

Online Appendix for “The Productivity of Professions: Evidence from the Emergency Department”

David Chan and Yiqun Chen

A.1 Diagnosis Coding: NPs versus Physicians

In this appendix, we examine whether NPs and physicians differ significantly in reporting three-digit ICD-10 diagnoses. All diagnoses in our data are coded in ICD-10 during our study period from January 2017 to January 2020. As OLS estimation is likely to be confounded by patient selection, we leverage IV regressions that instrument for whether a case is treated by an NP using the number of NPs on duty. Specifically, we first create indicators for each of the 836 different three-digit ICD-10 primary diagnosis categories in our data (including one for the missing category). Then, for each diagnosis indicator, we run a separate 2SLS regression as follows to estimate whether NPs and physicians are significantly different in reporting:

$$y_i = \delta \text{NP}_i + \mathbf{T}_i \boldsymbol{\eta} + \varepsilon_i, \quad (\text{A.1})$$

$$\text{NP}_i = \lambda Z_i + \mathbf{T}_i \boldsymbol{\zeta} + v_i, \quad (\text{A.2})$$

where, similar to Equations (1) and (2), NP_i indicates whether case i is treated by an NP and Z_i denotes the instrument (i.e., the number of NPs on duty between 8 a.m. and 6 p.m., our analysis time window, at the ED on the day case i visits). The vector $\mathbf{T}_i$ includes ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day indicators. The coefficient of interest is δ . As with the main specification, we cluster standard errors by provider.

Panel A of Appendix Figure A.14 plots the distribution of t -statistics for the estimated δ coefficients from the 836 separate regressions that use each three-digit diagnosis indicator as the outcome variable. The share of t -statistics indicating a p -value below or equal to 0.05 is only 0.07, close to the null hypothesis of no differential three-digit diagnosis coding between NPs and physicians (i.e., a share of 0.05 of t -statistics indicating a p -value less than or equal to 0.05). Both the Shapiro-Wilk normality test and the test for normality on the basis of skewness and kurtosis suggest that we cannot reject the null hypothesis that the t -statistics are normally distributed, at least at the 10 percent level. Panel B of Appendix Figure A.14 further plots t -statistics against the prevalence of the three-digit diagnosis among physicians, showing that NPs are not more likely to report diagnoses that are more (or less) common.¹

1. We measure the prevalence of the diagnosis as the share of cases with the diagnosis among cases treated by physicians on days without any NP, to restrict the potential influence of patient sorting between NPs and physicians.

The pattern of similar three-digit diagnosis coding between NPs and physicians could arise for the relatively straightforward cases that are compliers. Additional consults and diagnostics (Section III.C) may also help NPs reach the same three-digit diagnosis as physicians. It is also worth noting that VHA ED providers' reimbursements are independent of patient diagnoses, and NPs and physicians are unlikely to have differential financial incentives in diagnosis coding.

A.2 Characterizing Compliers and Never-Takers

This appendix describes the estimation of characteristics of compliers and never-takers. Following the approach developed by Abadie (2003), we characterize compliers by δ estimated through the 2SLS model specified in Equations (A.1) and (A.2), replacing the outcome variable y_i with $x_i \times \text{NP}_i$, i.e., the interaction between each patient characteristic x_i and the indicator for being treated by an NP. Results are discussed in Section III.D and shown in Column 2 of Appendix Table A.11.

To estimate the characteristics of never-takers, we follow a method by Dahl, Kostøl, and Mogstad (2014). We first collapse the data to the ED-day level. We then residualize the share of cases treated by NPs by ED-by-year, ED-by-month, and ED-by-day-of-the-week indicators. We define never-takers as cases treated by physicians in ED-day cells with the residual share of cases treated by NPs at least as high as the 90th percentile of ED-day cells with at least one case treated by NPs. There are no always-takers in our setting since patients cannot be assigned to NPs on days when no NPs are on duty.

Column 4 of Appendix Table A.11 reports the average characteristics of never-takers. For each characteristic, we compute the mean of never-takers as well as the ratio between the mean and the overall sample mean. We estimate standard errors for the means by bootstrap, blocking observations by provider with 500 replications. In line with the notion that NPs treat healthier cases than physicians do, Appendix Table A.11 shows that never-takers are the riskiest, followed by the overall sample, and finally, compliers. For example, the total number of Elixhauser comorbidities among never-takers, the overall sample, and compliers is, respectively, 4.0, 3.6, and 3.3; the average predicted 30-day mortality among these three types of cases is, respectively, 1.7, 1.2, and 0.9 percentage points.

A.3 Provider Value-Added and Practice-Style Measures

This appendix describes our construction of measures of provider value-added and practice styles, used to examine the exclusion restriction in Section III.E. We consider physician value-added as a measure of risk-adjusted mortality outcomes and form this measure using leave-out data. Specifically, for physician p on day d , we measure

$$A_{p,d} = \frac{\sum_{i \in I_p} \mathbf{1}(d(i) \neq d, Z_i = 0) \tilde{Y}_i}{\sum_{i \in I_p} \mathbf{1}(d(i) \neq d, Z_i = 0)}, \quad (\text{A.3})$$

where $\tilde{Y}_i$ is risk-adjusted 30-day mortality, or the difference between the patient's actual and predicted 30-day

mortality. To address possible finite-sample bias, we exclude patient cases on day d .² We also leave out cases visiting on days with any NPs on duty, to mitigate the concern of patient sorting between NPs and physicians.

Still, since cases are not experimentally assigned among physicians, $A_{p,d}$ may reflect both a physician's effect on patient outcomes and systematic patient-physician sorting under imperfect risk adjustment. As one way to assess the degree of such potential biases, we investigate the robustness of physician value-added estimates to patients' predicted mortality constructed using different risk adjusters, analogous to the test of student sorting biases in the teacher value-added literature (e.g., Chetty, Friedman, and Rockoff 2014). If patient sorting is important, the estimated physician value-added will change meaningfully with the addition of risk adjusters. Otherwise, our estimates should remain stable. Appendix Figure A.15 shows that physician value-added estimates are stable regardless of patient risk adjusters. We compare physician value-added measures constructed using (i) the most parsimonious set of risk adjusters that includes only age-bin and three-digit primary diagnosis indicators, (ii) the less parsimonious set that adds non-age demographics (gender, race, and marital status), and (iii) the set that further adds dummies for 31 Elixhauser comorbidities, with the baseline physician value-added constructed using the full set of patient covariates (i.e., demographics, Elixhauser comorbidities, prior health care use, vital signs, and three-digit diagnosis indicators). The correlations between measures (i)–(iii) and the baseline measure are all above 0.99. Note that these risk adjusters are important predictors of patient 30-day mortality: They alone explain 7 percent of the variation in 30-day mortality, with an F -statistic of 88 for joint significance.

We consider physician practice styles as measures of physician-chosen inputs to care. Specifically, we define practice style measures by Equation (A.3), but instead set $\tilde{Y}_i$ as the difference between the patient's actual and predicted log length of stay or log cost of the ED visit. As with value-added, we show the robustness of practice style estimates to different patient risk adjusters in Appendix Figure A.15.

We construct similar measures of value-added and practice style for NPs. As with physicians, we show the robustness of these estimates to different patient risk adjusters. Appendix Figure A.15 shows that NP value-added and practice-style estimates are highly stable among those constructed using (i) the most parsimonious set of risk adjusters that includes only age-bin and three-digit primary diagnosis indicators, (ii) the less parsimonious set that adds non-age demographics (gender, race, and marital status), (iii) the set that additionally includes dummies for 31 Elixhauser comorbidities, and (iv) the full set that further adds detailed controls for vital signs and prior health care use.

A.4 Augmenting Provider Supply with NPs

In this appendix, we ask, holding fixed the number of cases arriving and the number of physicians on duty, how additional NPs may affect patient wait time and downstream outcomes. The empirical specification takes the following form:

2. Specifically, there may be ED-day level shocks that are correlated with both the number of NPs on duty and the set of patients treated by a specific physician; these shocks can be influential in estimations with a finite sample.

$$y_i = \sum_{n=0}^N \delta_n \times \mathbf{1}(Z_i = n) + N_i^c \gamma_1 + N_i^p \gamma_2 + \mathbf{T}_i \eta + \mathbf{X}_i \beta + \varepsilon_i, \quad (\text{A.4})$$

where $\mathbf{1}(Z_i = n)$ is an indicator for $n \in \{0, \dots, 4\}$ NPs being on duty at the ED on the day case i visits,³ and N_i^c and N_i^p are, respectively, the number of cases arriving and the number of physicians on duty at the ED on the day case i visits. We apply Equation (A.4) to the following outcomes: (i) wait time (the time between arrival at the ED and assignment to a provider); (ii) length of stay (the time between assignment to a provider and discharge from the ED); and (iii) log total spending of the ED visit, which is the sum of the cost of ED care, hospital admission in the ED visit, and 30-day preventable hospitalization.

Appendix Figure A.16 presents the estimated trade-off based on Equation (A.4). Panel A shows the trade-off between wait time and length of stay. Each dot plots a pair of δ_n with the estimated δ_n for wait time on the y-axis and the estimated δ_n for length of stay on the x-axis. Panel B similarly plots the trade-off between wait time and total spending per case (we convert the estimated δ_n for log total spending to spending per case).

As implied in Figure A.16, decreasing wait time by 30 minutes per case by hiring additional NPs increases length of stay by 10 minutes per case (Panel A) and increases total medical spending by \$436 per case, not including the cost of additional NP wages (Panel B; inflation-adjusted to 2020 dollars).⁴ Including the wage costs of additional NPs brings this figure to about \$480 per case.⁵ That is, roughly nine-tenths of the additional spending to reduce wait time by hiring NPs comes from the lower productivity of NPs, while only one-tenth comes from additional NP wage costs.

However, the cost of reducing wait time by hiring additional NPs declines with case complexity. As implied in Appendix Figure A.17, among the lowest-complexity quartile cases, reducing wait time by 30 minutes per case by hiring additional NPs increases total medical spending by about \$280 per case (not including the wage cost of additional NPs). Though reducing wait time by 30 minutes increases length of stay by a slightly greater magnitude in the lowest-complexity quartile cases than in the average case (12 versus 10 minutes per case), the difference is not statistically significant.

