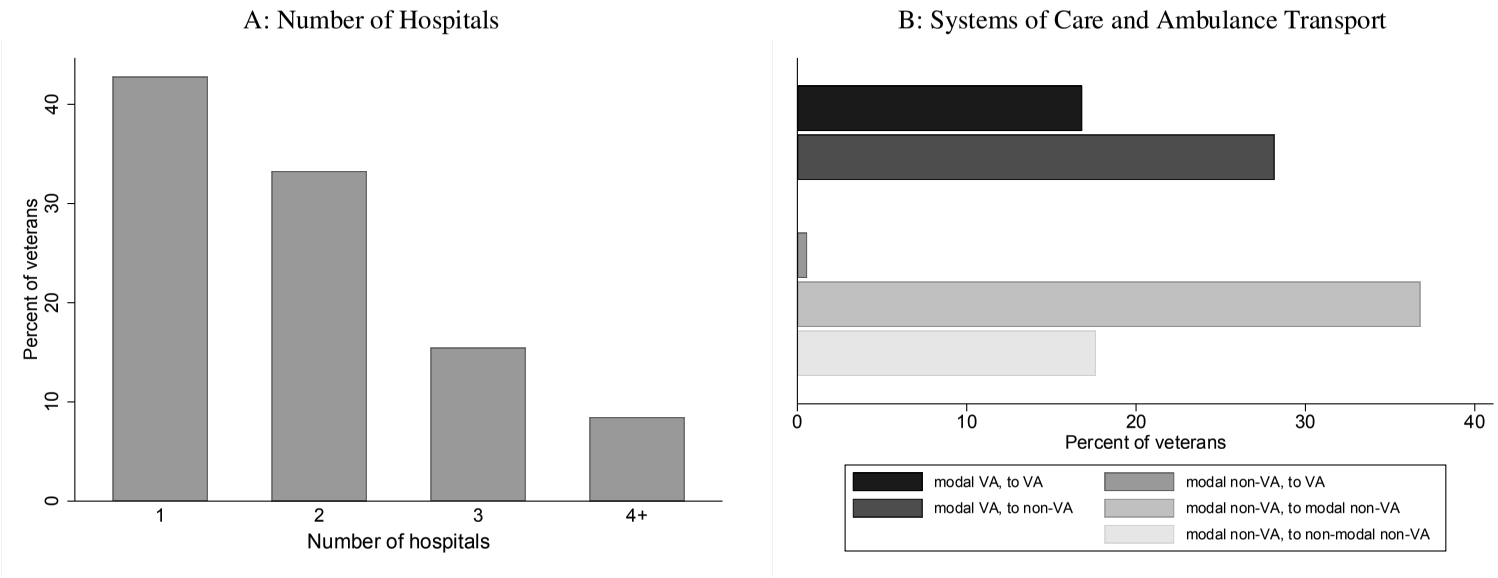


B: Marginal Treatment Effects



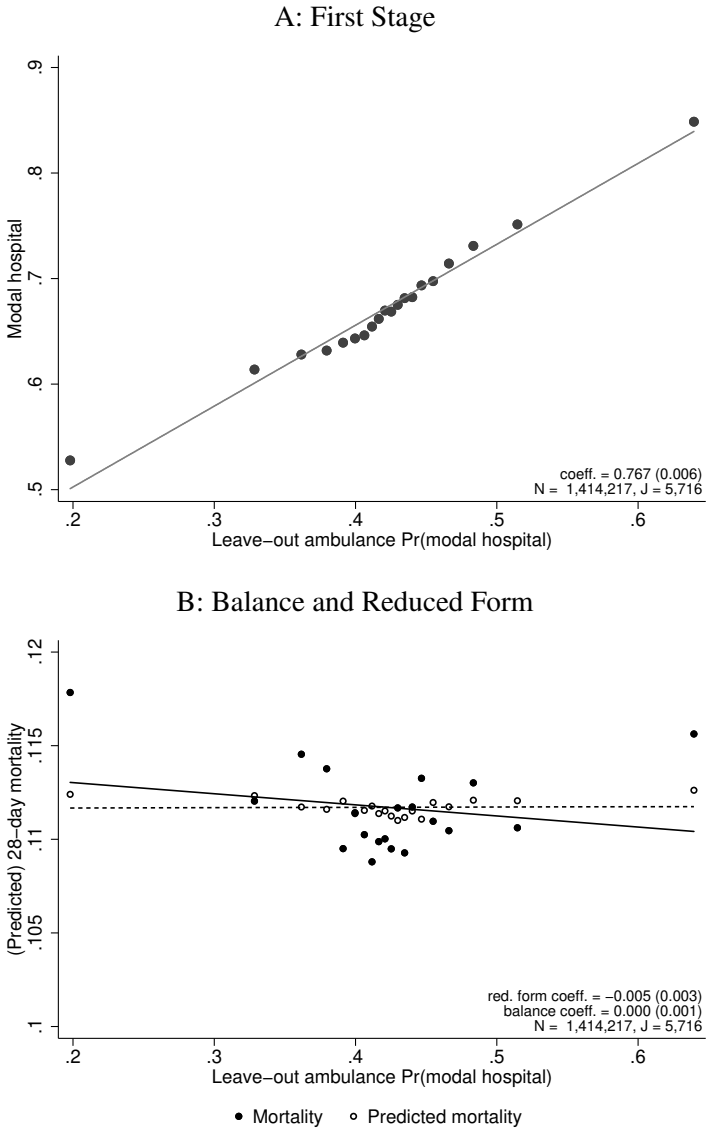
Note: This figure shows a flexible fit of the IV relationship between 28-day mortality and the ambulance propensity to transport to a VA hospital. Panel A shows the visual IV relationship with residual 28-day mortality on the *y*-axis and residual probability of being transported to a VA hospital on the *x*-axis. Both objects are residualized by baseline controls, described in Appendix Table A.2. The probability of being transported to a VA hospital is calculated from the first-stage relationship in Equation (3). The data underlying the fit in Panel A are similar to those in the linear visual IV plot in Appendix Figure A.1. The fit is based on five Gaussian basis splines. Panel B shows the implied marginal treatment effects, which are the analytical derivatives at each point on the fit in Panel A. 95% confidence intervals are calculated by 50 bootstrapped interations (drawn by zip codes, with replacement). Details are given in Appendix A.4.

