Online Appendix for

"Is There a VA Advantage?

Evidence from Dually Eligible Veterans"

David C. Chan

David Card

Lowell Taylor

October 2023

Appendix for Online Publication

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 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:

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

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 the 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.[51] 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 to which the ambulance company delivers its patients. 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.14). 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 IV.B, 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

---

51. 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.

same baseline controls as in our benchmark analyses in Section II with the addition of various ambulance-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, Dobbie, and Yang 2018; Bhuller et al. 2020). We define eight pairs of subsamples based on several important patient characteristics: (i) age ≤ 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 with 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):

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

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 $\mathcal{I}_m$:

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

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.[52] 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 and reverse-sample instruments. Coefficient sizes

---

52. We use the analytical sample to 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).

are generally smaller for the reverse-sample instruments. In Appendix Table A.6, we show first-stage relationships using two other instruments based on the smaller analytical sample. Specifically, we construct a "baseline" instrument, $\tilde{Z}_i$, and an "in-sample" instrument, $\tilde{Z}_i^m$, from the analytical sample:

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

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

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 the 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 III, let $s_{IV}\,(t;d) \equiv E\,[\,S_i\,(t;d)|\,i \in C]$ 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 $t \in \{0,1,\dots,52\}$. We then define the corresponding estimands of the potential mortality *hazards* as follows:

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

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

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

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

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

We construct similar objects for potential survival and hazard rates under OLS: $\hat{s}_{OLS}\,(t;d)$ and $\hat{h}_{OLS}\,(t;d)$, respectively. Using the same set of block bootstrap samples, we compute $\hat{s}_{OLS}^r\,(t;d)$ and

$\hat{h}_{OLS}^r(t;d)$ in each bootstrap replication sample r.

A.2.2. Test of Mortality Displacement

To detect "mortality displacement" (Schwartz 2000), in which deaths of VA patients are only 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

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

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):

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

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:

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

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

$$(A.9) \qquad Q_1^r \equiv \sum_{t=1}^{52} w_{1,t} \mathbf{1}\left(\tilde{h}_{IV}^r(t;0) - \tilde{h}_{IV}^r(t;1) < 0\right)\left(\tilde{h}_{IV}^r(t;0) - \tilde{h}_{IV}^r(t;1)\right)^2.$$

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

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

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:

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

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.[53]

We show results in Panel A of Appendix Figure A.5. We find that Q_1 is within the distribution of bootstrapped values of Q_1^r. Therefore, we 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 II.B 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

$$(A.12) \qquad \Pr(Y \in B, D = 1 | Z = 1) - \Pr(Y \in B, D = 1 | Z = 0) \quad \geq \quad 0;$$
$$(A.13) \qquad \Pr(Y \in B, D = 0 | Z = 0) - \Pr(Y \in B, D = 0 | Z = 1) \quad \geq \quad 0.$$

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.[54]

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

---

53. 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 bootstrapped 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.

54. Chan, Gentzkow, and Yu (2019) provide an applied example, in the setting of radiologists. In this paper, standard monotonicity tests in Appendix A.1.2 are satisfied, but a simple version of this extended test of validity is strongly rejected. 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 $B \subset \mathcal{Y}$,

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

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

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, $h_{IV}(t;d)$, 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:

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

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

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

where $w_{2,t}^{-1} = \left(\hat{\sigma}_{IV}^B(t;d)\right)^2$. 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

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

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

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

We show results in Panel B of Appendix Figure A.5. We find that Q_2 is lower than any bootstrapped value of Q_2^r. This suggests that we cannot reject the null hypothesis $H_{0,2}$ in equation (A.16) and 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 $h_{IV}(t;d) = 0$, for all $t \geq 1, d \in \{0,1\}$, which means that $h_{IV}(t;d) > 0$ for at least 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, $h_1(t)$ and $h_2(t)$, for $t \geq 2$. We consider two types of null hypotheses.

First, we assess mean differences in hazard rates between $\{h_1(t)\}_t$ and $\{h_2(t)\}_t$, for $t \geq 1$, under the null hypothesis that the mean hazard rate is the same between the two sets:

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

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:

$$
\text{(A.19)} \qquad H_{0,4} : h_1(t) - h_2(t) = 0, \text{ for all } t \geq 1.
$$

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.6 and A.7, we consider five comparisons of hazard rates, for $t \geq 1$, 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 $t \geq 1$. 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 $t \geq 1$. While we show the 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 $t \geq 1$. 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 $t \geq 1$. 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, we describe a simple approach to calculate characteristics of non-compliers, following Dahl, Kostol, and Mogstad (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, $(\ell(i), \mathbf{X}_i^0)$. 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 5. 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: For many characteristics, those that are more common among compliers tend to be more common among always-takers and less common among never-takers. Compared to the overall population, always-takers are more likely to be Black and have lower income. Always-takers are more likely to have a 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. However, both always-takers and never-takers, as defined by this methodology, have slightly higher predicted mortality.

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 $(\ell(i), \mathbf{X}_i^0)$: $P(Z_i)$, where we have omitted the key controls for brevity. Following Heckman and

Vytlacil (2005), we can state the treatment rule as

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

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

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

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

The average treatment effect (ATE) is

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

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 $MTE\left(u\right)$ 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 $MTE\left(u\right)$ can be stated as

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

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 $P\left(Z_i\right)$, 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.2 suggests a slightly convex shape in the relationship between mortality and $P\left(Z_i\right)$, which implies selection on gains.

We proceed with estimating a flexible relationship between Y_i and $P\left(Z_i\right)$ as follows. We compute $P\left(Z_i\right) = \hat{D}_i$ 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 $\hat{\beta}_{IV}$.[55]

---

55. This regression corresponds to the indirect least squares version of IV and is numerically identical to the visual IV coefficient corresponding to the two-stage least squares version of IV.

Rather than fitting a straight line through points $(\hat{D}_i^*, Y_i^*)$, 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 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

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

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.8.

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

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

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 II, regardless of the choice of $J(u)$.[56]

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

---

56. Kline and Walters (2019) show algebraic equivalence between the control function LATE implied by equation (A.21), $\underline{p}$, and $\overline{p}$, under a binary instrument and 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):[57]

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

This expectation corresponds to the following regression:

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

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 II. For inference on the difference between the ATE and the LATE, we recover a numerically equivalent LATE with the following control function regression:[58]

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

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 $g_d(\varepsilon) \equiv E\left[\varepsilon_{0,i} \mid D_i = d, \varepsilon_{1,i} = \varepsilon\right]$, where $\varepsilon_{0,i}$ is the structural error term in equation (2). Specifically, we estimate regressions of the following form:

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

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. ATE estimates are all smaller in magnitude than the LATE estimate from Section II. We compute standard errors on this difference with 50 bootstrapped iterations (selecting samples by zip code, with replacement).

---

57. 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 $\lambda_d(p) \equiv E\left[J(U_i) - \mu_J \mid D_i = d, P(Z_i) = p\right]$. 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 $\varepsilon_{1,i}$: $E\left[J(U_i) - \mu_J \mid \varepsilon_{1,i} = \varepsilon\right] = -\frac{\varepsilon}{2}$.

58. 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.5.   Hospital Characteristics

This appendix provides further details on hospital characteristics that we use in our heterogeneity analyses in Section IV.A. These characteristics are listed in Table 6 and Appendix Tables A.10 to A.12. 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) number of facility admissions; (iii) number of available hospital beds; (iv) teaching hospital status; (v) trauma center status; (vi) number of privileged ED staff, which we use to construct ED staff per 100 ED visits given (i); (vii) number of full-time registered nurses, which we use to construct nurses per 100 admissions given (ii); (viii) number of privileged hospitalists, which we use to construct hospitalists per 100 admissions; and (ix) number of privileged intensivists, 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. Relative spending is therefore not comparable between VA and non-VA hospitals. The AHA average of relative spending for VA hospitals is not 1 because not all VA hospitals are found in the AHA Annual Survey. 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, 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 any electronic health record certified products on the Certified Health IT Product List (CHPL) reported on `healthIT.gov`. Additional characteristics in Table 6 are also obtained from the AHA Annual Survey: (i) average daily census, (ii) urban location (i.e., the hospital

is not classified as either "micro" or rural), (iii) capitated lives covered, and (iv) Preferred Provider Organization (PPO) affiliation.

A.6.   Heterogeneity by Observable Characteristics

This appendix describes our analytical approach to estimating treatment effect heterogeneity by observable hospital or patient characteristics. As described in Section IV.A, we have three categories of characteristics: (i) characteristics of non-VA hospitals serving a given zip code, weighting the hospitals by volume of rides from the zip code; (ii) characteristics of the VA hospital serving a given zip code; and (iii) patient characteristics. Hospital characteristics are described in further detail in Appendix A.5.

For each characteristic x, we construct a binary indicator variable, $I_{x,i} \in \{0,1\}$. For example, for the non-VA hospital characteristic of the 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 the 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:

$$(A.26) \qquad Y_i = \beta_x D_i + \rho_x D_i \tilde{I}_{x,i} + \pi_x \tilde{I}_{x,i} + \gamma_x \hat{\varepsilon}_{1,i} + \mathbf{X}_i^0 \delta_x + \zeta_{x,\ell(i)} + \epsilon_{x,i},$$

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. This approach enables greater statistical power than performing separate IV regressions in subsamples defined by $I_{x,i} \in \{0,1\}$. For a discussion of this general approach, see Wooldridge (2015), Section III. 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.

A.7.   OLS Heterogeneity in Station-Specific VA Advantage

In analyses described in this appendix, we estimate OLS heterogeneity in the station-specific VA advantage and validate this heterogeneity with our quasi-experiment. As in our heterogeneity analyses in Section IV.A, we assign each zip code ℓ to a VA station $s(\ell)$ based on the station that the most veterans living in that zip code use. This assignment of zip codes to VA stations matches station catchment areas for 92% of zip codes.

In separate OLS regressions, we estimate the VA advantage for each station s as β_s in

$$(A.27) \qquad Y_i = \beta_{OLS}^s D_i + \mathbf{X}_i^0 \delta^s + \zeta_{\ell(i)}^s + \varepsilon_i,$$

using ambulance rides i such that the zip code $\ell(i)$ maps to station s (i.e., $s(i) \equiv s(\ell(i)) = s$). The

ride-weighted variance of $\hat{\beta}^s_{OLS}$ is 3.4×10^{-4}, while the ride-weighted variance of the sampling error for each $\hat{\beta}^s_{OLS}$ is 2.1×10^{-4}. This implies a sampling-error-adjusted, ride-weighted variance of β^s_{OLS} of $A = (3.4 - 2.1) \times 10^{-4} = 1.4 \times 10^{-4}$, or a standard deviation of β^s_{OLS} of $\sqrt{A} = 0.012$.