A.5 Distribution of Provider Effects on Total Spending

In this appendix, we estimate the distribution of provider effects on log total spending associated with the ED visit. We start by identifying provider effects using a just-identified IV model. Next, we estimate the variance of provider effects, using a split-sample approach to account for the bias due to sampling error in the estimated provider effects. We then apply an empirical Bayes deconvolution method, adapted by Kline, Rose, and Walters (2022) from Efron (2016), to recover the underlying population distribution of provider effects.

3. The maximum n is 6, but only a small share (0.8 percent) of ED-day cells has 5 or 6 NPs on duty (Appendix Figure A.1). We thus group $n = 5$ or 6 with $n = 4$.

4. Figure A.16 shows that a 33-minute decline in wait time increases length of stay by 11 minutes and total spending by \$472 per case (comparing the first and last dots in each panel). We rescale the trade-off to a 30-minute decline in wait time for easier interpretation.

5. For this wage cost estimation, we divide the yearly NP wage by the average number of cases per ED-year and then multiply this figure by the number of additional NPs required.

We present the receiver operating characteristic (ROC) curves of provider productivity, case assignments, and wage payments as characteristics that may distinguish NPs and physicians. Finally, we assess the extent to which provider productivity relates to cases assigned and wages paid across individual providers.

A.5.1 Estimating Provider Effects

We generate a measure of total spending associated with each ED visit, which is the sum of the three main components of costs examined in Section III: ED costs, hospital admission, and 30-day preventable hospitalization (we multiply the latter two components by the average cost of a hospital stay, \$19,220).

We then estimate provider effects on total spending associated with the ED visit. To mitigate the impact of extreme values, we take the log of the medical expenditures. To account for the possibility that the treating provider is endogenous, we instrument for indicators for treating providers with indicators for on-duty providers in the ED-day cell of the patient's visit. The empirical specification is a just-identified 2SLS model as follows:

$$y_i = \sum_j \theta_j \mathbf{1}_{\{j(i)=j\}} + \mathbf{T}_i \eta + \mathbf{X}_i \beta + \varepsilon_i, \quad (\text{A.5})$$

$$\mathbf{1}_{\{j(i)=j\}} = \sum_j \lambda_j \mathbf{1}_{\{j \in \mathcal{I}_i\}} + \mathbf{T}_i \zeta + \mathbf{X}_i \gamma + v_i, \quad (\text{A.6})$$

where $\mathbf{1}_{\{j(i)=j\}}$ is an indicator for whether case i is treated by provider j , and $\mathbf{1}_{\{j \in \mathcal{I}_i\}}$ is an indicator for whether provider j is on duty in the ED-day cell of case i 's visit with $\mathcal{I}_i$ representing the set of providers on duty in the ED-day cell. The coefficients of interest are θ_j , representing provider effects. Since θ_j is only identified relative to one another for providers within the same ED, we make the natural normalization that the case-weighted mean of θ_j is zero within each ED, using linear constraints in the 2SLS estimation to yield valid standard errors.

The F -statistics for the joint significance of on-duty provider indicators in the first-stage regressions, i.e., Equation (A.6), have shares of 0.99 and 0.68 above 10 and 100, respectively, suggesting that provider availability is strongly predictive of the treating provider.⁶ Appendix Figure A.18 shows that patient characteristics are well balanced across the average characteristics (age, gender, and practice style) of on-duty providers, conditional on the baseline controls $\mathbf{T}_i$.⁷ In addition, the F -statistics for the joint significance of on-duty provider indicators from regressions of patient predicted log length of stay and predicted log cost of the ED visit on on-duty provider indicators conditioning on $\mathbf{T}_i$ are 2.4 and 2.1, respectively. These are much smaller than the corresponding F -statistics using the actual log length of stay and log cost of the ED visit as the outcome—which are 10.3 and 9.9, respectively. These results support the assumption that the set of on-duty providers is conditionally independent of the set of patients arriving, thereby validating our instruments.

6. Since $\mathbf{1}_{\{j(i)=j\}}$ is always zero for patients outside of the ED in which a provider practices, we report F -statistics from the first-stage regression in Equation (A.6) using observations in each ED separately.

7. We compute case-weighted average characteristics of on-duty providers, with the index case left out. For the practice style examined in this balance test, to deal with the concern about patient sorting between NPs and physicians, we use provider effects on patient log length of stay and log cost of the visit estimated by the 2SLS model in Equations (A.5) and (A.6).

A.5.2 Estimating Variance of Provider Effects

We estimate the variance of provider effects, within each professional class, on log total spending associated with the ED visit. The estimated provider effects $\hat{\theta}_j$ from Appendix A.5.1 yield a case-weighted variance of 0.054 for NPs, and 0.064 for physicians (see Appendix Table A.22).⁸ However, these estimates are upward-biased, due to sampling error resulting from the fact that provider effects are estimated on a finite sample. To account for such biases, we leverage a split-sample approach similar to that employed in earlier studies (e.g., Silver 2021). Specifically, we randomly split each provider's patients within each day into two approximately equal-sized partitions. We then estimate the 2SLS model in Equations (A.5) and (A.6) using each partition separately, yielding two fixed effect estimates for each provider $\hat{\theta}_{j,a}$ and $\hat{\theta}_{j,b}$. Suppressing the j subscript for simplicity, we have

$$\hat{\theta}_q = \theta + e_q, q \in \{a, b\},$$

where q indicates partitions, and e_q is partition-specific sampling error, such that $\text{Cov}(\theta, e_b) = \text{Cov}(e_a, \theta) = 0$. The random split of patients within each provider-day makes plausible the assumption that e_a and e_b are uncorrelated, i.e., $\text{Cov}(e_a, e_b) = 0$. We can therefore compute the variance of provider effects as the covariance of $\hat{\theta}_a$ and $\hat{\theta}_b$:

$$\begin{aligned} \text{Cov}(\hat{\theta}_a, \hat{\theta}_b) &= \text{Cov}(\theta + e_a, \theta + e_b) \\ &= \text{Cov}(\theta, \theta) + \text{Cov}(\theta, e_b) + \text{Cov}(e_a, \theta) + \text{Cov}(e_a, e_b) \\ &= \text{Var}(\theta). \end{aligned}$$

We perform this calculation separately for NPs and physicians.

Appendix Table A.22 reports the case-weighted variance of provider effects from the split-sample approach. The variance for physicians is estimated at 0.045, about 70 percent of the calculated variance without accounting for the bias due to sampling error. The variance from the split-sample approach suggests that, on average, a one-standard-deviation costlier physician raises medical spending by 21 percent per case. For NPs, the split-sample variance estimate is 0.048, suggesting that a one-standard-deviation costlier NP raises spending by 22 percent per case.

A.5.3 The Population Distribution of Provider Effects

We now estimate the distribution of provider effects by applying a non-parametric empirical Bayes deconvolution approach adapted by Kline, Rose, and Walters (2022) from Efron (2016). This approach extracts a flexible estimate of the distribution of population provider effects using provider effects $\hat{\theta}_j$ and their standard errors s_j estimated in Equations (A.5) and (A.6). Assuming provider z-scores $z_j = \hat{\theta}_j/s_j$ are distributed as

$$z_j|c_j \sim \mathcal{N}(c_j, 1), c_j \sim G_c,$$

where $c_j = \theta_j/s_j$ (i.e., the population analogue of z_j), the procedure first applies the Efron (2016)

8. Since provider effects are normalized to have a mean of 0 in each ED, the variance is interpretable as the within-ED variance of provider effects.

deconvolution procedure to yield a distribution of provider z -scores $\hat{G}_c$ with density function $\hat{g}_c(\cdot)$. The Efron (2016) procedure estimates $\hat{G}_c$ by maximum likelihood of parameters that represent coefficients on a set of splines, with a regularization parameter to tamp down excursions from a flat prior.

Next, assuming that s_j is independent of c_j , we can derive an estimate of the distribution of provider effects $\hat{F}$, with density function $\hat{f}(\cdot)$ for each value θ :

$$\hat{f}(\theta) = \frac{1}{J} \sum_{j=1}^J \frac{1}{s_j} \hat{g}_c(\theta/s_j). \quad (\text{A.7})$$

Following Kline, Rose, and Walters (2022), we assess the independence of z_j and s_j by reporting regressions of z_j on s_j . To account for possible correlated estimation errors in z_j and s_j , we also present split-sample versions of these regressions that randomly split cases for each provider into two approximately equal-sized partitions and regress z -scores from one partition on standard errors from the other partition. The results are reported in Appendix Table A.23, which show no significant relationship between z_j and s_j , suggesting that independence between z -scores and standard errors is plausible.

We apply the empirical Bayes deconvolution estimator to NPs and physicians separately.⁹ As in Kline, Rose, and Walters (2022), we calibrate the regularization parameter in the maximum likelihood estimation so that the variance of the deconvolved distribution of provider effects matches the corresponding split-sample variance estimates reported in Appendix Table A.22. We demean both the physician and NP distributions to have a mean of zero and then shift the distribution of NPs to the right by 0.067, where 0.067 is the 2SLS estimate of the NP effect on the log total spending associated with the ED visit obtained by Equations (1) and (2). Panel A of Figure 8 displays the deconvolved density of provider effects for NPs and physicians. For interpretation, we rescale provider effects on log total spending per case on the x -axis into annual spending, as follows: (i) estimating the empirical distribution of log total spending per case, where the log spending is residualized with respect to ED and provider fixed effects, with means added back; (ii) for each provider effect from the deconvolution, adding the provider effect to the spending distribution in (i); (iii) exponentiating the individual values of the distribution in (ii) and integrating over the distribution to calculate the mean total spending per case for each provider effect; and (iv) multiplying the mean in (iii) for each provider effect by the average number of cases a provider sees per year observed in the data.

Finally, using the deconvolved density of NP and physician effects on log total spending per ED case, we estimate the probability that a randomly drawn NP is costlier than a randomly drawn physician by

$$\Pr(\theta_j > \theta_{j'} | j \in \mathcal{J}_{NP}, j' \in \mathcal{J}_{MD}) = \int_0^1 \hat{F}_{MD}(\theta) d\hat{F}_{NP}(\theta), \quad (\text{A.8})$$

where $\hat{F}_{MD}(x)$ and $\hat{F}_{NP}(x)$ are the deconvolved cumulative distribution functions of physician effects and NP effects, respectively, and $\mathcal{J}_{MD}$ and $\mathcal{J}_{NP}$ are the sets of providers who are physicians and NPs, respectively. We find the probability that a randomly drawn NP is costlier than a randomly drawn physician, in terms of the total spending associated with the ED visit defined above, is 62 percent. Put differently, the probability that

9. To restrict the inclusion of noisy provider effects, our deconvolution excludes providers with fewer than 150 cases. We set the support of provider effects to $[\theta^5 - SD, \theta^{95} + SD]$, where θ^5 , θ^{95} , and SD are, respectively, the 5th percentile, 95th percentile, and standard deviation of estimated NP and physician effects for the NP and physician deconvolution, respectively.

a randomly drawn NP is less costly than a randomly drawn physician is as high as 38 percent. This statistic remains large when we adjust the deconvolved productivity distributions to account for possible differences in treatment effects between the overall population and compliers: When we assume the average treatment effect is as large as twice the LATE estimate (higher than the LATE among the highest-complexity quartile patients), the probability that NPs are less costly remains large at 28 percent.