Figure A.8: Sources of Prior Utilization



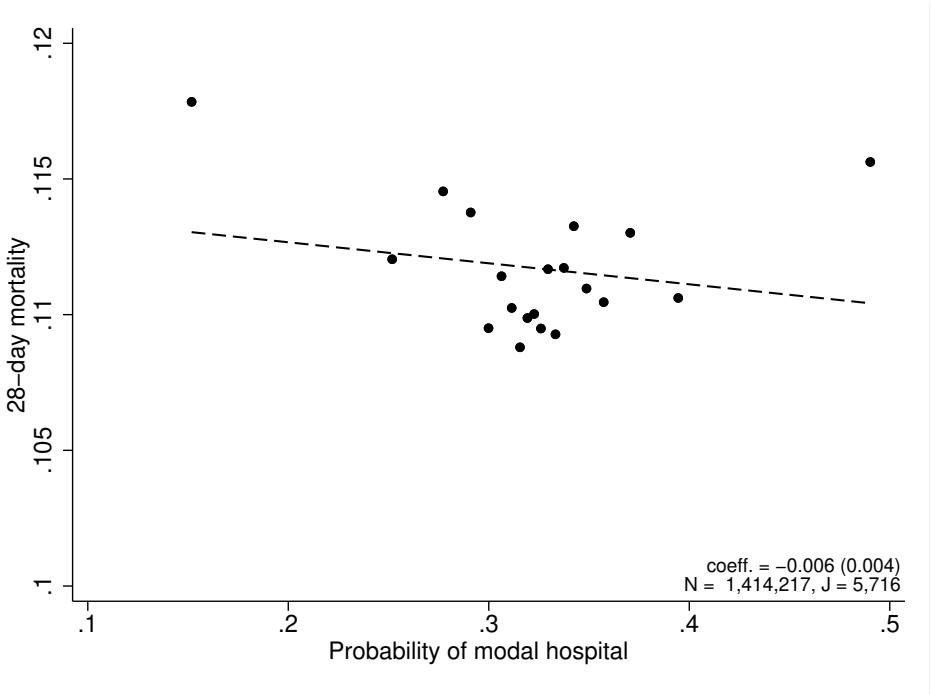
Note: This figure shows patterns of prior utilization and ambulance transport among a sample of patients who have some prior utilization either at the VA or affiliated with a non-VA hospital. Panel A shows the percentage of patients in this sample who utilize care associated with different numbers of hospitals. Panel B shows ambulance transport patterns to either the VA or a non-VA hospital depending on whether a patient's modal hospital in prior utilization was associated with the VA or with a non-VA hospital. If the patient's modal hospital utilization was at a non-VA hospital, the figure also shows the percentage of patients who were transported to their modal non-VA hospital or another non-VA hospital. The sample selection for this group of patients is given in Appendix Table A.13.

Figure A.9: Modal Hospital First Stage, Balance, and Reduced Form



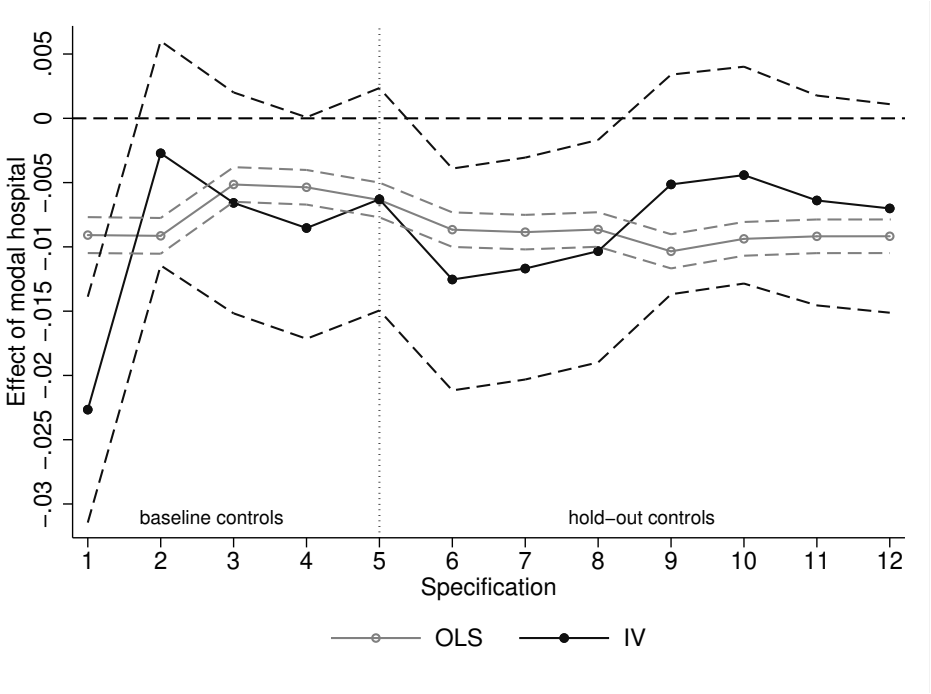
Note: Panel A shows a binned scatterplot of arrival at the veteran's modal hospital against the ambulance leave-out propensity to arrive at that hospital on the *x*-axis. The figure is a graphical representation of the first-stage regression in Equation (10). Panel B shows binned scatterplots of 28-day mortality and predicted 28-day mortality on the *y*-axis against the ambulance leave-out propensity to arrive at the veteran's modal hospital on the *x*-axis. Mortality bin means are shown in solid circles, while predicted mortality bin means are shown in hollow circles. The figure represents the reduced-form regression in Equation (11) and the corresponding balance regression replacing mortality with predicted mortality. The sample includes 1,421,612 ambulance rides and 5,923 combinations of ambulance company identifiers and Dartmouth Atlas Hospital Referral Regions (HRRs). The sample includes patients who have some utilization affiliated with a non-VA hospital and no utilization at the VA in the prior year. The selection details of this sample is given in Appendix Table A.13. Controls include patient zip code dummies, ALS/BLS dummies, source of the ambulance ride, time categories, and patient prior utilization, which are detailed in Appendix Table A.2.