In Appendix Figure A.9, we plot the distribution of $\hat{\beta}^s_{OLS}$ for 32 stations with at least 5,000 rides, forming a smple of 276,483 rides. We also plot the empirical Bayes posteriors for all stations, which we calculate as follows:

$$(A.28) \qquad \tilde{\beta}^s_{OLS} = (1 - B_\ell)\hat{\beta}^s_{OLS} + B_s\hat{\beta}_{OLS},$$

where $B_s = \frac{V_s}{s+A}$ is the shrinkage factor based on V_s, which is the variance of the sampling error for station s, and A, which is the variance of the prior distribution of β^s_{OLS}. $\hat{\beta}_{OLS} = -0.024$ is the overall OLS estimate reported in Section II.C. This figure shows that essentially all stations exhibit a VA advantage, at least when estimated by OLS.

We evaluate whether differences in $\tilde{\beta}^s_{OLS}$ imply differences in the treatment effects identified by our quasi-experiment. As a first analysis, we divide stations into two groups depending on whether $\tilde{\beta}^s_{OLS}$ is above- or below-median. We estimate by two-stage least squares $\hat{\beta}_{IV}$, based on equations (3) and (4), separate IV estimates for ambulance rides belonging to each of these two groups. $\hat{\beta}_{IV}$ estimated for stations with below-median (i.e., larger in magnitude) $\tilde{\beta}^s_{OLS}$ is 0.030 larger in magnitude than the same estimate for stations with above-median (i.e., smaller in magnitude) $\tilde{\beta}^s_{OLS}$. However, the difference is imprecise, with a bootstrapped standard error of 0.051.

For a more systematic validation of $\tilde{\beta}^s_{OLS}$, in the spirit of Angrist et al. (2017), we conduct a pooled analysis by indirect least squares. Specifically, denoting demeaned $\tilde{\beta}^s_{OLS}$ as $\tilde{\beta}^{s*}_{OLS}$, we estimate

$$Y_i = \beta D_i + \gamma D_i \times \tilde{\beta}^{s(i)*}_{OLS} + \mathbf{X}^0_i \delta + \zeta_{\ell(i)} + \varepsilon_i,$$

where we instrument D_i and $D_i \times \tilde{\beta}^{s(i)*}_{OLS}$ by Z_i and $Z_i \times \tilde{\beta}^{s(i)*}_{OLS}$. This regression reveals an imprecise and wrong-signed result of $\hat{\gamma} = -0.790$ (s.e., 1.351). The overall imprecision of these results suggests that there is little signal of heterogeneity across station-specific OLS measures of the VA advantage. The more precise results in Section IV.A also suggest little meaningful heterogeneity along binary characteristics of VA and non-VA hospitals in a given zip code.

A.8. Reported Utilization Patterns

This appendix details comparisons of reported utilization patterns between VA and non-VA hospitals. Our analyses are based on utilization from the VA and Medicare data corresponding to any patient in the baseline sample in the 28 days following his or her ambulance ride. Each item of utilization corresponds to a service defined by its Current Procedural Terminology (CPT) code.

Our first set of analyses examine the share of utilization originating from the VA across different CPT codes. Specifically, we define this share as the proportion of utilization for a CPT code originating from VA records out of the total utilization for that CPT code reported by both VA and non-VA

(i.e., Medicare) providers. Figure 6 shows VA shares for the top 25 (out of 5,167) CPT codes in the Medicare Physician Fee Schedule (MPFS), ranked by total utilization.

We find a wide range of VA shares even within this set of common CPT codes. At one extreme, only 4.1% of utilization for CPT code 99223, one of the codes for evaluation and management (E/M) performed in initial hospital care, originate from the VA. Also with a VA share of 4.1%, CPT code 99239 reports E/M care lasting more than 30 minutes on the discharge day of a hospitalization. For this code to be reported, the physician must report spending more than 30 minutes with the patient. In contrast, the complementary E/M CPT code that reports spending 30 or fewer minutes on discharge day (99238) is more than four times likelier (17.1%) to originate from the VA. Non-VA hospitals have a clear financial incentive to report the code 99239 over 99238 (the former reimburses close to 50% more), but differentiating between the two services has no clinical value. At the other extreme, 90.5% of utilization for CPT code 99211, which reports a simple outpatient E/M service not requiring the presence of a physician, originate from the VA. Strikingly, *all* of the reported utilization of CPT code 98966, for short calls made by a non-physician, occur in the VA.

Appendix Figure A.11 shows similar VA shares for the top 25 (out of 115) groups of Category I CPT codes, ranked by total utilization. This figure shows similar patterns, albeit for much larger aggregations of utilization. Non-VA providers much more commonly report hospital inpatient E/M services. The VA much more commonly reports physical therapy, rehabilitation, psychiatric services, and telephone (i.e., non-face-to-face) services provided by non-physicians. Pulmonary services— the vast majority of which comprise low-reimbursed services such as measuring oxygen levels and providing inhalation treatment—are also much more commonly reported in the VA.

We examine the relationship between reimbursement and the share of a CPT code's utilization coming from the VA. We measure reimbursement among CPT codes on the MPFS, multiplying year-specific relative value units (RVUs) with the year-specific dollar conversion factor. In Figure 7, we show a strong negative relationship between the median reimbursement (across years) for a given CPT code and its VA share. Importantly, reimbursement by Medicare for physician services is determined by the resource-based relative value scale (RBRVS), a system entirely based on the costliness of procedures and not on the benefit of procedures (American Medical Association 2015). Thus, services with high potential value relative to their costs (e.g., telephone calls) are reimbursed little and much less likely reported in the fee-for-service system outside the VA.

We finally focus on evaluation and management (E/M) CPT codes, which allow for reporting of complexity. E/M codes are among the most common CPT codes and reflect an integral part of clinical care, particularly for emergency patients. Reimbursement may vary widely across E/M codes reporting different levels of complexity. For example, the set of CPT codes 99201-99205, collectively for "office or other outpatient encounters for new patients," may range almost fivefold in reimbursement. The complexity allowed for an E/M code reported for an encounter depends on documentation, but much of the documentation is ultimately unverifiable. For these reasons, Fang and Gong (2017) devote much of their analysis to detecting potential overbilling to E/M codes.

In Appendix Figure A.12, we show the odds of reporting the highest level of complexity within

a type of E/M code relative to reporting the lowest level of complexity with that type among non-VA vs. VA providers. We display the odds ratio (i.e., the non-VA odds divided by the VA odds) on the *x*-axis for seven broad categories of E/M codes. We show that the odds of reporting the highest level of complexity are much higher among non-VA providers. In only one category (i.e., critical care) is the odds ratio close to one. The (volume-weighted) average odds ratio is 5.1.

A.9. Modal-Hospital Mechanisms

This appendix details analyses in Section IV.C, where we describe indirect evidence for the role of health IT and integrated care. We perform analyses on a sample of veterans who only use non-VA care in the year prior to their ambulance rides. Since no veteran in this sample has prior VA utilization, the sample is disjoint from our benchmark sample (Appendix Table A.1). We only include zip codes with at least two non-VA hospitals within 20 miles, but we make no requirement on proximity to a VA hospital. The probability of transport to a VA hospital in this sample is 0% (as opposed to 33% in the benchmark sample), but rates of weekend transport and 28-day mortality are remarkably similar. We detail the sample selection process for this analysis in Appendix Table A.14 and present patient and ride characteristics in Appendix Table A.15.

Quasi-Experimental Design. 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 $h^m(i)$ 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 or her modal hospital. Our instrumental variable for this treatment is

$$(A.29) \qquad Z_i^m = \frac{1}{K_{j(i),z(i)} - 1} \sum_{i' \in \mathcal{I}_{j(i),z(i)}} \frac{\mathbf{1}(k(i') \neq k(i)) D_{i'}^m}{N_{k(i'),z(i'),j(i')}}.$$

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 $k(i)$.[59] We use the following first-stage and reduced-form equations, similar to equations (3) and (4):

$$(A.30) \qquad D_i^m = \pi_1^m Z_i^m + \gamma_1^m \overline{Z}_i^m + \mathbf{X}_i^0 \delta_1^m + \zeta_{1,\ell(i)}^m + \varepsilon_{1,i}^m;$$
$$(A.31) \qquad Y_i = \pi_2^m Z_i^m + \gamma_2^m \overline{Z}_i^m + \mathbf{X}_i^0 \delta_2^m + \zeta_{2,\ell(i)}^m + \varepsilon_{2,i}^m.$$

---

59. 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 Appendix Table A.15). 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."

We include in these equations an additional control variable:

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

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 $\ell(i)$ 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) while including hospital fixed effects in equations (A.30) and (A.31) and (ii) while splitting rides i into samples based on whether the ride was before or after 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.14), we demonstrate in Appendix Figure A.14 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 $\left(\overline{Z}_i^m, \mathbf{X}_i^0, \ell(i)\right)$.[60]

Results. The IV estimate 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.15 shows that the overall relationship between the reduced form and first stage is not particularly striking.[61] However, computing this IV estimate separately by years, we show in Figure 8 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.

To extend this analysis, we use hospital-specific dates of hospital health IT adoption or ACO participation (described in Appendix A.5). During our sample period, a sizable proportion of hospitals adopted health IT and, to a much lesser extent, participated in an ACO. We construct four subsamples defined by whether or not each veteran's modal hospital had adopted health IT at the time of his or her ambulance ride and similarly by whether or not each veteran's modal hospital had joined an ACO. In each subsample, we performed the same IV regression of the effect of transport to a veteran's modal hospital. Results are shown in Appendix Table A.16, 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 (A.30) and (A.31), respectively. Results are qualitatively unchanged regardless of their inclusion.

In Columns 3 and 6 of Appendix Table A.16, we also perform regressions in the overall sample (described in Panel B of Appendix Table A.14). We maintain all of the interactions implicit in our subsample results except that we allow hospital group 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

---

60. Analogously to Figure 1, this figure presents binned scatter plots of the first-stage regression in equation (A.30), the reduced-form regression in equation (A.31), and a balance regression with predicted mortality as the outcome variable and the same design matrix.