A.5.4 ROC Curve Representation

The probability in Equation (A.8) is equivalent to the area under the curve (AUC) of an ROC curve. The ROC curve displays the performance of an exercise classifying providers by a certain characteristic. In the case of provider effects on log total spending, the AUC of 0.62 indicates relatively poor performance in classifying providers as NPs versus physicians, depending on their (true) provider effects from their respective deconvolved distribution.

We construct ROC curves based on respective provider characteristics of productivity, the riskiness of assigned patients (measured by predicted 30-day mortality), and wages, where we consider physicians as the “positive” class and NPs as the “negative” class. For each characteristic of productivity, assigned patient predicted mortality, and wages, a provider with a higher value of the characteristic is more likely to be a physician (i.e., in the positive class). We define productivity as the additive inverse of the provider effect on log total spending per ED visit: $\mu_j = -\theta_j$. For a given characteristic x , we plot the ROC curve with $1 - \hat{F}_{MD}^x$ (i.e., the true positive rate) on the y-axis and $1 - \hat{F}_{NP}^x$ (i.e., the false positive rate) on the x-axis, where $\hat{F}_{MD}^x$ and $\hat{F}_{NP}^x$ are the empirical cumulative distribution functions of x among physicians and NPs, respectively. For productivity, we use the deconvolved distributions previously described in Appendix A.5.3, noting that $\hat{F}_{MD}^\mu = 1 - \hat{F}_{MD}^\theta$ and $\hat{F}_{NP}^\mu = 1 - \hat{F}_{NP}^\theta$. Since we observe actual patient assignments and provider wages, we do not estimate the risks of assigned patients or provider wage payments in a regression framework, nor do we apply deconvolution to obtain a population prior distribution. Rather, we apply the empirical cumulative distribution functions of the average predicted 30-day mortality for patients assigned to each provider and each provider’s annualized full-time equivalent (“yearly”) wage (inflation-adjusted to 2020 dollars).

We show the ROC curves in Appendix Figure A.12. As mentioned above, the AUC based on productivity is 0.62. The AUCs based on assigned patient risks and wage payments are 0.75 and 0.99, respectively. Taken together, a provider’s assigned patient risks are qualitatively more predictive of professional class than is the provider’s productivity; the provider’s yearly wages are extremely predictive of professional class.

A.5.5 Case Assignments and Wage Payments versus Productivity

Separately for NPs and physicians, we assess the extent to which case assignments and wages relate to individual provider productivity. As before, we measure productivity as provider effects on log total spending per ED visit, with higher provider effects indicating lower productivity. The empirical specification takes the following form:

$$y_j = \alpha \tilde{\theta}_j + \mathbf{L}_j \pi + \varepsilon_j, \quad (\text{A.9})$$

where y_j is the average risk (measured by predicted 30-day mortality) of cases assigned to provider j or yearly wage of provider j (inflation-adjusted to 2020 dollars). $\mathbf{L}_j$ is a vector of ED indicators since provider effects are only identified relative to one another within EDs. Since provider effects $\hat{\theta}_j$ are estimated with noise, we calculate empirical Bayes posteriors of provider effects as in Kline, Rose, and Walters (2022):

$$\tilde{\theta}_j = s_j \times \frac{\int x \varphi(z_j - x) \hat{g}_c(x) dx}{\int \varphi(z_j - x) \hat{g}_c(x) dx}, \quad (\text{A.10})$$

where φ denotes the standard normal density.

A.6 ED-Specific NP Effects

In this appendix, we estimate heterogeneity in the ED-specific NP effect. In separate 2SLS regressions for each ED ℓ , we estimate the NP effect using only cases at that ED:

$$\begin{aligned} y_i &= \delta_\ell \text{NP}_i + \mathbf{t}_i \eta_\ell + \mathbf{X}_i \beta_\ell + \varepsilon_i, \\ \text{NP}_i &= \lambda_\ell Z_i + \mathbf{t}_i \zeta_\ell + \mathbf{X}_i \gamma_\ell + v_i, \end{aligned}$$

where $\mathbf{t}_i$ is a vector of indicators for patient arrival year, month, day of the week, and hour of the day.

In Appendix Figure A.11, we plot the distribution of $\hat{\delta}_\ell$ for all EDs in our sample. We also plot the empirical Bayes posterior mean $\tilde{\delta}_\ell$ for each ED, calculated as

$$\tilde{\delta}_\ell = w_\ell \hat{\delta}_\ell + (1 - w_\ell) \hat{\delta}. \quad (\text{A.11})$$

The shrinkage factor is given by $w_\ell = \frac{\hat{\pi}^2}{s_\ell^2 + \hat{\pi}^2}$, where $\hat{\pi}^2$ and s_ℓ^2 are, respectively, the variance of the prior distribution of $\hat{\delta}_\ell$ and the variance of the sampling error for each $\hat{\delta}_\ell$. We calculate s_ℓ^2 as the square of the standard error of $\hat{\delta}_\ell$. We calculate $\hat{\pi}^2$ as the difference between the case-weighted variance of $\hat{\delta}_\ell$ and the case-weighted mean of s_ℓ^2 . Finally, $\hat{\delta}$ is the case-weighted mean of $\hat{\delta}_\ell$.

The gray bins in Appendix Figure A.11 plot the empirical Bayes posterior $\tilde{\delta}_\ell$ for each ED in our sample.¹⁰ The distribution of posteriors is more compressed than that of the raw estimates of ED-specific effects, reflecting shrinkage due to sampling error in the raw estimates. The results show a fair amount of heterogeneity. Nonetheless, most EDs exhibit positive effects of NPs on raising patient length of stay, cost of the ED visit, and 30-day preventable hospitalization rate.

10. The figure reports results for all EDs in our main analysis sample for log length of stay and log cost (in total 44 such EDs). For 30-day preventable hospitalization, since it is relatively uncommon (occurs in less than 2 percent of the sample), the estimates are relatively imprecise when using observations from a specific ED; we thus include only EDs with at least 25,000 cases in the analysis sample (in total 20 such EDs).

A.7 Measuring Number of On-Duty Physicians

In this appendix, we describe the measure of on-duty physicians. If patients arriving during the analysis time window are riskier, providers on duty during that period may be more occupied, raising the likelihood that patients arriving late in the window are assigned to physicians starting shifts after 6 p.m., thereby increasing the observed number of physicians treating patients arriving during the analysis time window. As shown in Columns 1-2 of Appendix Table A.24, we find a positive correlation between the health risks of patients arriving during the analysis time window and the observed number of physicians treating these patients. Note, however, that the measured number of on-duty NPs is unlikely to be subject to the same endogeneity concerns as the measured number of on-duty physicians, since NPs hardly work outside of our analysis time window (only one percent of ED-day cells in our analyzed EDs and months have NPs treating patients arriving outside of 8 a.m.–6 p.m.). Further, as shown in Figure 2 and Appendix Figure A.4, patient characteristics are remarkably balanced across the number of NPs on duty.

To deal with the endogeneity in the observed number of on-duty physicians, we consider a measure excluding physicians who treat patients arriving at or after 10 p.m., to drop physicians starting shifts after 6 p.m. (Results are similar when using alternative cutoffs of 9 or 11 p.m.) As shown in Columns 3-4 of Appendix Table A.24, although the relationship between patient health risks and the number of on-duty physicians based on this adjusted physician measure is weaker than that based on the raw measure in Columns 1-2, there remains a statistically significant relationship between patients' number of comorbidities and the measured number of physicians.

To obtain a more quasi-experimental measure of on-duty physicians, we count them over a 24-hour window. That is, we include all physicians who treat patients arriving at any time during the day, rather than restricting to those treating patients arriving between 8 a.m. and 6 p.m. This measure would still be correlated with the total physician staffing during our analysis time window, and its broader time window can mitigate concerns about the relationship between patient health risks and the identified number of on-duty physicians. As shown in Columns 5-6 of Appendix Table A.24, patient health risks are balanced across values of this 24-hour physician count. Comparing the three measures of on-duty physicians described above, Appendix Table A.25 shows that, when patient risks are more balanced against a measure of physician availability, IV estimates using the NP share of on-duty providers as the instrument become closer to our main IV estimates in Table 2. In our robustness checks in Appendix Figure A.10 and Panel D of Appendix Table A.18, we count on-duty physicians using the 24-hour physician count, the most balanced measure with respect to patient risks.

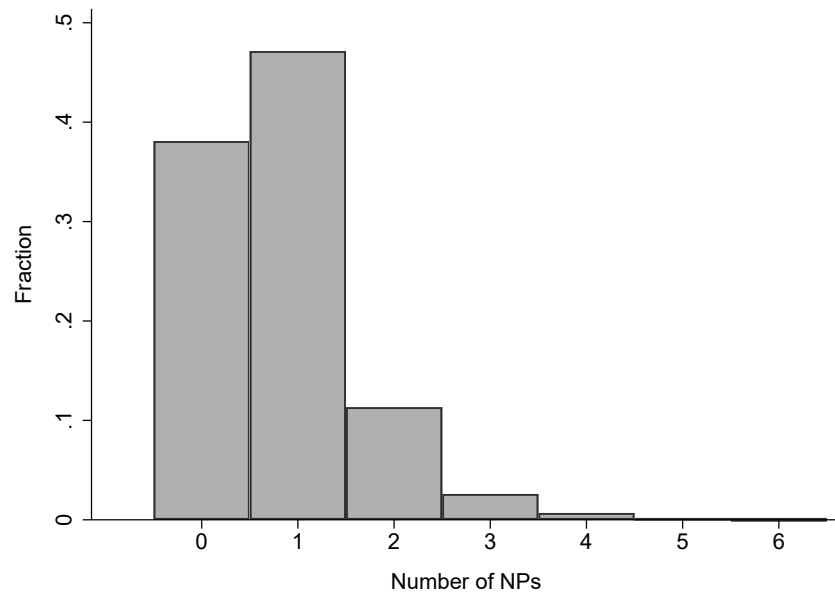
In Appendix Table A.26, we report results using the NP share of physician equivalents, assuming different substitution rates between NPs and physicians. Both the first-stage and the reduced-form estimates decline with the assumed NP-physician substitution rate, which is intuitive: If the assumed substitution rate exceeds the true rate, the NP assignment probability will rise less with a unit increase in the estimated share of physician equivalents who are NPs than it does with the actual share; conversely, if the assumed substitution rate is below the true rate, the NP assignment probability will rise more with a unit increase in the estimated share of physician equivalents who are NPs than it does with the actual share. The more the assumed substitution rate is above (below) the true rate, the smaller (larger) the first-stage estimates can be relative

to the truth. The reduced-form estimates also decline with the assumed substitution rate, given that a lower share of patients is assigned to NPs with a unit increase in the estimated share of physician equivalents who are NPs. However, the resulting IV estimates are stable.