Figure A.10: Modal Hospital Visual IV



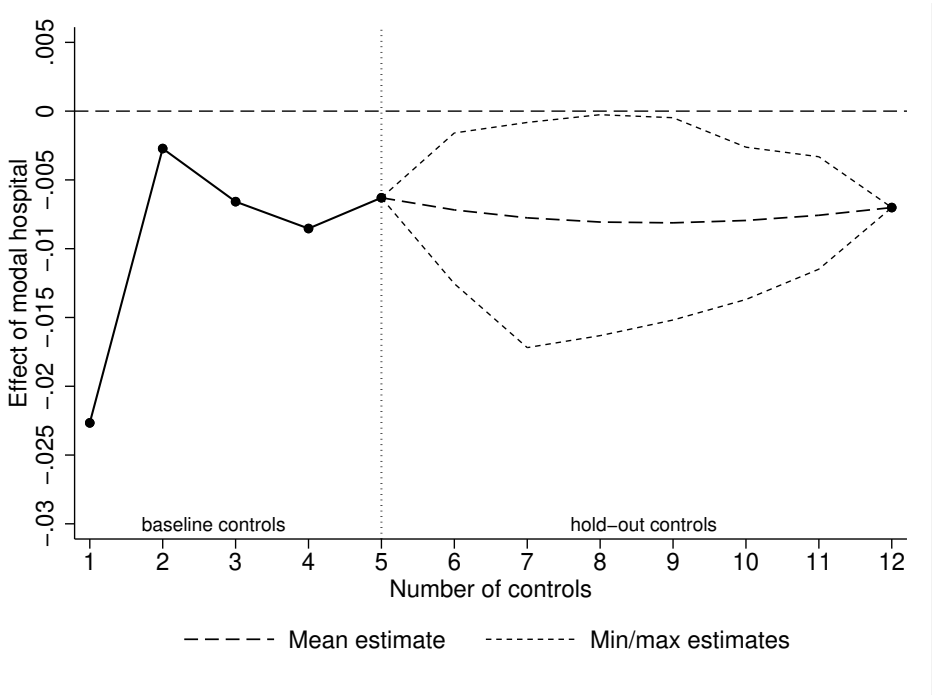
Note: This figure shows the visual IV plot corresponding to the IV regression of the effect of arrival at a patient's modal hospital on 28-day mortality. For each bin of the instrument, which is the ambulance leave-out propensity to arrive at the patient's modal hospital, we plot the mean 28-day mortality on the *y*-axis and the probability that the index patient arrives at his modal hospital on the *x*-axis. Modal hospital arrival predictions correspond to a first-stage regression in Equation (10), and mortality predictions correspond to a reduced-form regression in Equation (11). The best-fit line in the visual IV plot replicates the IV estimate of the effect of arrival at a patient's modal hospital on 28-day mortality, which we perform to obtain the standard error (in parentheses). This IV regression uses 1,421,612 observations and 5,923 combinations of ambulance company identifiers and Dartmouth Atlas Hospital Referral Regions (HRRs). We use the sample of non-VA-only utilizers, given in Appendix Table A.13. Controls include patient zip code dummies, ALS/BLS dummies, source of the ambulance ride, time categories, and patient prior utilization, which are detailed in Appendix Table A.2.

Figure A.11: Modal Hospital OLS and IV Specifications



Note: This figure shows the effect of arrival at a patient's modal hospital on 28-day mortality estimated from OLS and IV specifications, with progressive sets of controls. Numbered incremental controls correspond to categories or subcategories of variables that are presented in order in Appendix Tables A.2 and A.3. Estimates are shown along solid lines, while 95% confidence intervals are shown in dashed lines. All specification control for hospital identities and use the sample of non-VA-only utilizers, given in Appendix Table A.13.

Figure A.12: Modal Hospital Combinations of Controls



Note: This figure shows IV estimates of the effect of arrival at a patient's modal hospital on mortality on the y-axis, with first-stage and reduced-form Equations (10) and (11), varying the number of controls included in the IV regression. Control variables are detailed in Appendix Tables A.2 and A.3. All specifications include the five baseline controls. Specifications with fewer than 10 controls do not include any leave-out controls. Specifications with ten or more controls include the five baseline controls and index patient controls. Therefore, the figure represents $5 + \left(2^5 - 1\right) + \left(2^6 - 1\right) = 99$ specifications. For each number of controls n for $n \in [5, 10]$, we consider "5 choose $n-5$" specifications. For each $n \geq 10$, we consider "6 choose $n-10$" specifications. The mean IV estimate is shown with a dashed line; the minimum and maximum IV estimates are shown with a short dashed line. We use the sample of non-VA-only utilizers, given in Appendix Table A.13.