61. Analogously to Figure 2 and Appendix Figure A.3 in the benchmark analysis, Appendix Figure A.16 shows stability in OLS and two-stage least squares estimates with increasing controls, and Appendix Figure A.17 shows the robustness of two-stage least squares estimates under an exhaustive set of control combinations.

estimate a first-stage regression that interacts everything with adoption status, except for fixed effects for hospital groups, $g\left(h\left(i\right)\right)$, defined by whether a hospital ever adopts health IT or an ACO:

$$(A.32) \qquad D_i^m = \sum_{a \in \{0,1\}} \mathbf{1}\left(\text{Adopted}_i = a\right)\left(\pi_{1,a}^m Z_i^m + \gamma_{1,a}^m \overline{Z}_i^m + \mathbf{X}_i^0 \delta_{1,a}^m + \zeta_{1,\ell(i),a}^m\right) + \xi_{1,g(h(i))}^m + \varepsilon_{1,i}^m.$$

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

$$(A.33) \qquad Y_i = \sum_{a \in \{0,1\}} \mathbf{1}\left(\text{Adopted}_i = a\right)\left(\beta_a D_i^m + \gamma_a \hat{\varepsilon}_{1,i}^m + \mathbf{X}_i^0 \delta_a + \zeta_{\ell(i),a}\right) + \xi_{g(h(i))} + \epsilon_i.$$

As with our other control-function regressions, we compute standard errors by 50 bootstrapped iterations, drawing samples by zip code blocks, with replacement. While estimates control for hospital or hospital group fixed effects, we find that results are essentially unchanged regardless of their inclusion.

References

AMERICAN MEDICAL ASSOCIATION. 2015. "The Physicians' Guide: Medicare RBRVS." Edited by Sherry L. SMITH. Chicago: AMA.

———. 2017. "Principles of CPT Coding." 9th. American Medical Association, May.

ANGRIST, Joshua D., Peter D. HULL, Parag A. PATHAK, and Christopher R. WALTERS. 2017. "Leveraging Lotteries for School Value-Added: Testing and Estimation." *Quarterly Journal of Economics* 132 (May): 871–919.

BLUNDELL, Richard, and Rosa L. MATZKIN. 2014. "Control Functions in Nonseparable Simultaneous Equations Models." *Quantitative Economics* 5:271–295.

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,* 9099–9135. February.

CHANDRA, Amitabh, Amy FINKELSTEIN, Adam SACARNY, and Chad SYVERSON. 2016. "Healthcare Exceptionalism? Productivity and Allocation in the U.S. Healthcare Sector." *American Economic Review* 106 (August): 2110–2144.

KLINE, Patrick M, and Christopher R WALTERS. 2019. "On Heckits, LATE, and Numerical Equivalence." *Econometrica* 87 (March): 677–696.

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

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

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

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

Figure A.1: Balance Coefficient in Simulated Data
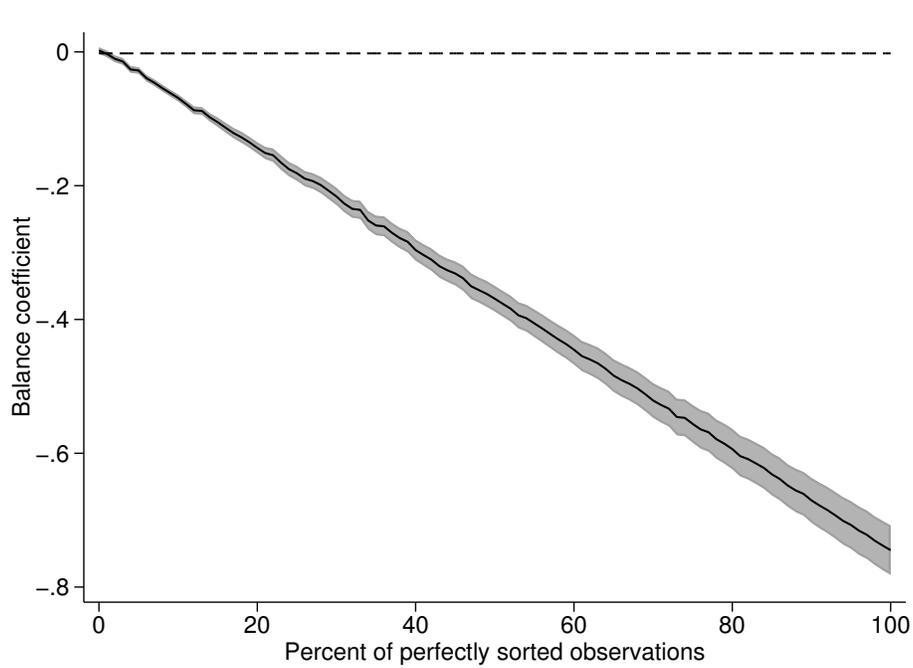
Note: This figure plots the balance coefficient in the baseline sample and in simulated data in which we perfectly sort a percent of ambulance rides and randomly assign the rest of the ambulance rides. The y-axis shows the balance coefficient, and the x-axis shows the percent of perfectly sorted observations in the simulated data, or $\iota \in \{0, 1, \ldots, 100\}$. The dashed horizontal line indicates the balance coefficient in the baseline sample, also shown in Panel B of 1. Each simulated dataset comprises observations of residuals of predicted mortality $\hat{Y}_i$ and residuals of Z_i, formed by regressions of each object on baseline controls. We form the simulated dataset by perfectly sorting ι percent (in expectation) of $\hat{Y}_i$ residuals according to ambulances that are sorted by their mean Z_i residual and randomly assigning the remaining $1 - \iota$ percent (in expectation). We assign rides only within their original zip code, also holding the number of rides assigned to each ambulance company within the zip code fixed. The regression of reassigned residual $\hat{Y}_i$ on residual Z_i gives the balance coefficient in each simulated dataset, shown in the solid black line. The shaded gray region indicates the 95% confidence interval, which we obtain by 20 bootstrapped replications drawn by zip code blocks with replacement. The upper confidence limit intersects with the actual balance coefficient between $\iota = 2$ and $\iota = 3$.

Figure A.2: Visual IV
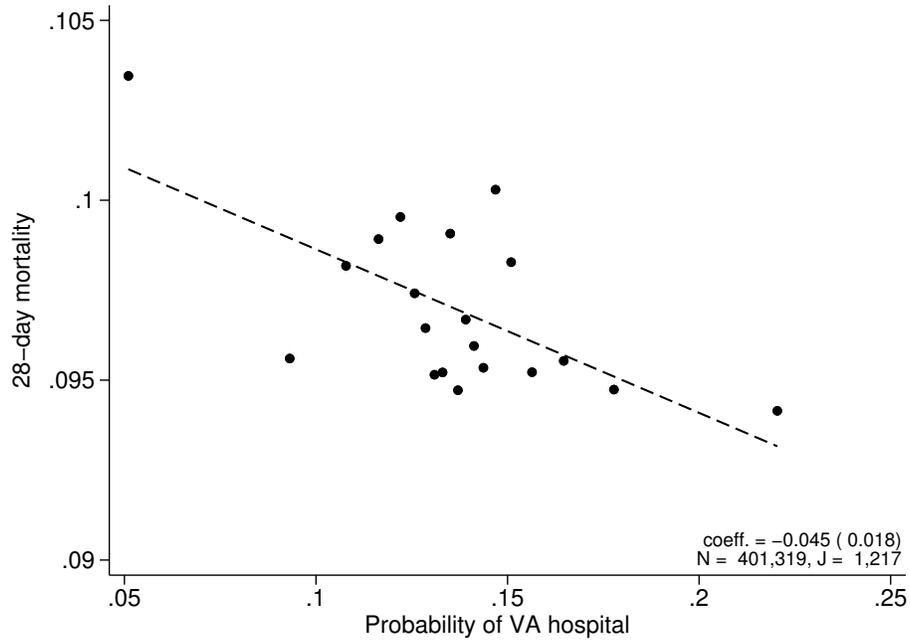
Note: This figure shows the visual IV plot corresponding to our baseline IV regression of the effect of the VA on 28-day mortality. For each bin of the instrument, which is the ambulance leave-out propensity to arrive at a VA hospital, we plot the mean 28-day mortality on the *y*-axis and the probability 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 $401,319$ observations and $1,217$ 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, detailed in Appendix Table A.2.

Figure A.3: Combinations of Controls
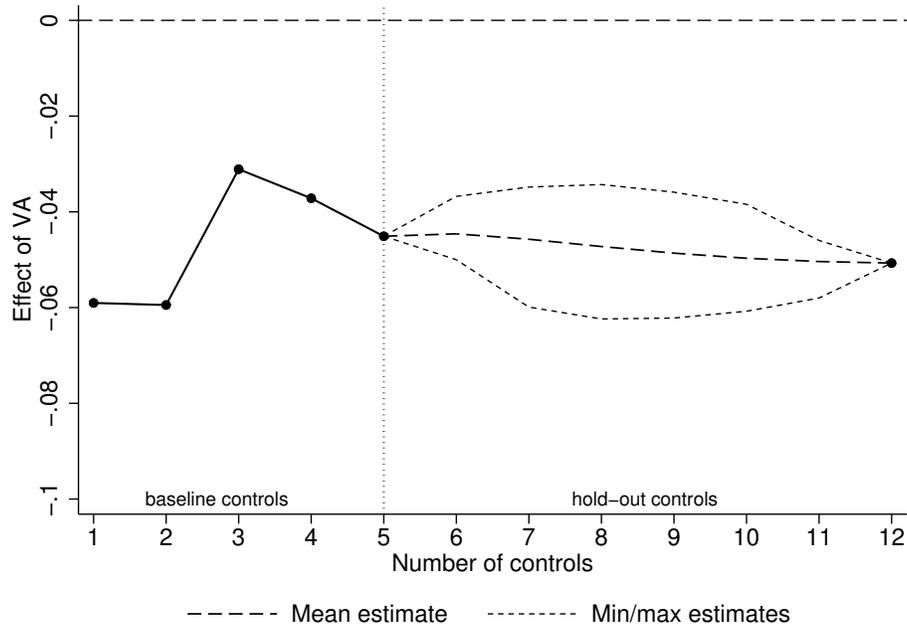
Note: This figure shows IV estimates of the VA effect on 28-day mortality on the *y*-axis, from equation (2), varying the number of controls included in the IV regression. Numbered incremental controls correspond to categories or subcategories of variables presented in order in Appendix Tables A.2 and A.3. All specifications include the five baseline controls. Therefore, the figure represents $5 + (2^7 - 1) = 132$ specifications. For each number of controls *n* for $n > 5$, we consider "7 choose $n - 5$" 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.4: Treatment Effects by Time and Sample