References

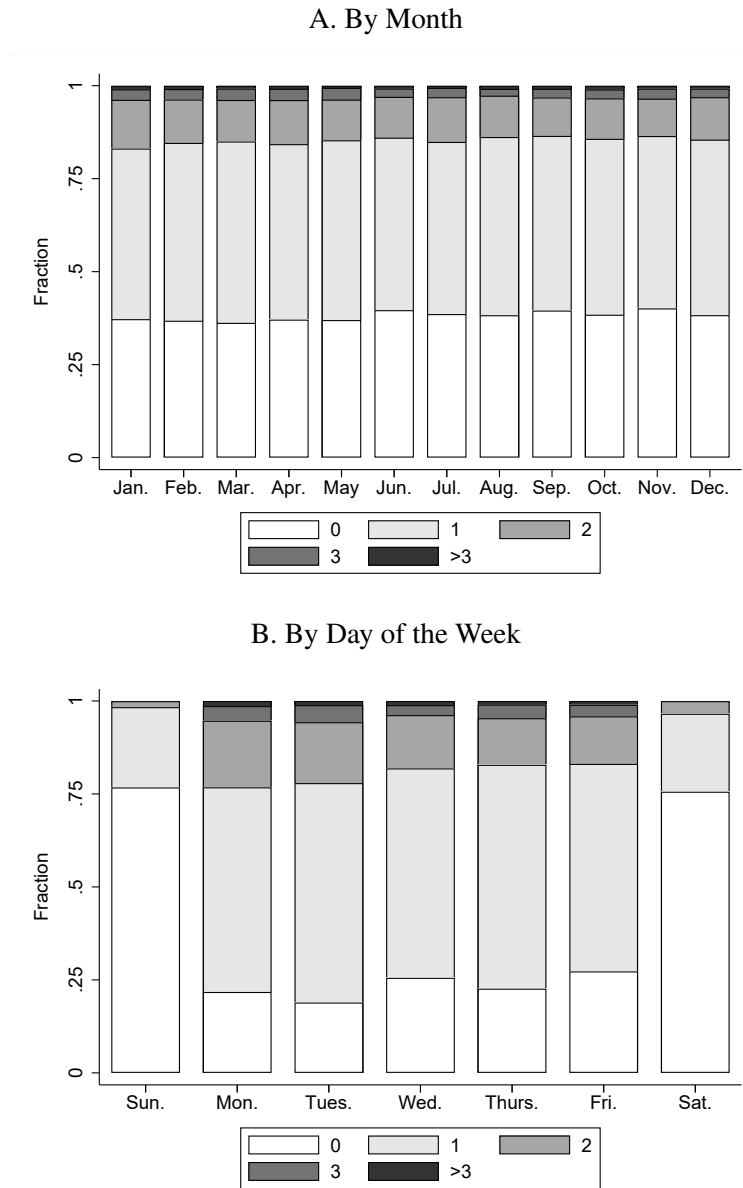
- Abadie, Alberto.** 2003. “Semiparametric Instrumental Variable Estimation of Treatment Response Models.” *Journal of Econometrics* 113 (2): 231–263.
- Chetty, Raj, John N. Friedman, and Jonah E. Rockoff.** 2014. “Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates.” *American Economic Review* 104 (9): 2593–2632.
- Dahl, Gordon B., Andreas Ravndal Kostøl, and Magne Mogstad.** 2014. “Family Welfare Cultures.” *Quarterly Journal of Economics* 129 (4): 1711–1752.
- Efron, Bradley.** 2016. “Empirical Bayes Deconvolution Estimates.” *Biometrika* 103 (1): 1–20.
- Kline, Patrick M., Evan K. Rose, and Christopher R. Walters.** 2022. “Systemic Discrimination Among Large U.S. Employers.” *Quarterly Journal of Economics* 137 (4): 1963–2036.
- Silver, David.** 2021. “Haste or Waste? Peer Pressure and Productivity in the Emergency Department.” *Review of Economic Studies* 88 (3): 1385–1417.

Figure A.1: Number of NPs on Duty



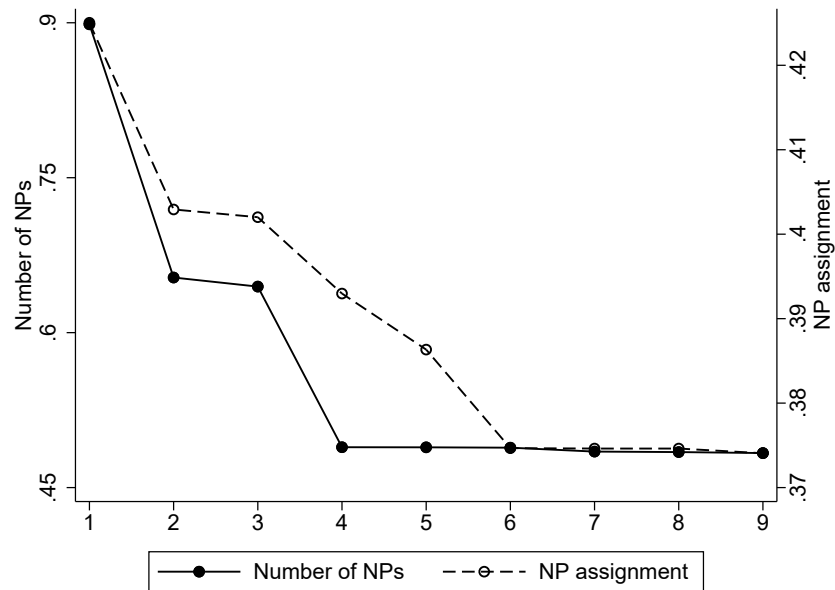
Notes: This figure shows the histogram of the number of NPs on duty in an ED-day cell. The unit of observation is at the ED-day level.

Figure A.2: Number of NPs on Duty by Month and Day of the Week



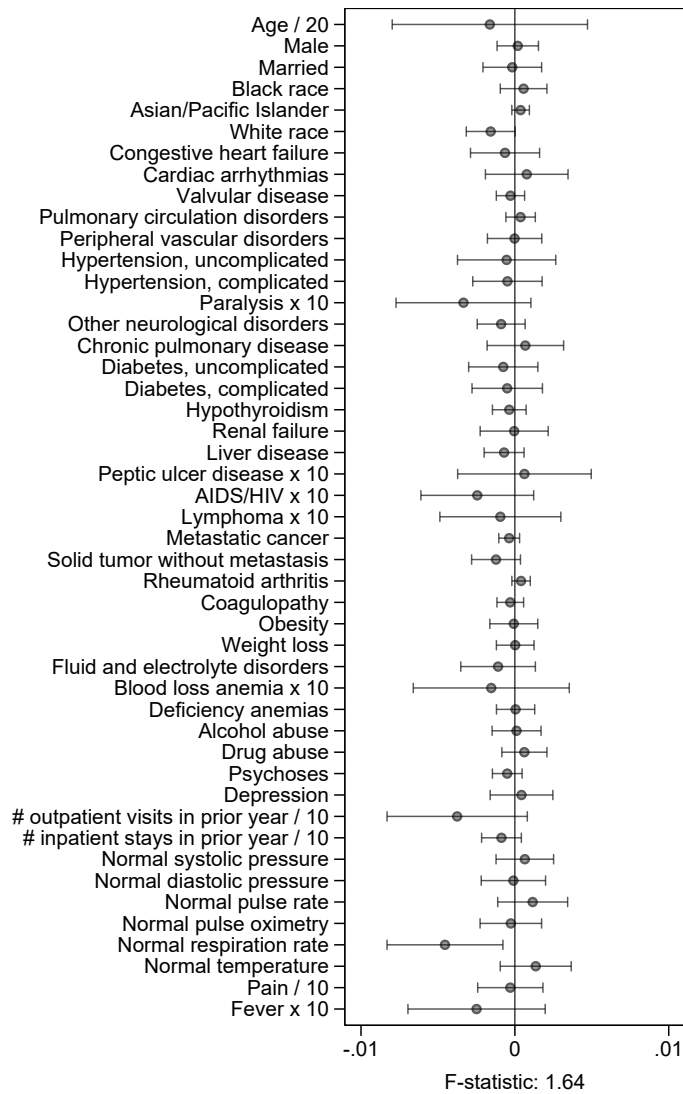
Notes: This figure plots the share of ED-day cells with different numbers of NPs on duty in each month (Panel A) and on each day of the week (Panel B). Each color of the stacked bars represents a different number of NPs on duty, with the lightest color at the bottom representing ED-day cells with the lowest number of NPs on duty (i.e., zero) and the darkest color at the top representing ED-day cells with the highest number of NPs on duty (i.e., more than three). Since only a small share of ED-day cells have more than three NPs on duty, this figure groups ED-day cells with more than three NPs into the group “>3”.