Table A.1: Selection of Baseline Sample

Sample step	Description	Rides	Patients	Ambulance companies	Hospitals	
					VA	Non-VA
1. Build initial sample of ambulance rides to EDs from January 1, 2001, to December 31, 2014.	Require ED visit within 24 hours after ambulance ride, non-missing demographic data, enrollment in Medicare Parts A and B for at least one year, date of death (if non-missing) weakly after the ambulance ride.	8,952,884	2,898,667	183,693	127	7,816
2. Clean sample	Drop rides linked to more than one ED visit (i.e., visits in different hospitals), with patients younger than 20 years or older than 99 years, with missing Health Referral Region, with missing ambulance diagnosis code, or from VA New Orleans (destroyed in 2005 due to Hurricane Katrina).	8,828,997	2,862,557	180,320	125	7,744
3. Distance restrictions	Drop patients who do not live within 20 miles of a VA hospital and within 20 miles of a non-VA hospital. Drop rides to a hospital over 50 miles from the patient's home.	3,465,588	1,118,302	14,662	118	3,071
4. Ambulance restrictions	Drop rides by an ambulance company with fewer than 20 patients in a given zip code. Drop rides by an ambulance company with less than 5% of rides in a given zip code to a VA hospital. Drop rides from zip codes with only one remaining ambulance company.	1,051,093	365,163	1,217	100	1,577
5. Prior utilization restriction	Drop rides for patients with no VA utilization (inpatient, ED, or primary care) in the prior year.	491,193	188,299	1,217	99	1,404
6. Prior ride restriction	Drop rides for patients with a prior ride within 30 days.	401,319	188,299	1,217	99	1,386

Note: This table details selection steps to create the baseline sample. At each step, the table lists the number of ambulance rides, patients, ambulance companies, and VA and non-VA hospitals. Table 1 shows average patient characteristics among observations at each sample step.

Table A.2: Baseline Control Variables

Category	Subcategory	Variables
Location (1,681 indicators)	Zip code (1,678 indicators)	Zip code indicators (1,678 indicators)
	Pickup source (3 indicators)	Indicators for whether pickup is from residence, residential (including domiciliary, custodial facility), skilled nursing facility, or scene of accident (omitted category)
Ambulance service (3 indicators)		Indicators for whether ambulance is ALS special (CPT codes A0427, A0330, A0370), ALS non-special (CPT codes Q3019, A0368, A0328), ALS level 2 (CPT code A0433), or BLS (omitted category; CPT codes A0429, A0362, A0322)
Time categories (182 indicators)		Day of the week (6 indicators) Month-year interactions (176 indicators)
Prior utilization (6 indicators)		Indicators for utilization in prior year of Medicare primary care, VA primary care utilization, Medicare ED, VA ED, Medicare inpatient, and VA inpatient services

Note: This table describes baseline controls variables, denoted as $\left(z\left(i\right), \mathbf{X}_i^0\right)$ in Condition 1 and throughout the text. We consider our quasi-experiment to be conditional on these variables, and we include these variables as controls in all of our analyses.

Table A.3: Hold-Out Control Variables

Category	Subcategory	Variables
Patient background (60 variables)	Demographics (31 indicators)	Age: 5-year age bins from 20-64 years, 2-year age bins from 65-100 years (27 indicators) Male gender Race: indicators for white, Black, Hispanic, and Asian/other (omitted category)
	Socioeconomic status, combat history, and eligibility (21 indicators)	Terciles of income and net worth (4 indicators) Period of combat: WWII, Korean, Vietnam, other (omitted category) (3 indicators) Indicator for aid and attendance for in-home care Priority group indicators (6 indicators) Service connection: service connected, not service connected, or non-veteran/other (omitted category) (2 indicators) 6 missing indicators for each of the above characteristics
	Extended prior utilization (8 variables)	Counts of VA primary care visits, outpatient visits, ED visits, and inpatient visits in prior year Analogous counts of Medicare visits in prior year
Prior diagnoses (93 indicators)		31 Elixhauser indicators (dividing hypertension indicator into 2 indicators for complicated and uncomplicated hypertension), in four categories: present in VA data only, present in Medicare data only, and present in both VA and Medicare data ($31 \times 3 = 93$ indicators)
3-digit ambulance diagnosis codes (778 indicators)		3-digit ambulance diagnosis (ICD-9) codes (778 indicators)
Co-rider characteristics (33 variables)	Co-rider baseline controls (12 variables)	Co-rider pickup source proportions (3 variables) Co-rider ambulance service proportions (3 variables) Co-rider prior utilization proportions (6 variables)
	Co-rider hold-out controls (21 variables)	Co-rider average continuous age Co-rider proportion male gender Co-rider race proportions (3 variables) Co-rider 1-digit ambulance code proportions (15 variables) Co-rider average predicted mortality

Note: This table describes hold-out control variables. These variables are used to test robustness of our findings, particularly in Figures 2, A.2, A.12, and A.11.