A: OLS

B: IV
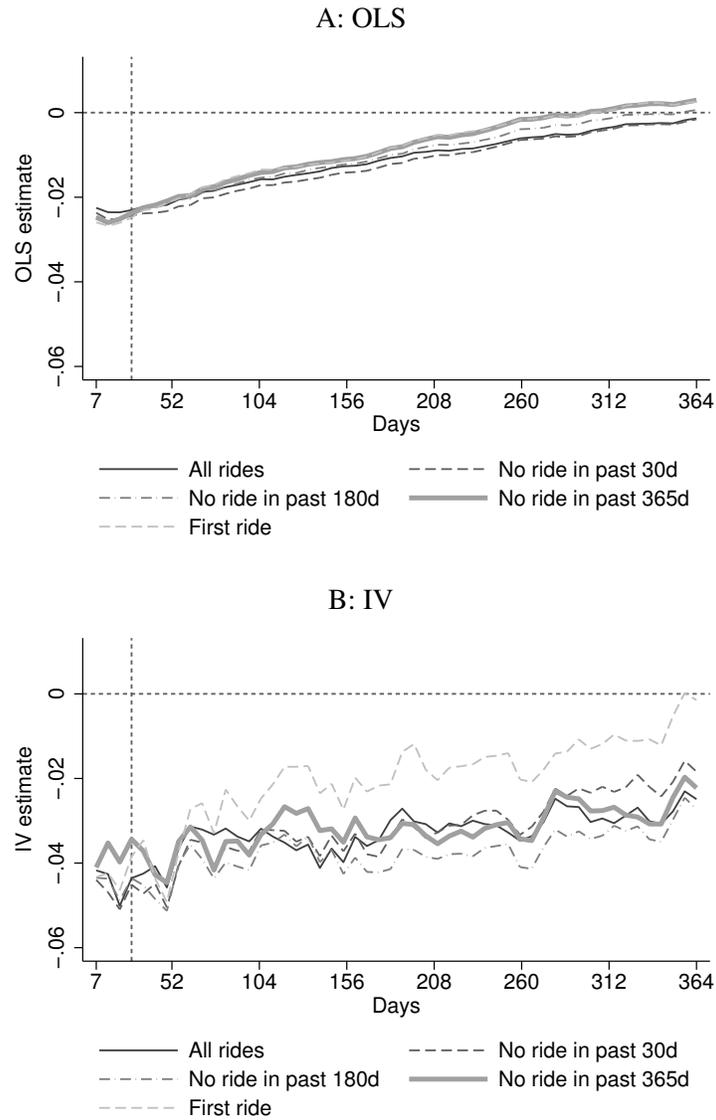
Note: This figure shows mortality treatment effects over varying days since the ambulance ride and in varying samples dropping patients with prior rides. "Days" indicate one-week intervals from the ambulance ride. Panel A shows OLS results corresponding to equation (6). Panel B shows IV results corresponding to equation (5). The vertical dashed line indicates treatment effects on 28-day mortality, our baseline outcome.

Figure A.5: Joint Inequality Constraints

A: No Mortality Displacement

test stat. = 29.8
crit. value = 41.9

B: No Negative Hazard Rate

test stat. = 4.3
crit. value = 72.7
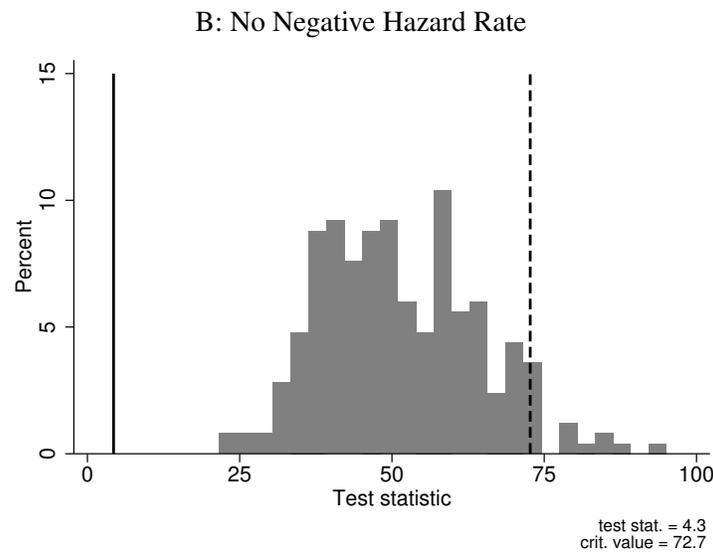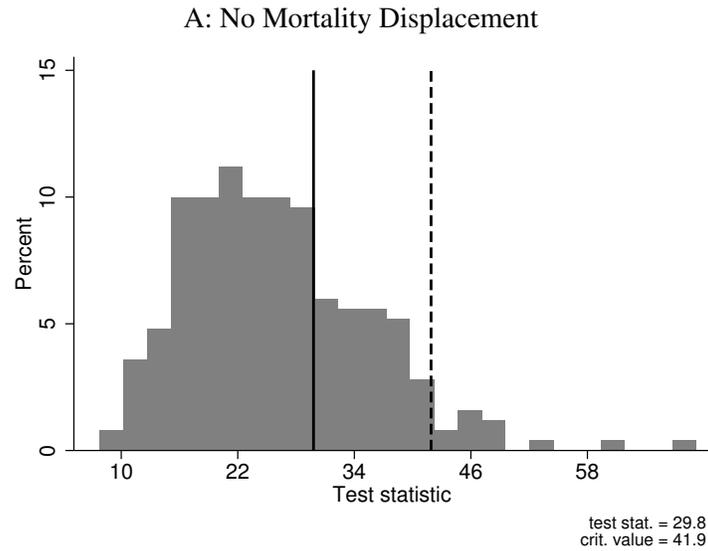
Note: This figure shows the test statistic for joint inequality constraints and bootstrapped-generated distributions of the test statistic under the least favorable version of the null hypothesis. Panel A shows the joint inequality test of no mortality displacement, as defined by the null hypothesis 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. The test statistic and the bootstrap procedure for Panels A and B are described further in Appendices A.2.2 and A.2.3, respectively.

Figure A.6: Mean Hazard Differences

A: VA vs. Non-VA Compliers

difference = 0.022
std. err. = 0.078

B: VA vs. Non-VA Users

difference = 0.059
std. err. = 0.006

C: VA Compliers vs. Users

difference = 0.055
std. err. = 0.048

D: Non-VA Compliers vs. Users

difference = 0.092
std. err. = 0.063
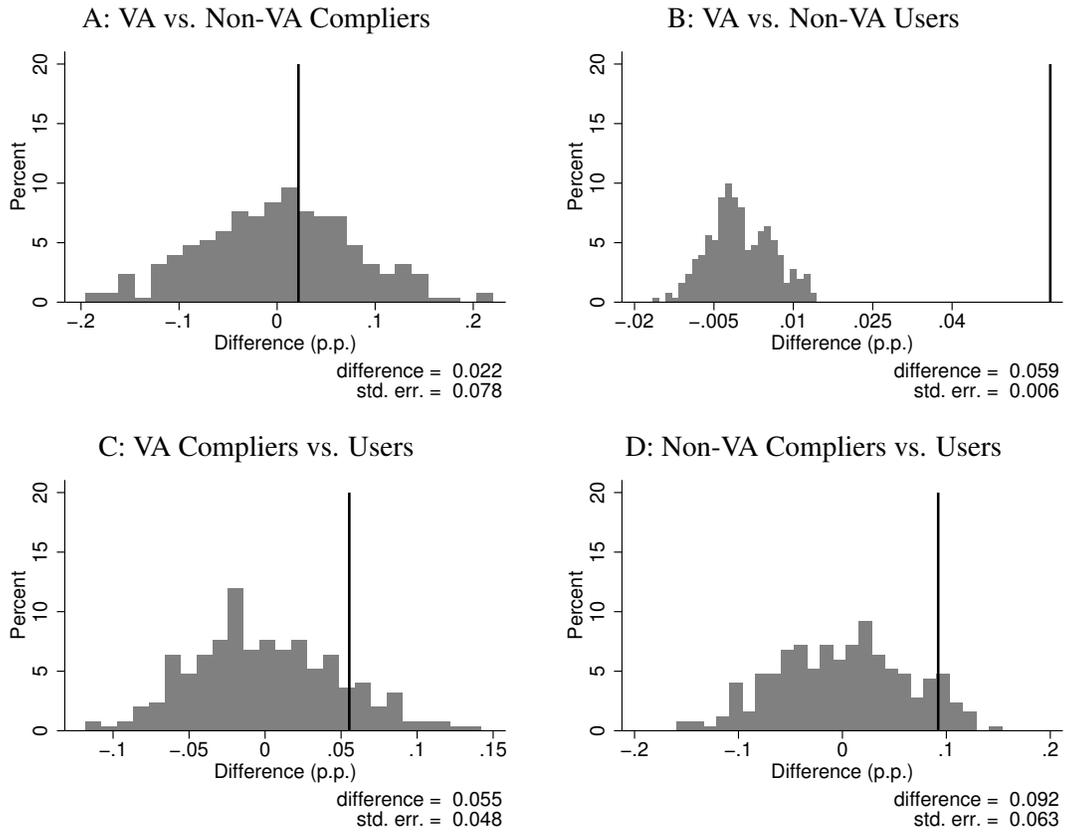
Note: This figure shows tests of equality of mean hazard rates for different sets of hazard rates, as defined by the null hypothesis in equation (A.18). Each panel corresponds to a comparison between sets of hazards corresponding to VA or non-VA compliers or users. Details of the statistical procedure 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.7: Joint Equality Constraints
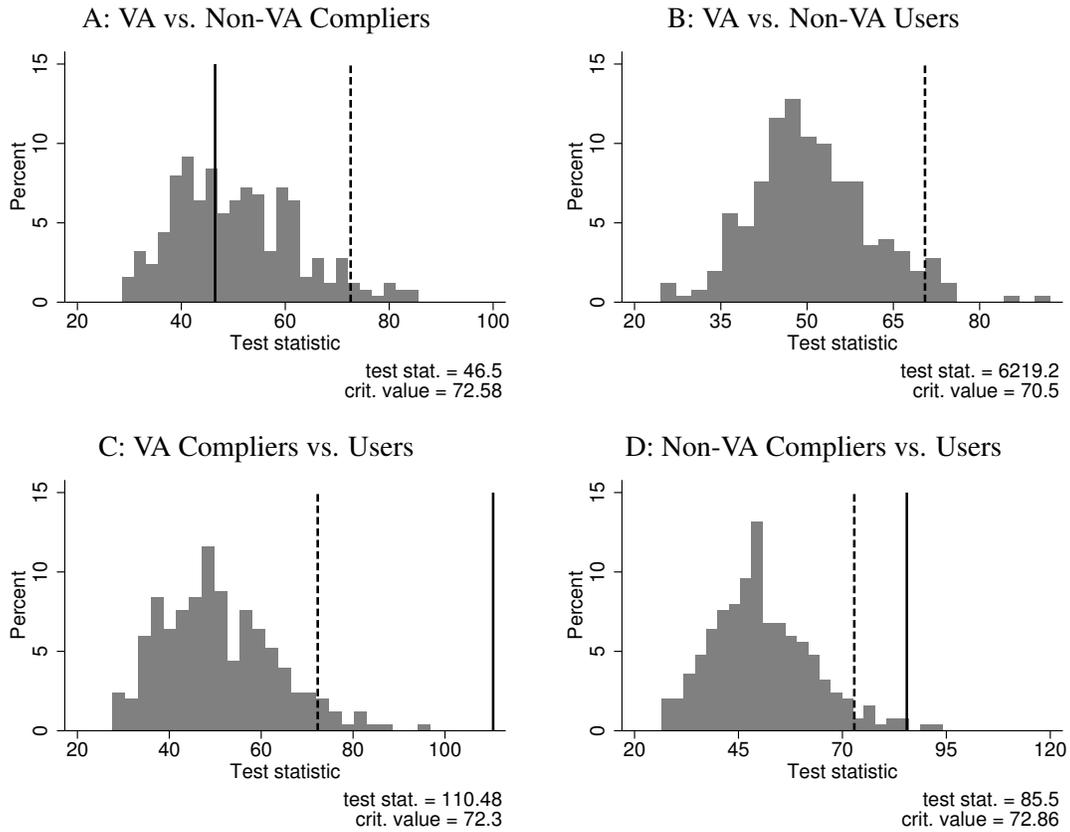
Note: This figure shows tests of joint equality of hazard rates for different sets of hazard rates, as defined by the null hypothesis in equation (A.19). Each panel corresponds to a comparison between sets of hazards corresponding to VA or non-VA compliers or users. Details of the statistical procedure 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. The dashed line shows the one-sided 95th percentile critical value.