Figure A.3: Standard Deviation of Number of NPs on Duty and NP Assignment



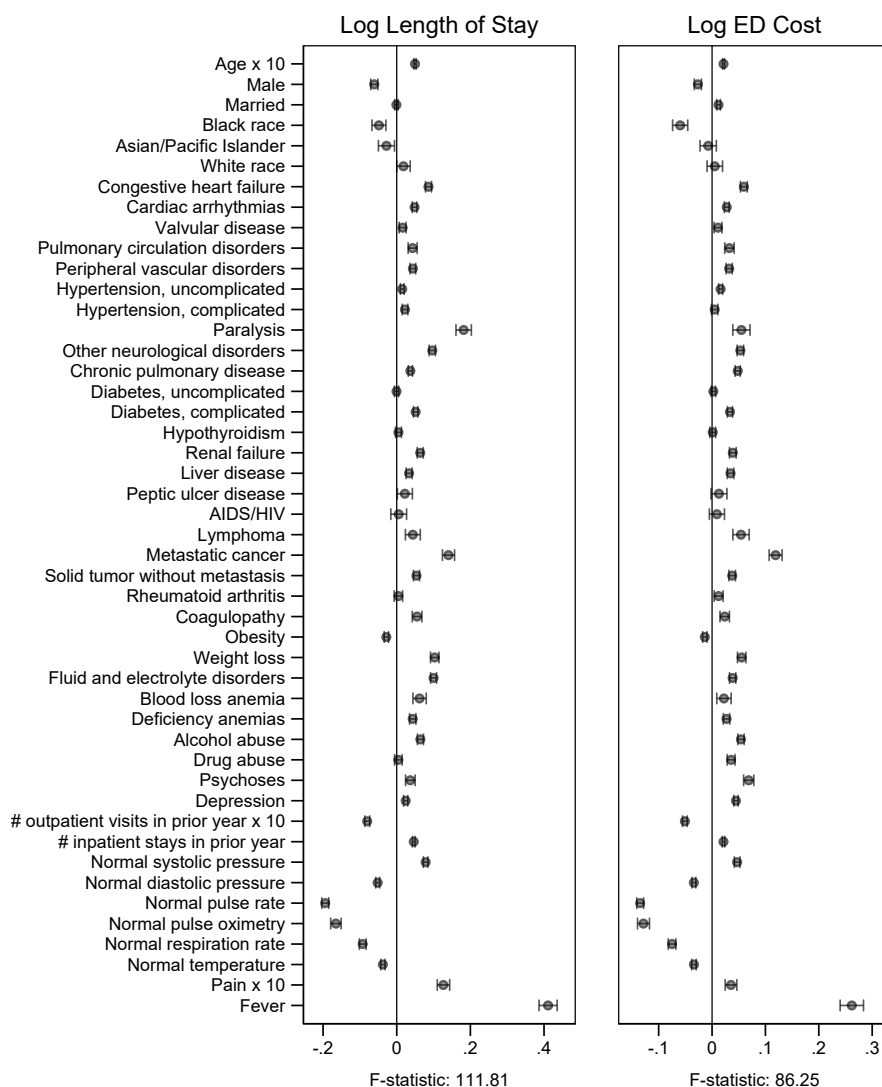
Notes: This figure shows standard deviations of the number of NPs on duty (solid dots) and whether the patient is assigned to an NP (hollow dots). The first dot shows the raw standard deviations without residualizing the number of on-duty NPs or NP assignment by any factors. The second through the last dots show the standard deviations of the residualized number of on-duty NPs and the residualized NP assignment. The following factors are added to the residualization stepwise (cumulatively) from the second to the last dot: (i) ED-by-year indicators; (ii) ED-by-month indicators; (iii) ED-by-day-of-the-week indicators; (iv) ED-by-hour-of-the-day indicators; (v) patient covariates $\mathbf{X}_i$ in Equations (1) and (2); (vi) patient volume in the ED-day cell; (vii) average wait time of patients in the ED-day cell (leaving out the index patient); and (viii) average characteristics of on-duty physicians (age, gender, value-added, and practice style; see Appendix A.3 for the construction of value-added and practice style). Average characteristics of on-duty NPs are not included in the residualization because by definition these characteristics are missing for the 38 percent of ED-day cells without any NPs on duty.

Figure A.4: Balance in Patient Characteristics



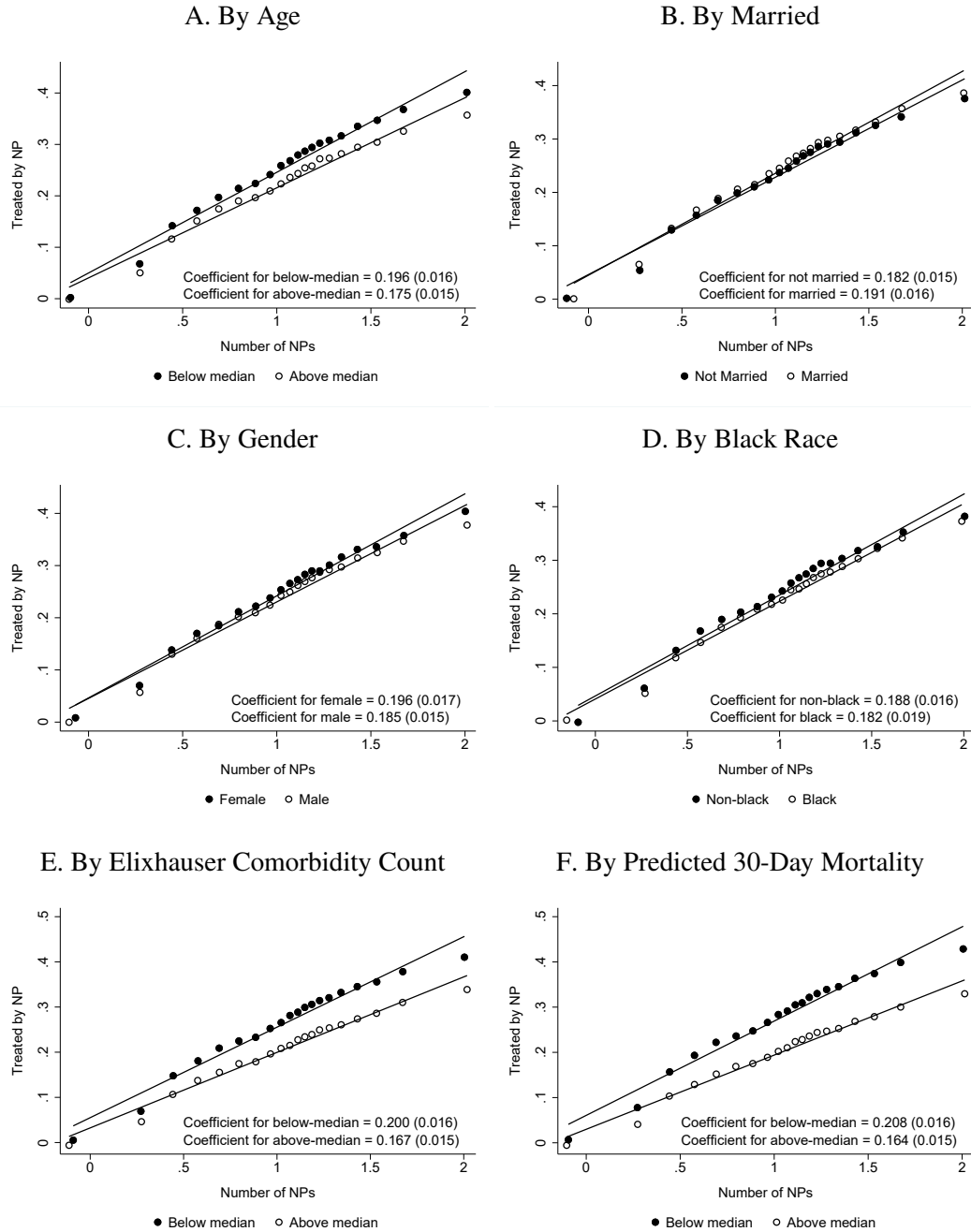
Notes: This figure shows estimated coefficients and 95 percent confidence intervals from regressions of each patient characteristic listed on the y-axis on the number of NPs on duty, controlling for the baseline control vector (i.e., ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day indicators). For improved readability, a few coefficients (and their confidence intervals) are scaled up and down by, e.g., 10, as shown by “x 10” and “/ 10” on the y-axis, respectively. At the bottom of the figure, we report the F -statistic from the joint F -test for all patient characteristics in a reverse regression with the number of NPs on duty as the dependent variable, conditioning on the baseline control vector. Standard errors are clustered by provider.

Figure A.5: Predicting Log Length of Stay and Log ED Cost



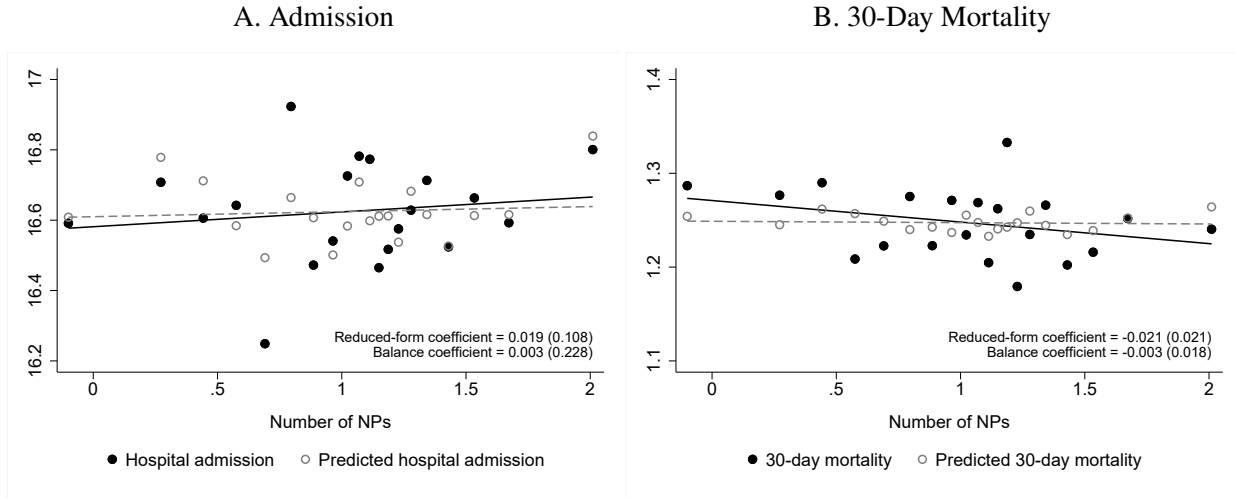
Notes: This figure shows estimated coefficients and 95 percent confidence intervals from regressions of patient log length of stay and log cost of the ED visit on patient characteristics, controlling for the baseline control vector (i.e., ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day indicators). For improved readability, a few coefficients (and their confidence intervals) are scaled up by 10, as shown by “x 10” on the y-axis. The bottom of each panel reports the F -statistic from the joint F -test of all patient characteristics, conditioning on the baseline control vector. Standard errors are clustered by provider.