Table A.4: Robustness of Exclusion Restriction

	(1)	(2)	(3)	(4)	(5)	(6)
	A: Dependent variable: 28-day mortality					
VA hospital	-0.053	-0.042	-0.045	-0.045	-0.045	-0.049
	(0.019)	(0.018)	(0.018)	(0.018)	(0.018)	(0.020)
Outcome mean	0.097	0.097	0.097	0.097	0.097	0.097
Observations	401,319	401,319	401,319	401,319	401,319	401,319
	B: Dependent variable: 28-day spending					
VA hospital	-2,245	-2,583	-2,748	-2,501	-2,569	-2,559
	(899)	(832)	(880)	(816)	(822)	(992)
Outcome mean	12,173	12,173	12,173	12,173	12,173	12,173
Observations	401,319	401,319	401,319	401,319	401,319	401,319
Ambulance charges splines	Yes	No	No	No	No	Yes
Non-VA hospitals in chosen set	No	Yes	No	No	No	Yes
Mileage splines	No	No	Yes	No	No	Yes
Out-of-sample non-VA mortality	No	No	No	Yes	No	Yes
Out-of-sample non-VA spending	No	No	No	No	Yes	Yes

Note: This table presents IV estimates of the effect of the VA on 28-day mortality (Panel A) and on 28-day spending (Panel B). In each panel, each column involves including a set of controls for ambulance actions on the specific ride (flexible functions of the charges incurred by the ambulance company, flexible functions of the mileage driven by the ambulance company) and for non-VA hospitals chosen by the ambulance company (non-VA hospitals chosen by the ambulance company in rides originating from the same zip code and "out-of-sample" averages of non-VA hospital mortality and spending). "Out-of-sample" refers to averages that are computed using patients outside of the main analytical sample (Appendix Table A.1) because they have no VA utilization in the prior year; specifically, they are computed using patients with only non-VA utilization in the prior year (Panel B of Appendix Table A.13). Regressions are run on the main analytical sample. Further details are given in Appendix A.1.1.

Table A.5: Monotonicity Tests

First stage sample	Observations	VA share	Instrument: Baseline	Instrument: Reverse-sample
Age ≤ 80	239,611	0.347	0.931	0.508
			(0.038)	(0.022)
Age > 80	161,707	0.305	0.789	0.464
			(0.041)	(0.022)
White	314,064	0.304	0.821	0.221
			(0.037)	(0.016)
Non-white	87,176	0.426	0.992	0.596
			(0.068)	(0.041)
Comorbidity count (high)	185,477	0.295	0.760	0.442
			(0.038)	(0.019)
Comorbidity count (low)	215,842	0.360	0.948	0.648
			(0.041)	(0.029)
Mental illness or substance abuse	188,961	0.354	0.931	0.543
			(0.040)	(0.027)
No mental illness or substance abuse	212,358	0.309	0.815	0.490
			(0.037)	(0.022)
VA visits in prior year (high)	183,087	0.508	1.038	0.790
			(0.050)	(0.040)
VA visits in prior year (low)	218,232	0.181	0.718	0.296
			(0.031)	(0.015)
Advanced Life Support	274,690	0.301	0.836	0.279
			(0.036)	(0.024)
No Advanced Life Support	126,616	0.393	0.840	0.209
			(0.048)	(0.047)
Predicted VA user (high)	200,659	0.543	1.113	0.952
			(0.054)	(0.058)
Predicted VA user (low)	200,660	0.117	0.559	0.220
			(0.030)	(0.012)
Predicted mortality (high)	200,659	0.328	0.835	0.393
			(0.036)	(0.022)
Predicted mortality (low)	200,660	0.333	0.898	0.557
			(0.046)	(0.026)
Instrument sample			Dual eligibles	Analytical sample

Note: This table presents first-stage coefficients on different subsamples of patients. For each subsample, we present results for two different instruments: (i) the baseline leave-out instrument, Z_i, given in Equation (1) and calculated from observations among dually eligible veterans (Step 1 of Appendix Table A.1), and (ii) a reverse-sample instrument, $\tilde{Z}_i^{-m}$, given in Equation (A.2) and calculated from observations in the analytical sample (Step 6 of Appendix Table A.1) that are outside of the regression subsample. Each regression uses baseline controls defined in Appendix Table A.2. Further details are given in Appendix A.1.2.

Table A.6: Monotonicity Tests (Continued)

First stage sample	Observations	VA share	Instrument	
			Baseline	In-sample
Age ≤ 80	239,611	0.347	0.586	0.513
			(0.021)	(0.019)
Age > 80	161,707	0.305	0.494	0.376
			(0.023)	(0.021)
White	314,064	0.304	0.504	0.513
			(0.019)	(0.020)
Non-white	87,176	0.426	0.676	0.440
			(0.032)	(0.033)
Comorbidity count (high)	185,477	0.295	0.490	0.361
			(0.019)	(0.018)
Comorbidity count (low)	215,842	0.360	0.592	0.456
			(0.022)	(0.018)
Mental illness or substance abuse	188,961	0.354	0.592	0.458
			(0.021)	(0.019)
No mental illness or substance abuse	212,358	0.309	0.501	0.386
			(0.020)	(0.018)
VA visits in prior year (high)	183,087	0.508	0.691	0.493
			(0.026)	(0.019)
VA visits in prior year (low)	218,232	0.181	0.421	0.382
			(0.018)	(0.018)
Advanced Life Support	274,690	0.301	0.523	0.436
			(0.020)	(0.018)
No Advanced Life Support	126,616	0.393	0.531	0.286
			(0.025)	(0.019)
Predicted VA user (high)	200,659	0.543	0.743	0.566
			(0.028)	(0.020)
Predicted VA user (low)	200,660	0.117	0.331	0.352
			(0.016)	(0.027)
Predicted mortality (high)	200,659	0.328	0.513	0.386
			(0.020)	(0.017)
Predicted mortality (low)	200,660	0.333	0.570	0.410
			(0.023)	(0.019)
Instrument sample			Analytical sample	Analytical sample

Note: This table presents first-stage coefficients on different subsamples of patients. For each subsample, we present results for two different instruments: (i) the baseline leave-out instrument, $\tilde{Z}_i$, given in Equation (1), and (ii) a in-sample instrument, $\tilde{Z}_i^m$, given in Equation (A.2) and calculated from leave-out observations in the same regression subsample. Both instruments are calculated using observations in the analytical sample (Step 6 of Appendix Table A.1). Each regression uses baseline controls defined in Appendix Table A.2. Further details are given in Appendix A.1.2.