Figure A.8: Marginal Treatment Effects

A: Flexible Visual IV

B: Marginal Treatment Effects
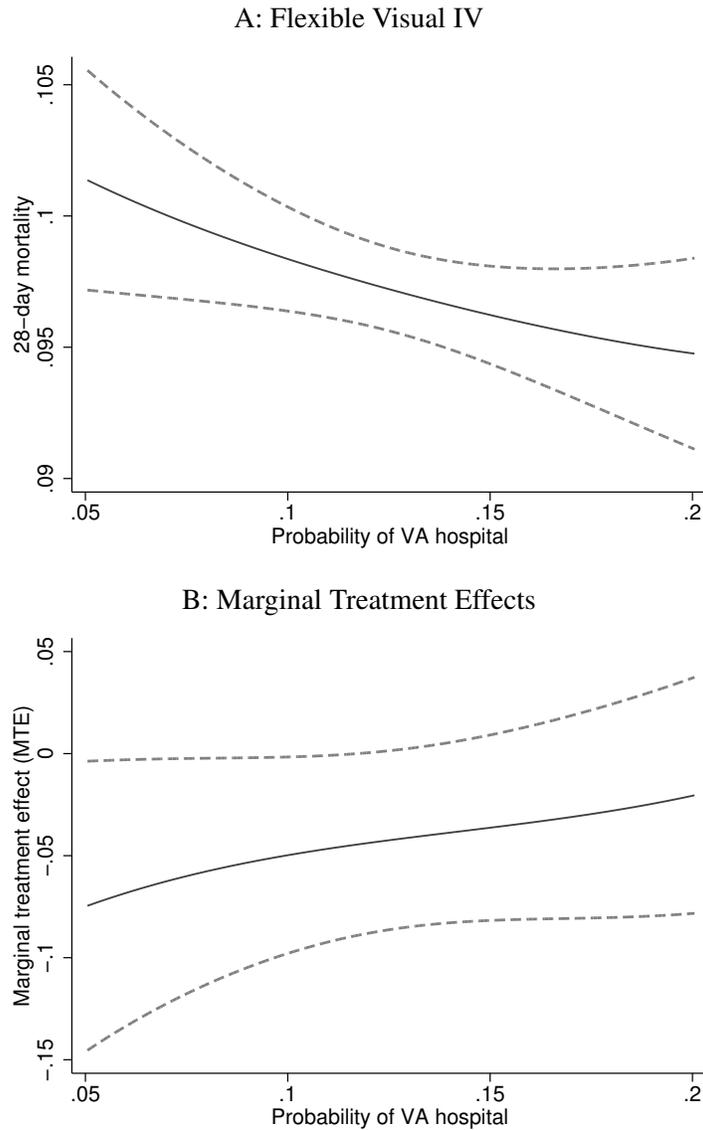
Note: This figure shows a flexible fit of the IV relationship between 28-day mortality and the ambulance propensity to transport to a VA hospital. Panel A shows the visual IV relationship with residual 28-day mortality on the *y*-axis and residual probability of being transported to a VA hospital on 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.2. 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.9: Station-Specific OLS Estimates of VA Advantage

A: Stations with at Least 5,000 Rides

B: All Stations, Shrunken Estimates
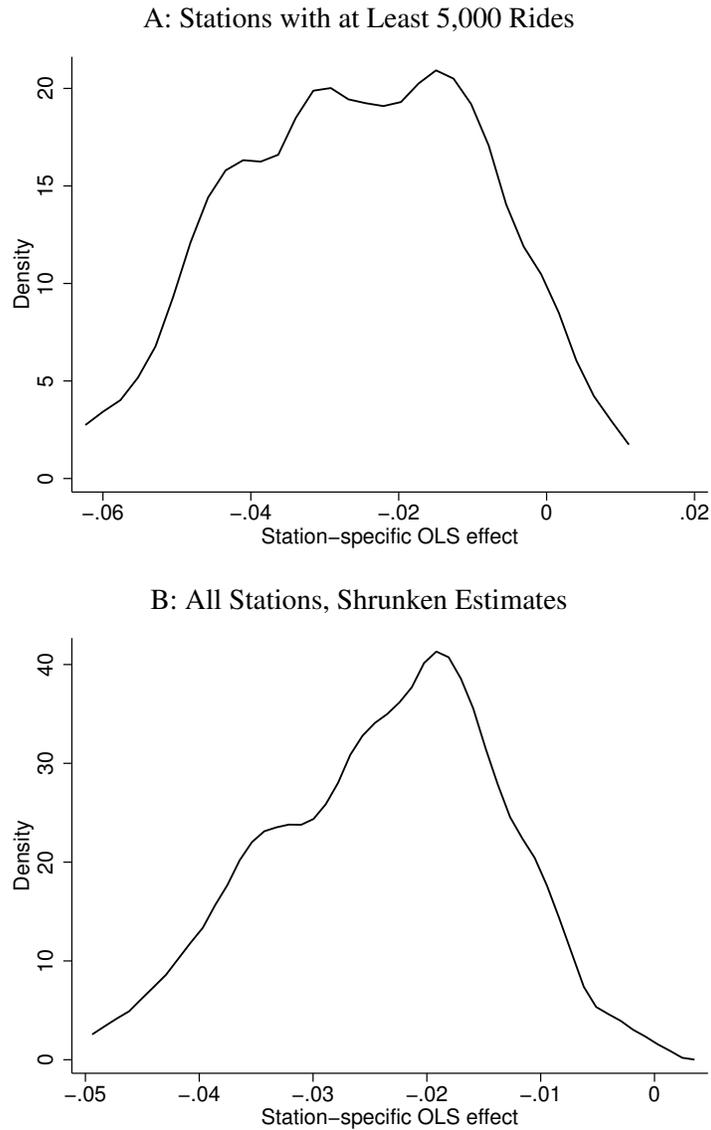
Note: Panel A of this figure shows the kernel density distribution of station-specific OLS estimates of the VA advantage, or $\hat{\beta}^s_{OLS}$ estimated from equation (A.27) for rides corresponding to each station. We include estimates from 32 stations with at least 5,000 rides, comprising a sample of $276,483$ rides. Panel B of this figure shows the kernel density distribution of empirical Bayes posteriors of the station-specific OLS estimates of the VA advantage. These posteriors are given by $\tilde{\beta}^s_{OLS}$ in equation (A.28). The figure displays posteriors from all 94 stations in our baseline sample in Appendix Table A.1, comprising $401,319$ rides.

Figure A.10: Complier Spending, Fixed Prices

A: Cumulative Spending

B: Conditional Spending Rate
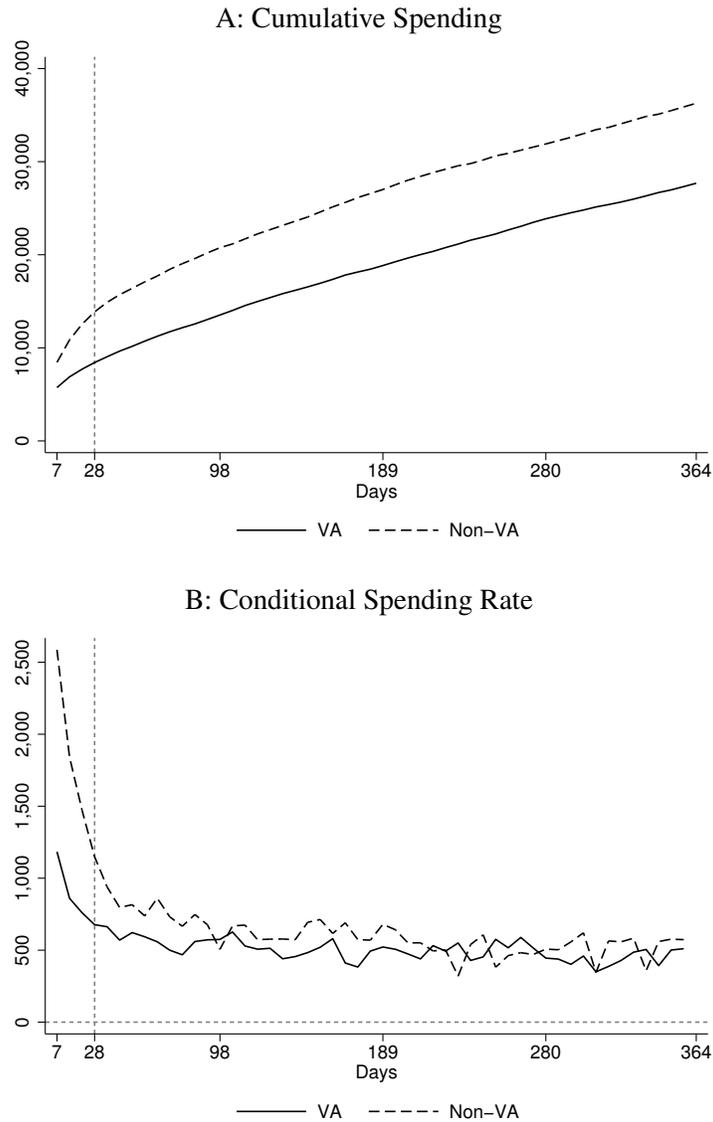
Note: This figure shows potential spending outcomes for ambulance compliers who arrive at a VA hospital and those who arrive at a non-VA hospital. Panel A shows cumulative spending per patient as a function of days from the ambulance ride. Panel B presents implied weekly spending rates for compliers, conditional on survival. Instead of actual spending by the government, insurers, and patients, as shown in Figure 5, this figure considers imputed spending with fixed prices based on methodology in Gottlieb et al. (2010) and Finkelstein, Gentzkow, and Williams (2016). Specifically, we impute spending for physician services based on Relative Value Units (RVUs) for service procedures with CPT codes, for other outpatient procedures based on average reimbursements for (non-CPT) HCPCS codes, and for inpatient stays based on Diagnosis-Related Group (DRG) weights. We scale prices by a constant so that imputed total Medicare spending equals actual total Medicare spending. The note for Figure 5 provides further details.