Figure A.6: Monotonicity Test



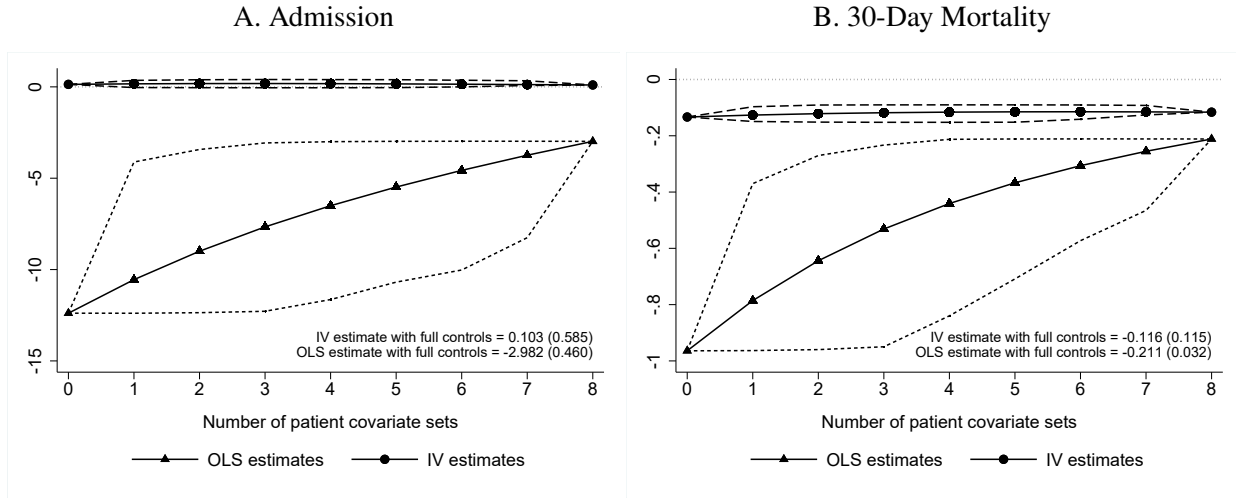
Notes: This figure shows the first-stage regressions for cases of different characteristics. Panels A-F split the sample by, respectively, age (above versus below the median of the sample), marital status, gender, race (Black versus non-Black), total number of Elixhauser comorbidities (above versus below the median of the sample), and predicted 30-day mortality (above versus below the median of the sample). Predicted 30-day mortality is generated from a linear regression of actual 30-day mortality on patient characteristics $\mathbf{X}_i$ included in Equations (1) and (2), including patient demographics, comorbidities, prior health care use, vital signs, and three-digit primary diagnosis indicators. To construct these binned scatter plots, we residualize both the y-axis and x-axis variables with respect to the baseline control vector (i.e., ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day indicators) within each subsample and then add means back. The coefficients report the first-stage estimates for each subset of patients conditional on the baseline control vector, with standard errors clustered by provider reported in parentheses.

Figure A.7: Reduced Form and Balance



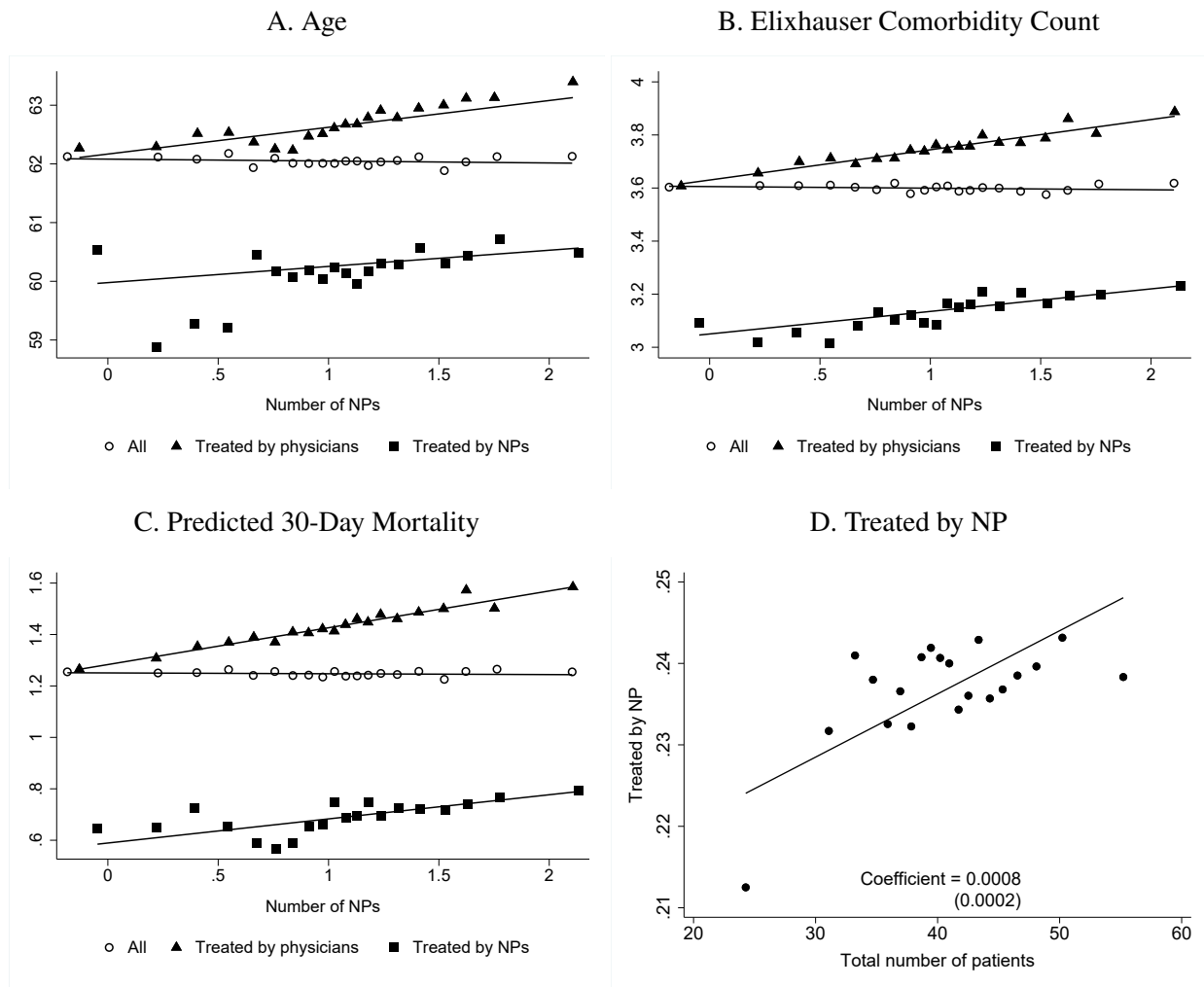
Notes: This figure shows binned scatter plots of patients' actual and predicted outcomes on the y-axis versus the number of NPs on duty on the x-axis, controlling for the baseline control vector (i.e., ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day indicators). Panels A and B report results for hospital admission in the ED visit and 30-day mortality, respectively. The solid circles and lines represent patients' actual outcomes. The hollow circles and dashed lines represent patients' predicted outcomes generated based on patient characteristics $\mathbf{X}_i$ included in Equations (1) and (2), including patient demographics, comorbidities, prior health care use, vital signs, and three-digit primary diagnosis indicators. The reduced-form coefficients are estimated using Equation (2), with patients' actual outcomes as the dependent variable; the balance-test coefficients are estimated by regressing patients' predicted outcomes on the number of NPs on duty, conditional on the baseline control vector. Standard errors clustered by provider are reported in parentheses.

Figure A.8: Stability of OLS and IV Estimates



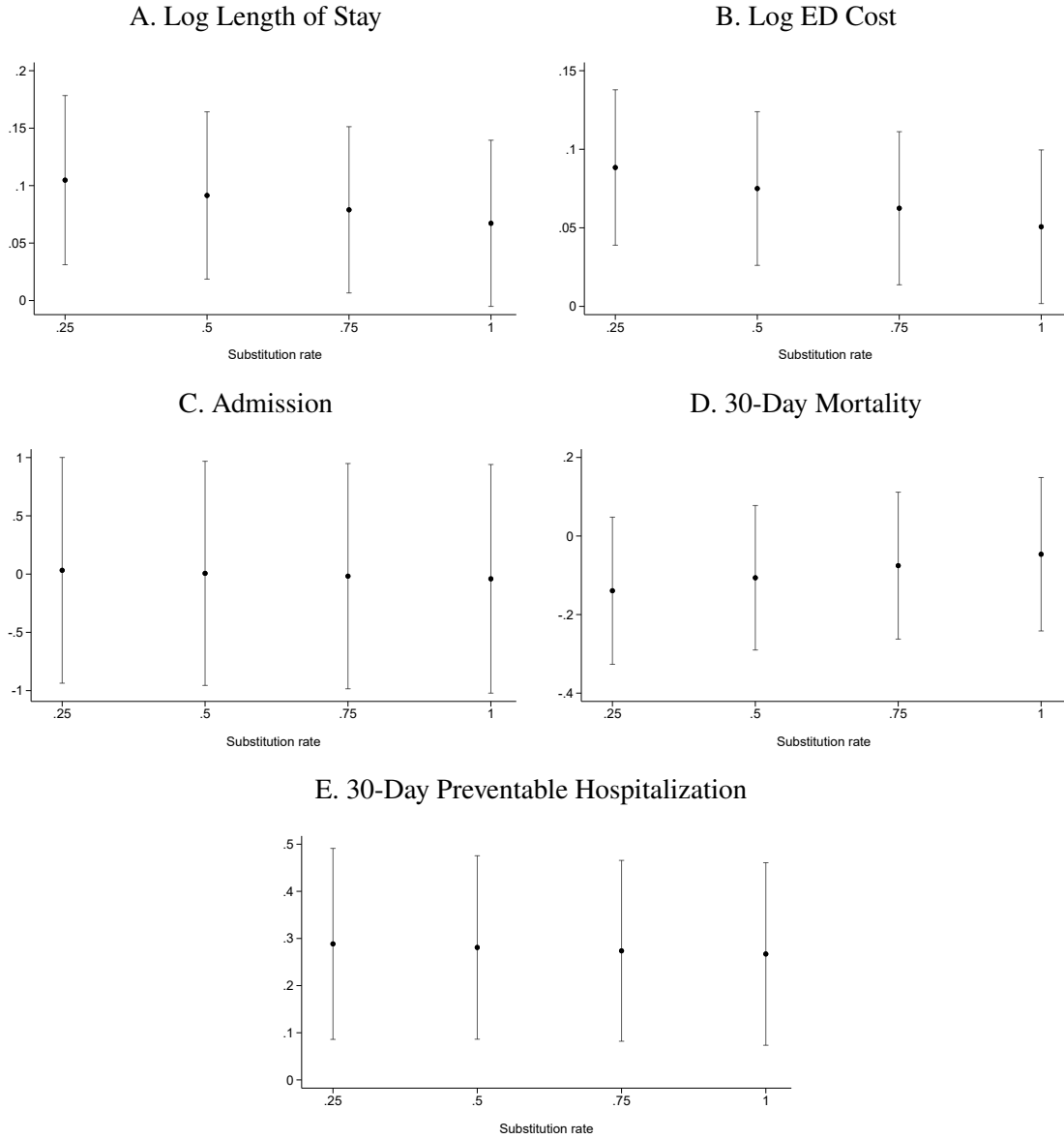
Notes: This figure shows the robustness of our OLS and IV estimates to the inclusion of different sets of patient controls. We divide patient observable characteristics into eight subsets: (i) five-year age-bin indicators; (ii) marital status; (iii) gender; (iv) race indicators; (v) indicators for 31 Elixhauser comorbidities; (vi) vital signs; (vii) prior health care use; and (viii) indicators for three-digit patient primary diagnosis of the visit. We then run separate regressions that control for each of the $2^8 = 256$ different combinations of patient covariates for each outcome. Each n on the x -axis indicates the number of covariate subsets included. For each n , we plot the maximum, mean, and minimum of the estimated coefficients for the effect of NPs using all possible combinations with n (out of eight) subsets of patient covariates. The connected triangles and circles show the means of the estimated coefficients from OLS and IV regressions, respectively. The dashed lines connect the maximum and minimum of the estimated IV coefficients. The dotted lines connect the maximum and minimum of the estimated OLS coefficients. The coefficients at the bottom of each panel show the IV and OLS estimates with the full set of patient controls, with standard errors reported in parentheses. Panels A and B report results for hospital admission in the ED visit and 30-day mortality, respectively.