Table A.7: Always-Taker and Never-Taker Characteristics

	Always takers		Never takers	
	Mean	Ratio	Mean	Ratio
Male	0.961	1.00	0.965	1.00
	(0.002)	[0.99 - 1.00]	(0.001)	[1.00 - 1.00]
Age	75.6	0.99	76.3	1.00
	(0.158)	[0.99 - 1.00]	(0.153)	[1.00 - 1.01]
Black	0.222	1.14	0.184	0.95
	(0.012)	[1.02 - 1.26]	(0.010)	[0.85 - 1.05]
Income	$18,039	0.86	$22,397	1.07
	($200)	[0.84 - 0.88]	($232)	[1.05 - 1.09]
Rural zip code	0.064	1.27	0.053	1.04
	(0.015)	[0.67 - 1.87]	(0.011)	[0.62 - 1.46]
Residential source	0.685	0.97	0.667	0.95
	(0.011)	[0.94 - 1.00]	(0.009)	[0.92 - 0.97]
Comorbidity count	5.85	0.95	6.44	1.05
	(0.046)	[0.94 - 0.97]	(0.032)	[1.04 - 1.06]
Mental illness	0.469	1.10	0.420	0.98
	(0.006)	[1.07 - 1.13]	(0.004)	[0.97 - 1.00]
Substance abuse	0.150	1.04	0.137	0.95
	(0.005)	[0.97 - 1.11]	(0.004)	[0.90 - 1.00]
Prior VA ED visit	0.823	1.56	0.383	0.72
	(0.004)	[1.54 - 1.57]	(0.006)	[0.70 - 0.75]
Prior Medicare ED visit	0.262	0.54	0.613	1.27
	(0.006)	[0.52 - 0.57]	(0.004)	[1.26 - 1.29]
Ambulance rides in prior year	2.212	1.03	2.210	1.03
	(0.030)	[1.00 - 1.05]	(0.025)	[1.00 - 1.05]
Advanced Life Support	0.576	0.84	0.707	1.03
	(0.013)	[0.81 - 0.88]	(0.010)	[1.01 - 1.06]
Predicted VA user	0.969	1.14	0.778	0.92
	(0.001)	[1.14 - 1.15]	(0.002)	[0.91 - 0.92]
Predicted mortality	0.096	0.99	0.103	1.07
	(0.002)	[0.96 - 1.03]	(0.001)	[1.05 - 1.08]

Note: This table presents average characteristics for always takers and never takers. Always takers are defined as patients who present to the VA even when they receive a residualized instrument below the 20th percentile; never takers are defined as patients who present to a non-VA hospital even when they receive a residualized instrument above the 80th percentile. To form these residualized instruments, we residualize the baseline instrument, Z_i, given in Equation (1), by baseline controls, described in Appendix Table A.2. Observations are drawn from the baseline sample described in Appendix Table A.1. For each row corresponding to a characteristic, the table presents average characteristics and the ratio between this average and the overall sample average. Overall sample means are given in Table 4. Standard errors are calculated by bootstrap, blocking observations by zip codes, and are shown in parentheses. Corresponding 95% confidence intervals of the ratio are presented in brackets. Further details are given in Appendix A.3.

Table A.8: Treatment Effects from Selection Model

	Dependent variable: 28-day mortality				
	(1)	(2)	(3)	(4)	(5)
ATE	-0.043	-0.033	-0.033	-0.033	-0.033
	(0.017)	(0.006)	(0.006)	(0.005)	(0.005)
ATE-LATE difference	0.003	0.013	0.012	0.012	0.012
	(0.001)	(0.011)	(0.011)	(0.011)	(0.011)
Control function	Linear	Cubic	Cubic	Gaussian basis	Gaussian basis
Knots		3	5	3	5
Outcome mean	0.097	0.097	0.097	0.097	0.097
Observations	401,319	401,319	401,319	401,319	401,319

Note: This table presents estimates of the average treatment effect (ATE) from the selection model in Equation (A.21), under different specifications. Column 1 presents results from a control function that is linear in the first-stage residual, corresponding to the regression in Equation (A.23). Columns 2 to 5 present results from semiparametric control functions, corresponding to regressions of the form in Equation (A.25). The columns vary in whether the spline functions are cubic functions or Gaussian basis functions and in the number of knots. In addition to the ATE, each column presents the difference between the ATE and the local average treatment effect (LATE). The LATE is estimated from Equation (A.24) and is numerically equivalent to the LATE from our benchmark analysis in Section 3. We compute standard errors (shown in parentheses) for the ATE and the ATE-LATE difference by bootstrap, blocking by zip codes. Appendix A.4 provides further details.