Figure A.11: VA Shares Within Top Reported Procedure Groups
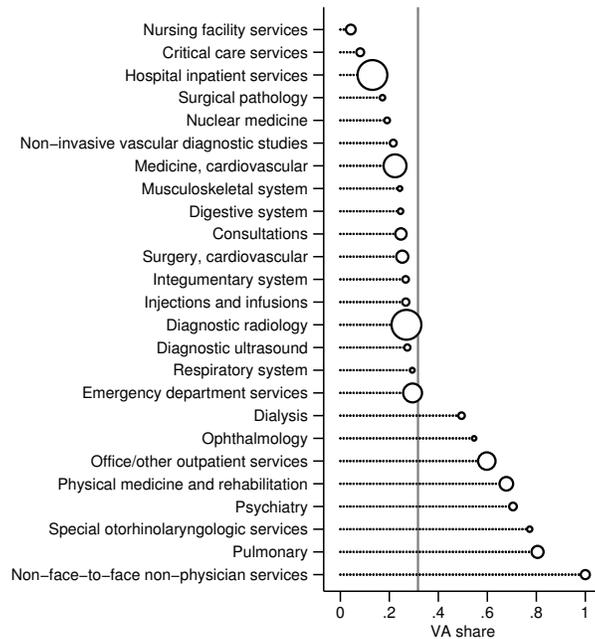
Note: This figure shows the VA share of reported utilization in each of the top 25 groups of procedures, defined by Current Procedural Terminology (CPT) codes. We form groups based on the list of 115 groups of Category I CPT codes at `https://en.wikipedia.org/wiki/Current_Procedural_Terminology`, which in turn is based on the organization of CPT codes by the American Medical Association (2017). We include utilization for any patient in our baseline sample in the 28 days following his or her ambulance ride. The area of each circle indicates the relative utilization volume of each CPT code group. For scale, the largest circle represents a service utilization of 4.159 times per ambulance ride; the smallest circle represents a service utilization of 0.095 times per ambulance ride. The top 25 groups of procedures represent 65.1% of the total utilization of all Category I CPT codes on the MPFS. The gray vertical line indicates the overall VA share of utilization of any CPT code on the MPFS.

Figure A.12: High-Complexity E/M Utilization in Non-VA vs. VA Providers
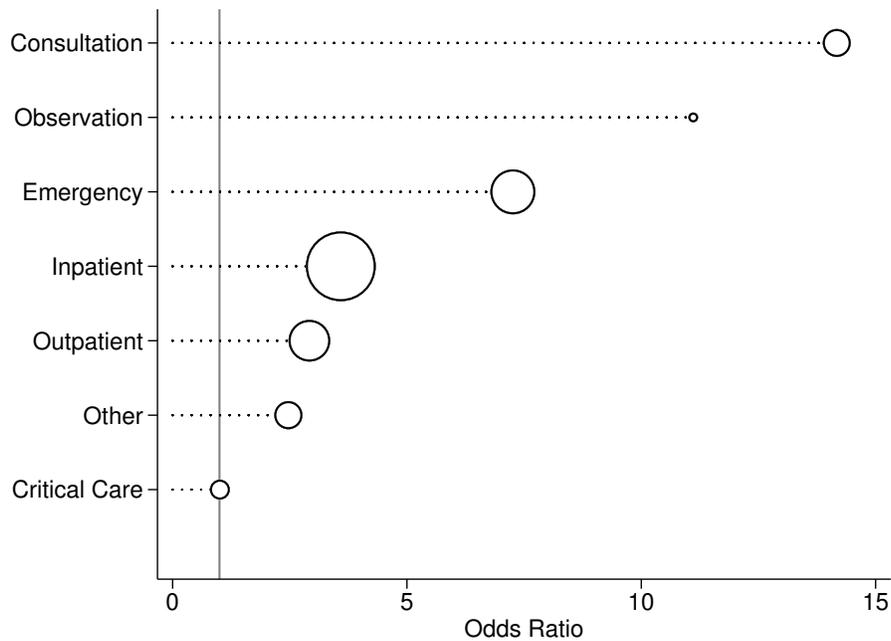
Note: This figure shows the odds ratio of high-complexity evaluation and management (E/M) Current Procedural Terminology (CPT) codes billed by non-VA vs. VA providers. We include utilization for any patient in our baseline sample in the 28 days following his or her ambulance ride. Within each type of E/M code, defined by the setting and the type of patient (e.g., "office or other outpatient visit for the evaluation and management of an established patient" for CPT codes 99211 to 99215), E/M codes are distinguished by "level" of complexity. We calculate the odds of highest to lowest complexity for non-VA providers and for VA providers and present the odds ratio on the *x*-axis. An odds ratio of one indicates that non-VA and VA providers are equally likely to bill the highest- vs. the lowest-complexity E/M code within the type. An odds ratio greater than one indicates that non-VA providers are more likely to do so. We present results within seven categories of E/M-code types defined by setting. The area of each circle is proportional to the total utilization volume in each of these categories.

Figure A.13: Sources of Prior Utilization

A: Number of Hospitals

B: Systems of Care and Ambulance Transport
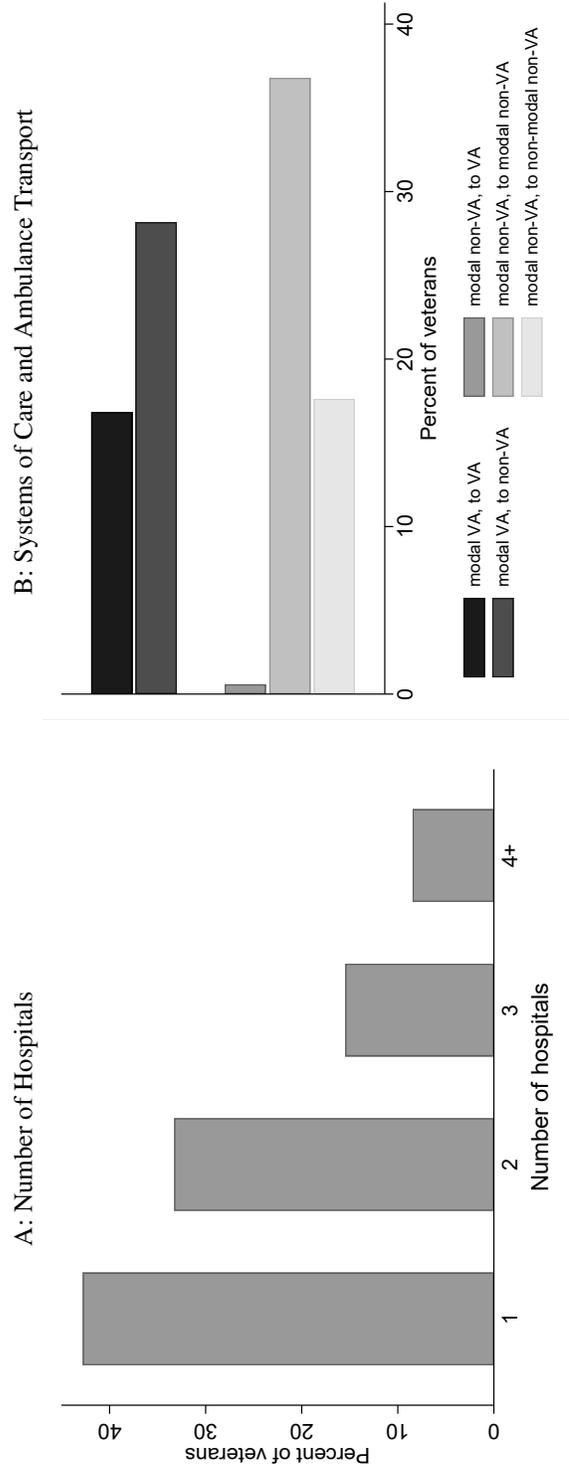
Note: This figure shows patterns of prior utilization and ambulance transport among a sample of patients who have some prior utilization either at the VA or affiliated with a non-VA hospital. Panel A shows the percentage of patients in this sample who utilize care associated with different numbers of 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 transported to their modal non-VA hospital or to another non-VA hospital. The sample selection for this group of patients is given in Appendix Table A.14.

Figure A.14: Modal Hospital First Stage, Balance, and Reduced Form

A: First Stage

coeff. = 0.767 (0.006)
N = 1,414,217, J = 5,716

B: Balance and Reduced Form

red. form coeff. = −0.005 (0.003)
balance coeff. = 0.000 (0.001)
N = 1,414,217, J = 5,716

● Mortality   ○ Predicted mortality
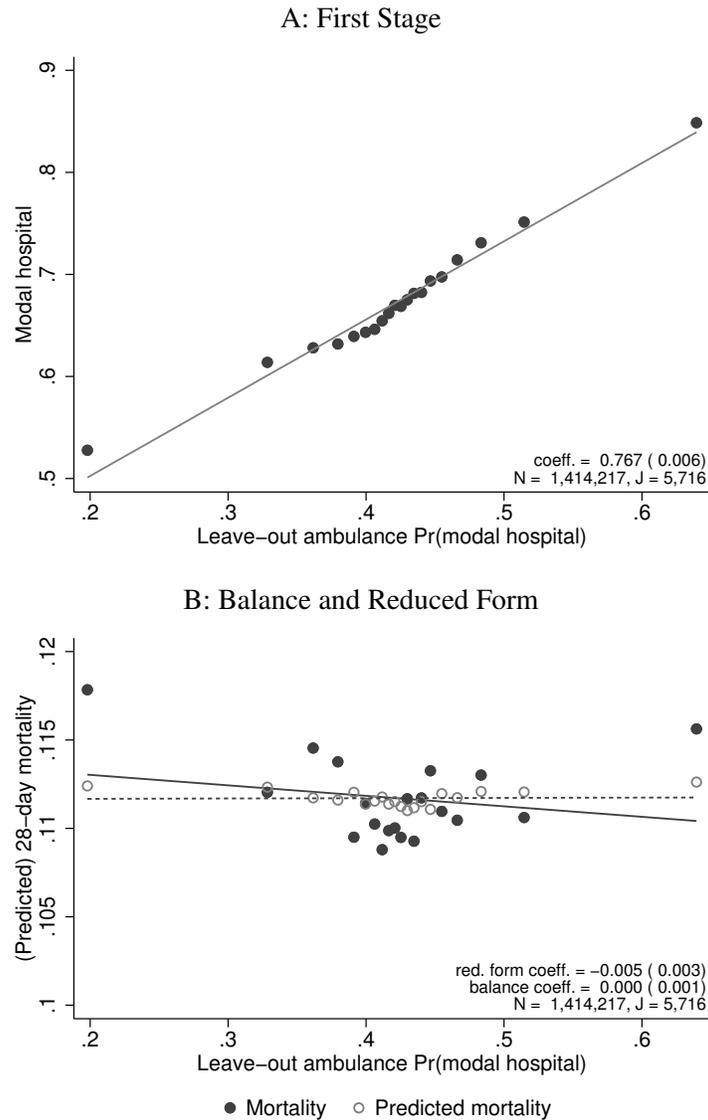
Note: Panel A shows a binned scatter plot of arrival at the veteran's modal hospital against the ambulance leave-out propensity to arrive at that hospital on the *x*-axis. The figure is a graphical representation of the first-stage regression in equation (A.30). Panel B shows binned scatter plots of 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 (A.31) and the corresponding balance regression replacing mortality with predicted mortality. The sample includes $1,414,217$ ambulance rides and $5,716$ 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 are given in Appendix Table A.14. Controls include patient zip code dummies, ALS/BLS dummies, source of the ambulance ride, time categories, and patient prior utilization, detailed in Appendix Table A.2.