Figure A.9: Patient Assignment



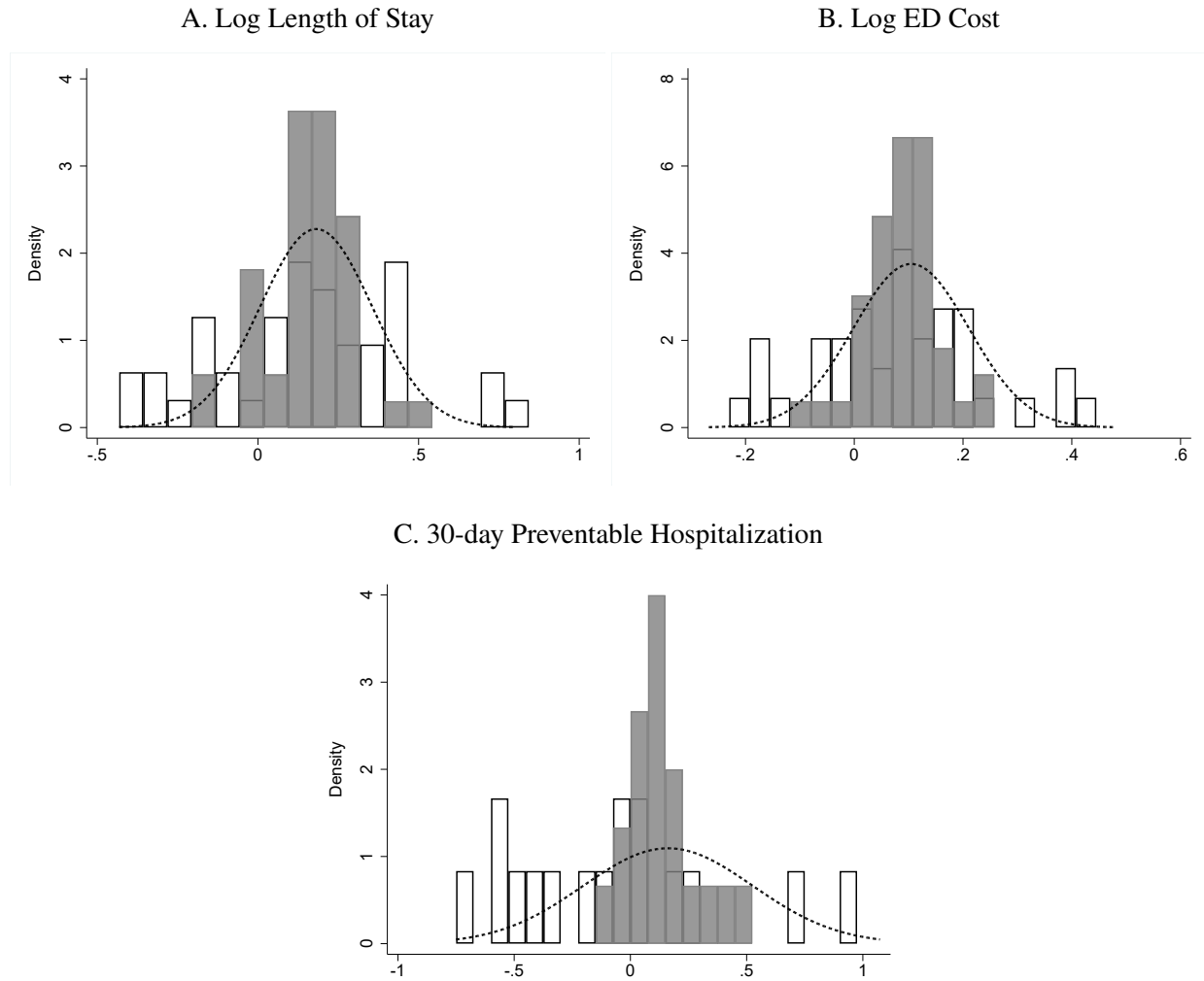
Notes: Panels A-C report patient characteristics (age, total number of Elixhauser comorbidities, and predicted 30-day mortality, respectively) on the y-axis against the number of NPs on duty on the x-axis. The unit of observation is at the case level. Both the y-axis and x-axis variables are residualized with respect to the baseline control vector (ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day indicators), with sample means added back to aid in interpretation. The circles, triangles, and squares show binned scatter plots for all cases, cases treated by physicians, and cases treated by NPs, respectively. Panel D plots whether the case is treated by an NP against the number of cases arriving in the analysis time window (i.e., 8 a.m. to 6 p.m.) of the ED-day cell of the case's visit. Both the y-axis and x-axis variables are residualized with respect to the baseline control vector, with means added back.

Figure A.10: Controlling for Physician Equivalents



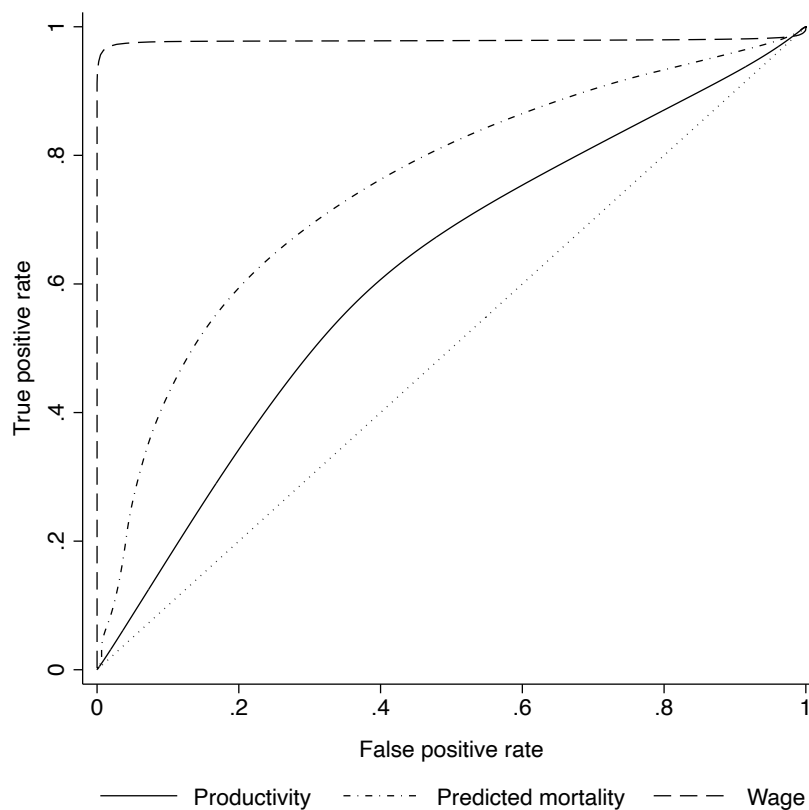
Notes: This figure reports IV estimates along with 90 percent confidence intervals based on Equations (1) and (2) with an additional control: physician equivalents. We calculate physician equivalents as the sum of the number of on-duty physicians and the number of on-duty NPs multiplied by the assumed NP-physician substitution rate. See Section III.E for details. Each panel reports IV estimates assuming that one NP substitutes for 0.25, 0.5, 0.75, and 1 physician, respectively, from left to right as noted on the x -axis. To address endogeneity in the observed number of on-duty physicians, we measure on-duty physicians as those treating patients arriving at any time in the ED-day cell of the case's visit, rather than only those treating patients arriving during the analysis time window (8 a.m.–6 p.m.), as discussed in Appendix A.7.