Table A.9: Heterogeneity by Non-VA Hospital Characteristics

	Regression estimates		Characteristic means	
	VA	VA $\times \tilde{I}_{x,i}$	$I_{x,i} = 0$	$I_{x,i} = 1$
Volume, Size, and Capabilities				
ED visits	-0.045	0.000	39,387	85,592
	(0.017)	(0.002)		
Admissions	-0.045	-0.000	13,504	29,745
	(0.016)	(0.002)		
Total staffed beds	-0.045	-0.000	276	626
	(0.017)	(0.002)		
Teaching hospital	-0.045	-0.000	0.02	0.51
	(0.017)	(0.002)		
Trauma center	-0.045	0.004	0.28	0.93
	(0.016)	(0.002)		
Advanced cardiac care	-0.046	-0.000	0.64	1.00
	(0.017)	(0.002)		
Stroke center	-0.045	0.001	0.03	0.65
	(0.017)	(0.002)		
Staffing				
ED staff per 100 ED visits	-0.045	-0.001	0.03	0.07
	(0.017)	(0.002)		
Nurses per 100 admissions	-0.045	0.003	2.18	3.41
	(0.017)	(0.002)		
Physicians per 100 admissions	-0.045	-0.000	0.04	0.34
	(0.016)	(0.002)		
Hospitalists per 100 admissions	-0.045	0.004	0.06	0.19
	(0.017)	(0.002)		
Intensivists per 100 admissions	-0.045	0.002	0.03	0.12
	(0.017)	(0.002)		

Note: This table presents regression results investigating heterogeneous treatment effects along binary indicators of average non-VA hospital characteristics associated with each zip code. For each zip code, hospital characteristics are averaged with weights proportional to the number of rides going to each non-VA hospital from the zip code. We then divide observations i, based on whether their zip codes $z(i)$ have below- vs. above-median averages, denoted by $I_{x,i} = 0$ and $I_{x,i} = 1$, respectively. Regression results correspond to Equation (8). The coefficient on VA represents the LATE of going to the VA, and the coefficient on VA $\times \tilde{I}_{x,i}$ represents the difference in the LATE between observations with $I_{x,i} = 1$ and observations with $I_{x,i} = 0$. Appendix Table A.10 presents results for additional characteristics. Appendix A.5 provides further details on the hospital characteristics.

Table A.10: Heterogeneity by Non-VA Hospital Characteristics (Continued)

	Regression estimates		Characteristic means	
	VA	VA $\times \tilde{I}_{x,i}$	$I_{x,i} = 0$	$I_{x,i} = 1$
Spending and Outcomes				
Relative spending	-0.045	-0.002	0.97	1.04
	(0.017)	(0.002)		
Mortality rate	-0.045	-0.003	11.62	12.89
	(0.017)	(0.002)		
Readmission rate	-0.045	-0.002	17.30	18.90
	(0.017)	(0.002)		
Organization and IT				
Network or hospital system	-0.045	-0.002	0.65	1.00
	(0.017)	(0.002)		
HMO or ACO	-0.045	-0.002	0.00	0.47
	(0.017)	(0.002)		
Health IT	-0.046	-0.002	0.00	0.80
	(0.016)	(0.002)		
Largest non-VA $\geq$ 80%	-0.045	0.006	0.52	0.90
	(0.016)	(0.003)		

Note: This table presents regression results investigating heterogeneous treatment effects along binary indicators of average non-VA hospital characteristics associated with each zip code. For each zip code, hospital characteristics are averaged with weights proportional to the number of rides going to each non-VA hospital from the zip code. We then divide observations i, based on whether their zip codes $z(i)$ have below- vs. above-median averages, denoted by $I_{x,i} = 0$ and $I_{x,i} = 1$, respectively. Regression results correspond to Equation (8). The coefficient on VA represents the LATE of going to the VA, and the coefficient on VA $\times \tilde{I}_{x,i}$ represents the difference in the LATE between observations with $I_{x,i} = 1$ and observations with $I_{x,i} = 0$. Appendix Table A.9 presents results for additional characteristics. Appendix A.5 provides further details on the hospital characteristics.

Table A.11: Heterogeneity by VA Hospital Characteristics

	Regression estimates		Characteristic means	
	VA	VA $\times \tilde{I}_{x,i}$	$I_{x,i} = 0$	$I_{x,i} = 1$
Volume, Size, and Capabilities				
ED visits	-0.045	-0.001	8,626	23,107
	(0.017)	(0.002)		
Admissions	-0.044	-0.003	3,247	8,146
	(0.016)	(0.002)		
Total staffed beds	-0.044	-0.007	139	463
	(0.017)	(0.002)		
Teaching hospital	-0.045	-0.003	0.00	0.93
	(0.017)	(0.002)		
Trauma center	-0.052	0.006	0.00	1.00
	(0.018)	(0.004)		
Advanced cardiac care	-0.051	-0.004	0.00	1.00
	(0.018)	(0.002)		
Staffing				
ED staff per 100 ED visits	-0.050	-0.002	0.02	0.12
	(0.022)	(0.003)		
Nurses per 100 admissions	-0.045	-0.000	5.26	16.77
	(0.017)	(0.002)		
Physicians per 100 admissions	-0.045	-0.001	1.49	5.39
	(0.017)	(0.002)		
Hospitalists per 100 admissions	-0.049	0.005	0.04	0.30
	(0.022)	(0.003)		
Intensivists per 100 admissions	-0.050	0.001	0.00	0.14
	(0.022)	(0.003)		
Spending and Outcomes				
Relative spending	-0.045	-0.002	0.95	1.22
	(0.016)	(0.002)		
Mortality rate	-0.045	0.005	7.11	7.98
	(0.017)	(0.003)		
Readmission rate	-0.045	-0.003	11.70	12.70
	(0.017)	(0.002)		

Note: This table presents regression results investigating heterogeneous treatment effects along characteristics of the VA hospital associated with each zip code. For each VA hospital characteristic x, we divide observations i, based on whether x is below vs. above the median, denoted by $I_{x,i} = 0$ and $I_{x,i} = 1$, respectively. Regression results correspond to Equation (8). The coefficient on VA represents the LATE of going to the VA, and the coefficient on VA $\times \tilde{I}_{x,i}$ represents the difference in the LATE between observations with $I_{x,i} = 1$ and observations with $I_{x,i} = 0$. Appendix A.5 provides further details on the hospital characteristics.