Figure A.15: Modal Hospital Visual IV
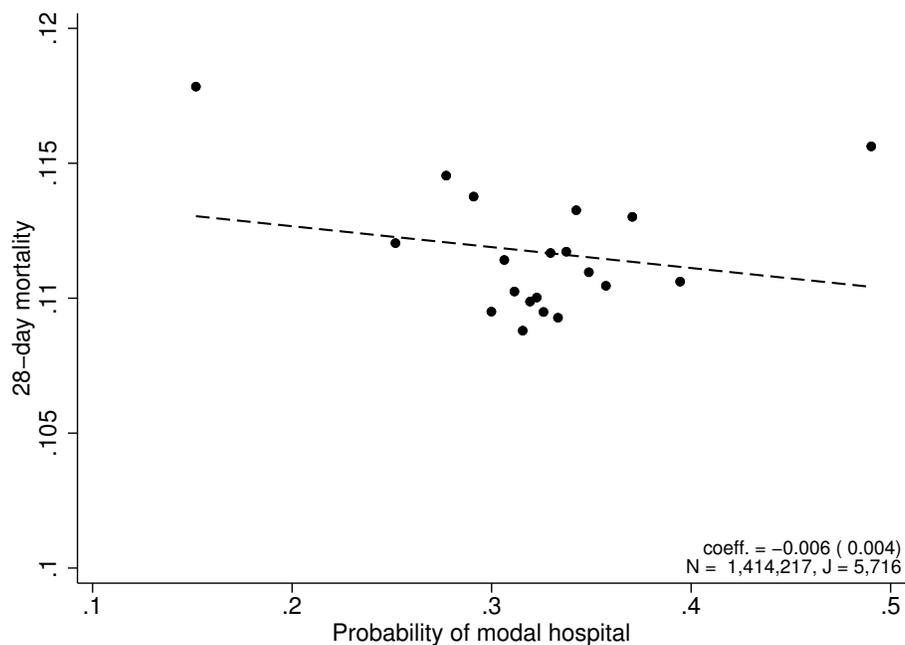
Note: This figure shows the visual IV plot corresponding to the IV regression of the effect of arrival at a patient's modal hospital on 28-day mortality. For each bin of the instrument, which is the ambulance leave-out propensity to arrive at the patient's modal hospital, we plot the mean 28-day mortality on the *y*-axis and the probability that the index patient arrives at his or her modal hospital on the *x*-axis. Modal hospital arrival predictions correspond to a first-stage regression in equation (A.30), and mortality predictions correspond to a reduced-form regression in equation (A.31). 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 $1,414,217$ observations and $5,716$ 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.14. Controls include patient zip code dummies, ALS/BLS dummies, source of the ambulance ride, time categories, and patient prior utilization, detailed in Appendix Table A.2.

A.34

Figure A.16: Modal Hospital OLS and IV Specifications
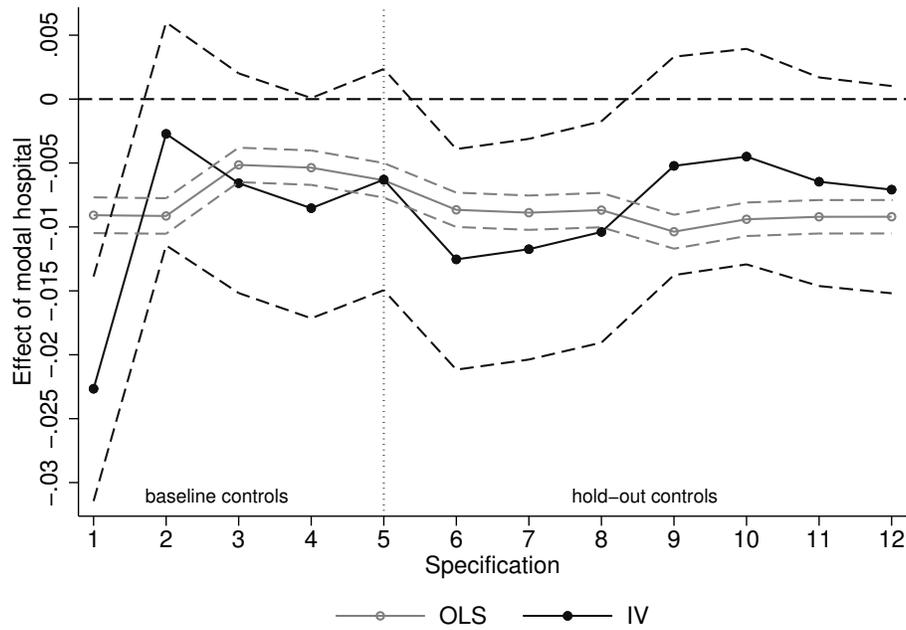
Note: This figure shows the effect of arrival at a patient's modal hospital on 28-day mortality estimated from OLS and IV specifications, with progressive sets of controls. Numbered incremental controls correspond to categories or subcategories of variables presented in order in Appendix Tables A.2 and A.3. Control sets are as follows: (1) zip code; (2) pickup source; (3) ambulance service; (4) time categories; (5) prior utilization; (6) demographics; (7) socioeconomic status, combat history, and eligibility; (8) extended prior utilization; (9) prior diagnoses; (10) 3-digit ambulance diagnosis codes; (11) co-rider baseline controls; and (12) co-rider hold-out controls. Estimates are shown along solid lines, while 95% confidence intervals are shown in dashed lines. All specifications control for hospital identities and use the sample of non-VA-only utilizers, given in Appendix Table A.14.

Figure A.17: Modal Hospital Combinations of Controls
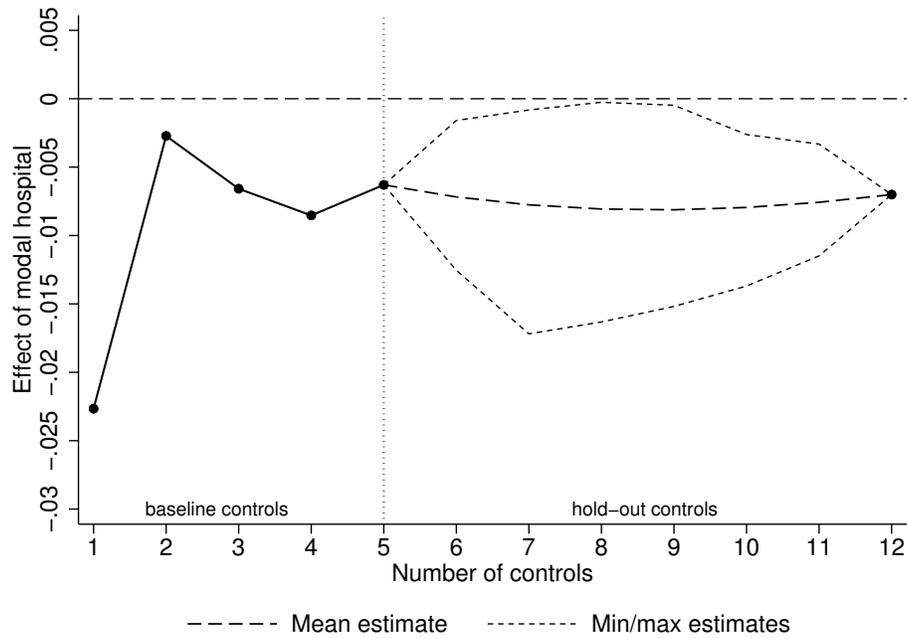
Note: This figure shows IV estimates of the effect of arrival at a patient's modal hospital on mortality on the y-axis, with first-stage and reduced-form equations (A.30) and (A.31), varying the number of controls included in the IV regression. Control variables are detailed in Appendix Tables A.2 and A.3. All specifications include the five baseline controls. Therefore, the figure represents $5 + (2^7 - 1) = 132$ specifications. For each number of controls n for $n > 5$, we consider "7 choose $n - 5$" 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.14.

Table A.1: Selection of Baseline Sample

Sample step	Description	Rides	Patients	Ambulance companies	Hospitals VA	Hospitals 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, non-missing ambulance diagnosis code, 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	183,692	126	7,815
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, or from VA New Orleans (destroyed in 2005 due to Hurricane Katrina).	8,828,997	2,862,557	180,320	124	7,743
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	117	3,070
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	99	1,576
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	98	1,403
6. Prior ride restriction	Drop rides for patients with a prior ride within 30 days.	401,319	188,299	1,217	98	1,385

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

Table A.2: Baseline Control Variables

Category	Subcategory	Variables
Location (1,633 indicators)	Zip code (1,630 indicators)	Zip code indicators (1,630 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 (173 indicators)		Day of the week (6 indicators); Month-year interactions (167 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 control 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. Numbers of non-collinear indicators are given in parentheses.

Table A.3: Hold-Out Control Variables

Category	Subcategory	Variables
Patient background (60 variables)	Demographics (30 indicators)	Age: 5-year age bins from 20-64 years, 2-year age bins from 65-100 years (26 indicators); Male gender; Race: indicators for white, Black, Hispanic, and Asian/other (omitted category)
	Socioeconomic status, combat history, and eligibility (22 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 (7 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 (641 indicators)		3-digit ambulance diagnosis (ICD-9) codes (641 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 Figure 2 and Appendix Figures A.3, A.16, and A.17. Numbers of non-collinear indicators or variables are given in parentheses.

Table A.4: Robustness of Exclusion Restriction

	(1)	(2)	(3)	(4)	(5)	(6)
	A: Dependent variable: 28-day mortality					
VA hospital	-0.053	-0.045	-0.039	-0.045	-0.045	-0.045
	(0.019)	(0.018)	(0.018)	(0.018)	(0.018)	(0.021)
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	-2,421	-2,805	-2,144	-2,549	-2,598	-2,257
	(897)	(876)	(809)	(814)	(820)	(974)
Outcome mean	12,265	12,265	12,265	12,265	12,265	12,265
Observations	401,319	401,319	401,319	401,319	401,319	401,319
Ambulance charges splines	Yes	No	No	No	No	Yes
Mileage splines	No	Yes	No	No	No	Yes
Out-of-sample mortality	No	No	Yes	No	No	Yes
Chosen non-VA hospitals						
Out-of-sample mortality	No	No	No	Yes	No	Yes
Out-of-sample 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), for "out-of-sample" outcomes by the ambulance company, and for non-VA hospitals chosen by the ambulance company ("out-of-sample" averages of mortality and spending for these non-VA hospitals). "Out-of-sample" refers to 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.14). 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 (0.038)	0.497 (0.022)
Age > 80	161,707	0.305	0.789 (0.041)	0.456 (0.022)
White	314,064	0.304	0.821 (0.037)	0.221 (0.016)
Non-white	87,176	0.426	0.992 (0.068)	0.596 (0.041)
Comorbidity count (high)	167,332	0.292	0.758 (0.041)	0.427 (0.019)
Comorbidity count (low)	233,987	0.358	0.938 (0.039)	0.553 (0.027)
Mental illness or substance abuse	188,961	0.354	0.931 (0.040)	0.514 (0.024)
No mental illness or substance abuse	212,358	0.309	0.815 (0.037)	0.456 (0.020)
VA visits in prior year (high)	183,087	0.508	1.038 (0.050)	0.710 (0.035)
VA visits in prior year (low)	218,232	0.181	0.718 (0.031)	0.284 (0.014)
Advanced Life Support	274,690	0.301	0.836 (0.036)	0.249 (0.018)
No Advanced Life Support	126,616	0.393	0.840 (0.048)	0.301 (0.032)
Predicted VA user (high)	200,659	0.543	1.113 (0.054)	0.865 (0.055)
Predicted VA user (low)	200,660	0.117	0.559 (0.030)	0.218 (0.011)
Predicted mortality (high)	200,659	0.328	0.835 (0.036)	0.368 (0.019)
Predicted mortality (low)	200,660	0.333	0.898 (0.046)	0.502 (0.024)
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	Instrument In-sample
Age ≤ 80	239,611	0.347	0.586 (0.021)	0.525 (0.020)
Age > 80	161,707	0.305	0.494 (0.023)	0.394 (0.022)
White	314,064	0.304	0.504 (0.019)	0.513 (0.020)
Non-white	87,176	0.426	0.676 (0.032)	0.440 (0.033)
Comorbidity count (high)	167,332	0.292	0.493 (0.020)	0.438 (0.021)
Comorbidity count (low)	233,987	0.358	0.583 (0.022)	0.504 (0.020)
Mental illness or substance abuse	188,961	0.354	0.592 (0.021)	0.518 (0.020)
No mental illness or substance abuse	212,358	0.309	0.501 (0.020)	0.426 (0.020)
VA visits in prior year (high)	183,087	0.508	0.691 (0.026)	0.572 (0.021)
VA visits in prior year (low)	218,232	0.181	0.421 (0.018)	0.445 (0.021)
Advanced Life Support	274,690	0.301	0.523 (0.020)	0.518 (0.021)
No Advanced Life Support	126,616	0.393	0.531 (0.025)	0.433 (0.024)
Predicted VA user (high)	200,659	0.543	0.743 (0.028)	0.619 (0.021)
Predicted VA user (low)	200,660	0.117	0.331 (0.016)	0.423 (0.027)
Predicted mortality (high)	200,659	0.328	0.513 (0.020)	0.458 (0.019)
Predicted mortality (low)	200,660	0.333	0.570 (0.023)	0.479 (0.021)
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) an 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 - 1.10]	(0.004)	[0.90 - 1.00]
Prior VA ED visit only	0.593	2.02	0.145	0.49
	(0.007)	[1.97 - 2.06]	(0.003)	[0.47 - 0.52]
Prior non-VA ED visit only	0.032	0.13	0.376	1.52
	(0.002)	[0.12 - 0.14]	(0.005)	[1.48 - 1.56]
Prior VA and non-VA ED visit	0.230	0.98	0.237	1.01
	(0.006)	[0.93 - 1.03]	(0.004)	[0.98 - 1.05]
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	0.103	1.07	0.100	1.03
	(0.002)	[1.03 - 1.10]	(0.001)	[1.01 - 1.05]

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 5. 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: Average Treatment Effect (ATE) 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)
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 linear specification of $J(u)$, corresponding to the regression in equation (A.23). Columns 2 to 5 present results from semiparametric specifications of $J_d(u)$, 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. We compute standard errors (shown in parentheses) for the ATE by bootstrap, blocking by zip codes. Appendix A.4 provides further details.

Table A.9: 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	-0.044	0.003	$10,651	$31,159
	(0.017)	(0.002)		
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 (A.26). 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.10: 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	-0.046	-0.002	28,082	53,849
	(0.016)	(0.002)		
Admissions	-0.046	-0.003	9,664	17,859
	(0.017)	(0.002)		
Total staffed beds	-0.046	-0.004	199	375
	(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 1,000 ED visits	-0.045	-0.001	0.30	0.75
	(0.017)	(0.002)		
Nurses per 1,000 patient-days	-0.046	0.006	4.13	6.58
	(0.016)	(0.002)		
Physicians per 1,000 patient-days	-0.045	0.002	4.36	10.79
	(0.017)	(0.002)		
Hospitalists per 1,000 patient-days	-0.045	0.003	0.12	0.39
	(0.017)	(0.002)		
Intensivists per 1,000 patient-days	-0.045	0.003	0.05	0.23
	(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 (A.26). 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.11 presents results for additional characteristics. Appendix A.5 provides further details on the hospital characteristics.

Table A.11: 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)		
Share of non-VA rides (max.)	-0.045	0.002	0.42	0.73
	(0.017)	(0.002)		

Note: This table presents regression results investigating heterogeneous treatment effects along binary indicators based on non-VA hospital characteristics associated with each zip code. For "Share of non-VA rides (max.)", we take the maximum non-VA hospital share of non-VA rides as the zip code characteristic. Hospital characteristics are averaged within each zip code for the remaining characteristics 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 statistics, denoted by $I_{x,i} = 0$ and $I_{x,i} = 1$, respectively. Regression results correspond to equation (A.26). 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.12: 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	8,625	23,111
	(0.017)	(0.002)		
Admissions	-0.044	-0.003	3,247	8,148
	(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 1,000 ED visits	-0.050	-0.001	0.19	1.21
	(0.022)	(0.003)		
Nurses per 1,000 patient-days	-0.045	0.003	3.80	8.60
	(0.017)	(0.002)		
Physicians per 1,000 patient-days	-0.050	-0.000	1.12	7.95
	(0.022)	(0.003)		
Hospitalists per 1,000 patient-days	-0.051	0.006	0.03	0.30
	(0.022)	(0.003)		
Intensivists per 1,000 patient-days	-0.050	0.001	0.00	0.15
	(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	7.13	7.96
	(0.017)	(0.003)		
Readmission rate	-0.045	-0.003	11.73	12.72
	(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 (A.26). 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.13: Heterogeneity by Difference Between VA and 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	-0.046	-0.000	-38,214	-8,950
	(0.017)	(0.002)		
Admissions	-0.045	-0.001	-12,217	-2,690
	(0.017)	(0.002)		
Total staffed beds	-0.046	-0.000	-152	236
	(0.017)	(0.002)		
Teaching hospital	-0.045	-0.001	-0.15	0.78
	(0.017)	(0.002)		
Trauma center	-0.052	-0.001	-0.86	-0.05
	(0.018)	(0.002)		
Advanced cardiac care	-0.052	-0.003	-0.34	0.34
	(0.018)	(0.002)		
Staffing				
ED staff per 1,000 ED visits	-0.051	-0.001	-0.00	0.00
	(0.022)	(0.003)		
Nurses per 1,000 patient-days	-0.046	-0.000	-0.00	0.00
	(0.017)	(0.002)		
Physicians per 1,000 patient-days	-0.051	-0.001	-0.01	0.00
	(0.022)	(0.003)		
Hospitalists per 1,000 patient-days	-0.051	-0.002	-0.00	0.00
	(0.022)	(0.002)		
Intensivists per 1,000 patient-days	-0.051	-0.003	-0.00	0.00
	(0.022)	(0.003)		
Spending and Outcomes				
Relative spending	-0.044	0.000	-0.01	0.30
	(0.017)	(0.002)		
Mortality rate	-0.046	0.005	-5.29	-3.97
	(0.017)	(0.002)		
Readmission rate	-0.045	-0.002	-6.76	-4.83
	(0.017)	(0.002)		

Note: This table presents regression results investigating heterogeneous treatment effects along the difference between VA and non-VA hospital characteristics associated with each zip code. Non-VA hospital characteristics are averaged within each zip code with weights proportional to the number of rides going to each non-VA hospital from the zip code. VA hospital characteristics are taken for the VA hospital with the largest share of rides from the zip code. Each difference is formed by subtracting the value for the (average) non-VA hospital characteristic from the value for the associated VA hospital characteristic. For each difference 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 (A.26). 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.14: 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 Appendix 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 Appendix 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 used to study descriptive utilization patterns in Appendix Figure A.13. Panel B shows selection steps for the sample of patients with only non-VA prior utilization, which we use in Section C to study the effect of receiving care at a modal non-VA hospital. The table lists the number of ambulance rides, patients, ambulance companies, and VA and non-VA hospitals at each step. Appendix Table A.15 shows average patient characteristics among observations at each step.

Table A.15: Characteristics of Non-VA-Only Sample

Restrictions	Dually eligible	Sample characteristics			
		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	$21,724	$21,763	$22,253	$22,800	$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 only	0.048	0.052	0.052	0.000	0.000
Prior non-VA ED visit only	0.607	0.606	0.602	0.797	0.752
Prior VA and non-VA ED visit	0.088	0.089	0.085	0.000	0.000
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 C to study the effect of receiving care at a modal non-VA hospital. Each step is detailed in Panel B of Appendix Table A.14.

Table A.16: 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 (A.30) and (A.31); 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.9. 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.14, Panel B.