Table A.12: Heterogeneity by Patient Characteristics

	Regression estimates		Characteristic means	
	VA	VA $\times \tilde{I}_{x,i}$	$I_{x,i} = 0$	$I_{x,i} = 1$
Older than 80	-0.047	0.004	0.00	1.00
	(0.017)	(0.003)		
Black	-0.043	-0.002	0.00	1.00
	(0.017)	(0.003)		
Hispanic	-0.045	-0.008	0.00	1.00
	(0.017)	(0.008)		
Income	-0.048	0.000	$12,005	$33,760
	(0.018)	(0.000)		
Comorbidity count	-0.044	-0.014	3.90	9.28
	(0.016)	(0.002)		
Mental illness or substance abuse	-0.045	-0.005	0.00	1.00
	(0.017)	(0.002)		
VA visits in prior year	-0.044	-0.004	2.15	11.88
	(0.017)	(0.002)		
Ambulance rides in prior year	-0.043	-0.008	1.00	3.55
	(0.017)	(0.002)		
Advanced Life Support	-0.046	-0.013	0.00	1.00
	(0.017)	(0.002)		
Predicted VA user	-0.044	-0.005	0.70	1.00
	(0.017)	(0.003)		
Predicted mortality	-0.045	-0.018	0.04	0.15
	(0.016)	(0.002)		

Note: This table presents regression results investigating heterogeneous treatment effects along patient characteristics. For each VA hospital characteristic x, we divide observations i, based on whether x is below vs. above the median, denoted by $I_{x,i} = 0$ and $I_{x,i} = 1$, respectively. Regression results correspond to Equation (8). The coefficient on VA represents the LATE of going to the VA, and the coefficient on VA $\times \tilde{I}_{x,i}$ represents the difference in the LATE between observations with $I_{x,i} = 1$ and observations with $I_{x,i} = 0$.

Table A.13: Selection of Alternative Samples

Sample step	Description	Rides	Patients	Ambulance companies	Hospitals	
					VA	Non-VA
A: Sample for Descriptive Utilization Patterns						
3. Start from distance restrictions in baseline sample	See step #3 in Table A.1.	1,051,093	365,163	1,217	100	1,577
4. Prior utilization restriction	Keep rides for patients with some non-VA or VA utilization (inpatient, ED, or primary care).	977,826	340,371	1,217	100	1,565
5. Prior ride restriction	Drop rides for patients with a prior ride within 30 days.	794,940	340,371	1,217	100	1,548
B: Non-VA-Only Sample						
2. Start from clean sample	See step #2 in Table A.1.	8,828,997	2,862,557	180,320	125	7,744
3. Distance restrictions	Drop rides to a hospital over 50 miles from the patient's home. Drop zip codes without at least two non-VA hospitals within 20 miles that receive at least 5% from that zip code.	6,424,120	2,131,152	29,100	122	5,498
4. Ambulance restrictions	Drop rides by an ambulance company with fewer than 20 patients in a given zip code. Drop rides from zip codes with only one remaining ambulance company.	3,919,572	1,372,499	5,716	119	3,999
5. Prior utilization restriction	Keep only rides for patients with some non-VA utilization (inpatient, ED, or primary care) but no VA utilization in the prior year.	1,735,141	644,917	5,716	97	3,812
6. Prior ride restriction	Drop rides for patients with a prior ride within 30 days.	1,414,217	644,917	5,716	96	3,799

Note: This table details selection steps to create two alternative samples. Panel A shows selection steps for the sample that we use to study descriptive utilization patterns in Appendix Figure A.8. Panel B shows selection steps for the sample of patients with only non-VA prior utilization, which we use in Section 5.5 to study the effect of receiving care at a modal non-VA hospital. At each step, the table lists the number of ambulance rides, patients, ambulance companies, and VA and non-VA hospitals. Table A.14 shows average patient characteristics among observations at each sample step.

Table A.14: Characteristics of Non-VA-Only Sample

Restrictions	Sample characteristics				
	Dually eligible	Add zip × hospital	Add zip × ambulance	Add non-VA-only prior utilization	Add no ride in prior month
Male	0.899	0.898	0.897	0.824	0.825
Age	77.04	77.12	77.32	77.68	78.05
Black	0.111	0.124	0.129	0.125	0.118
Income	$22,222	$22,243	$22,243	$23,393	$23,393
Rural zip code	0.255	0.169	0.125	0.120	0.120
Residential Source	0.610	0.619	0.657	0.614	0.636
Comorbidity count	6.53	6.62	6.60	6.96	6.57
Prior VA ED visit	0.136	0.141	0.138	0.000	0.000
Prior Medicare ED visit	0.695	0.694	0.687	0.797	0.752
Ambulance rides in prior year	2.77	2.83	2.82	3.13	2.28
Advanced Life Support	0.696	0.695	0.699	0.676	0.684
Weekend rate	0.272	0.271	0.271	0.270	0.269
28-day mortality	0.115	0.116	0.113	0.117	0.112
Present at VA	0.044	0.049	0.049	0.002	0.002
Number of patients	2,862,557	2,131,152	1,372,499	644,917	644,917
Number of ambulance rides	8,828,997	6,424,120	3,919,572	1,735,141	1,414,217

Note: This table presents characteristics of observations remaining at each step of creating the sample of patients with only non-VA prior utilization, which we use in Section 5.5 to study the effect of receiving care at a modal non-VA hospital. Each step is detailed in Panel B of Appendix Table A.13.