Figure A.11: ED-Specific Estimates of NP Effect



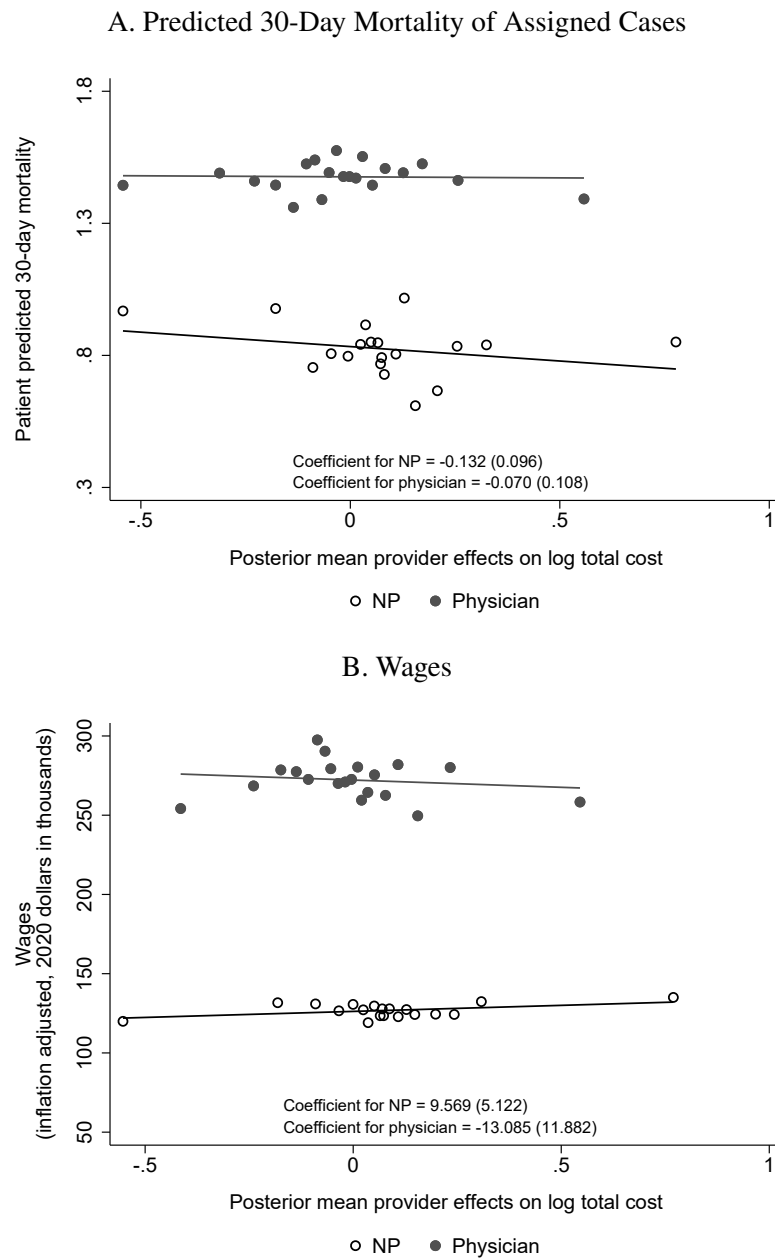
Notes: This figure reports the distribution of ED-specific IV estimates of the NP effect. Panels A, B, and C report results for the NP effect on log length of stay, log cost of the ED visit, and 30-day preventable hospitalization, respectively. The white bins show the histogram of ED-specific IV estimates without any adjustment to account for estimation noise. The gray bins show the histogram of ED-specific IV estimates with empirical Bayes adjustments (see details in Appendix A.6). The dashed lines show the standard normal density with a variance of the prior distribution of ED-specific IV estimates for each outcome.

Figure A.12: Receiver Operating Characteristic (ROC) Curves



Notes: This figure displays the ROC curves for provider productivity (in the solid line), predicted 30-day mortality of cases assigned to the provider (in the dot-dashed line), and provider wages (in the dashed line). The dotted line plots the 45-degree line. Physicians are defined as the “positive” class, and NPs are defined as the “negative” class. The ROC curves are computed based on the distributions displayed in Figure 8. See Appendix A.5.4 for details.

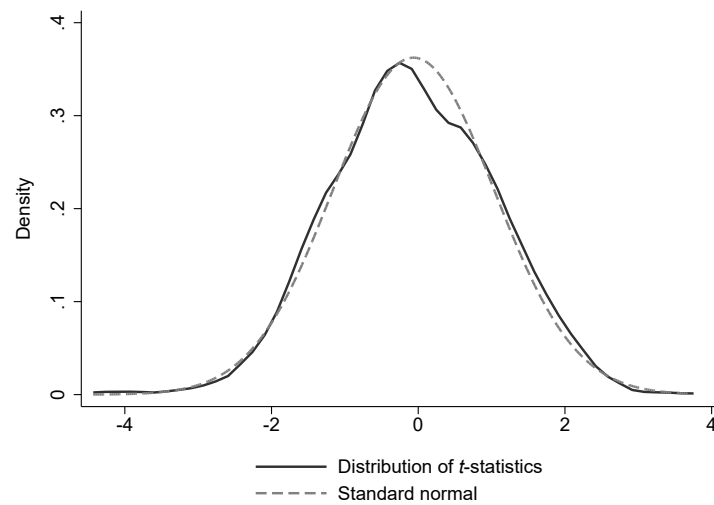
Figure A.13: Case Assignment and Wage Payment versus Productivity



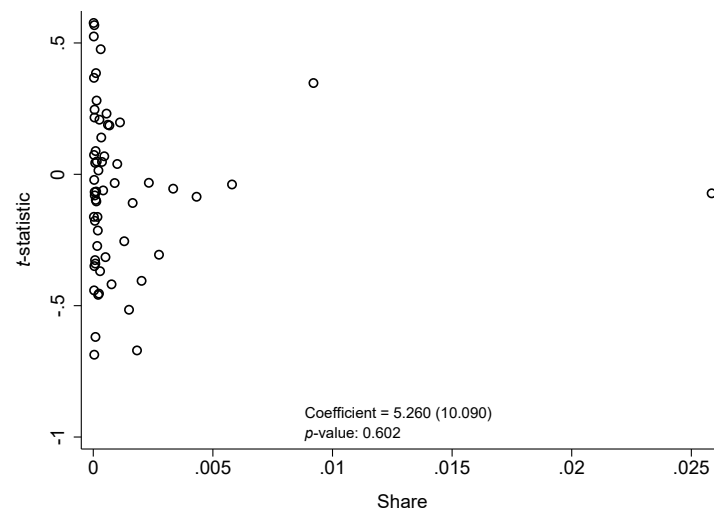
Notes: Panel A shows binned scatter plots of average risks (measured by predicted 30-day mortality) of cases assigned to a provider on the y-axis versus posterior mean provider effects on log total cost per ED visit on the x-axis. Panel B shows binned scatter plots of provider yearly wage on the y-axis versus posterior mean provider effects on log total cost per ED visit on the x-axis. Both the y-axis and x-axis variables are residualized with respect to ED indicators, with means added back for ease of interpretation. Wages are inflation-adjusted to the year 2020. Coefficients from regressions of wages on posterior mean provider effects controlling for ED indicators are reported, with standard errors clustered by ED shown in parentheses. The hollow circles report results for NPs; the solid circles report results for physicians. See more details in Appendix A.5.5.

Figure A.14: Diagnosis Coding: NPs versus Physicians

A. Distribution of t -Statistics

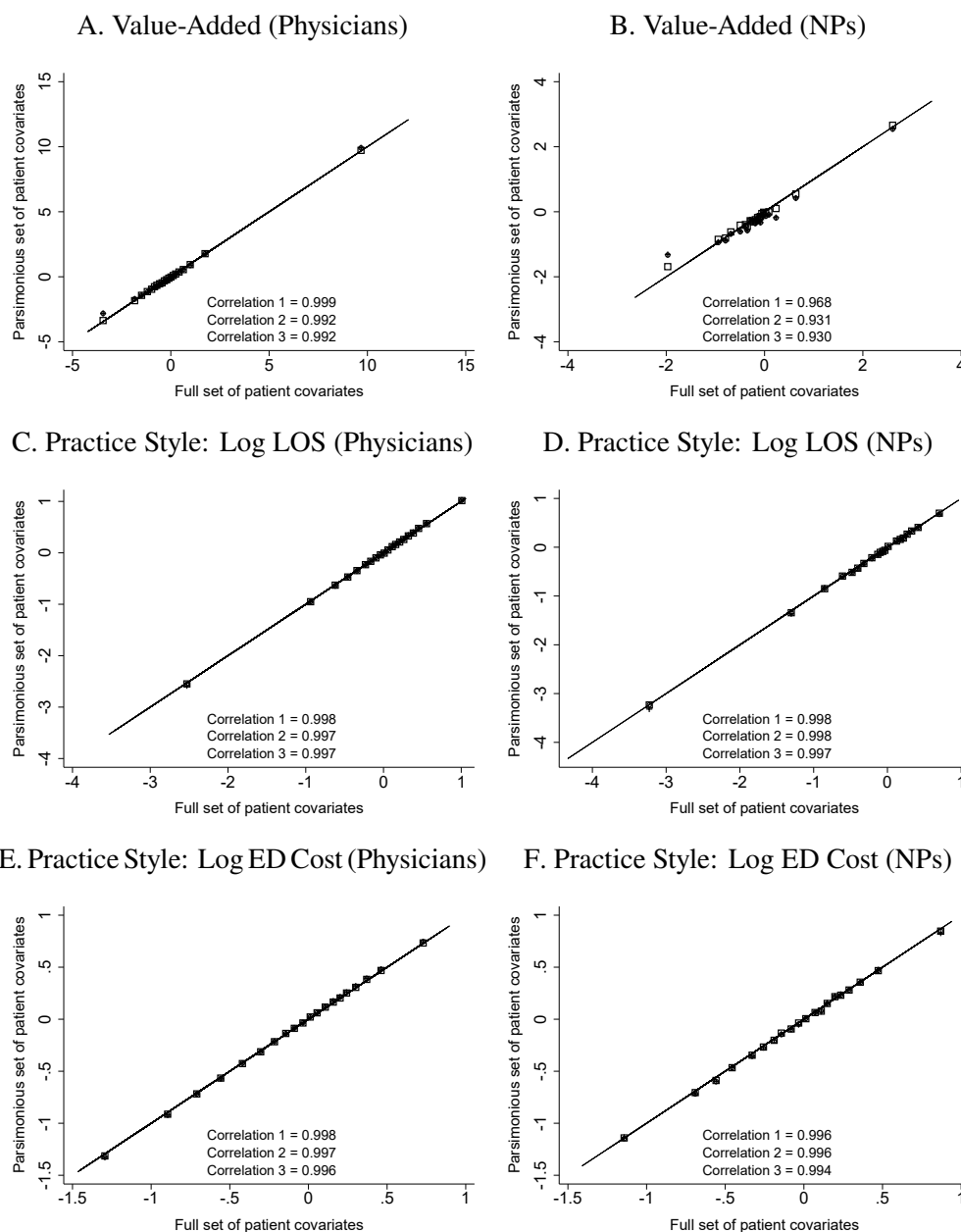


B. t -Statistics versus Diagnosis Prevalence



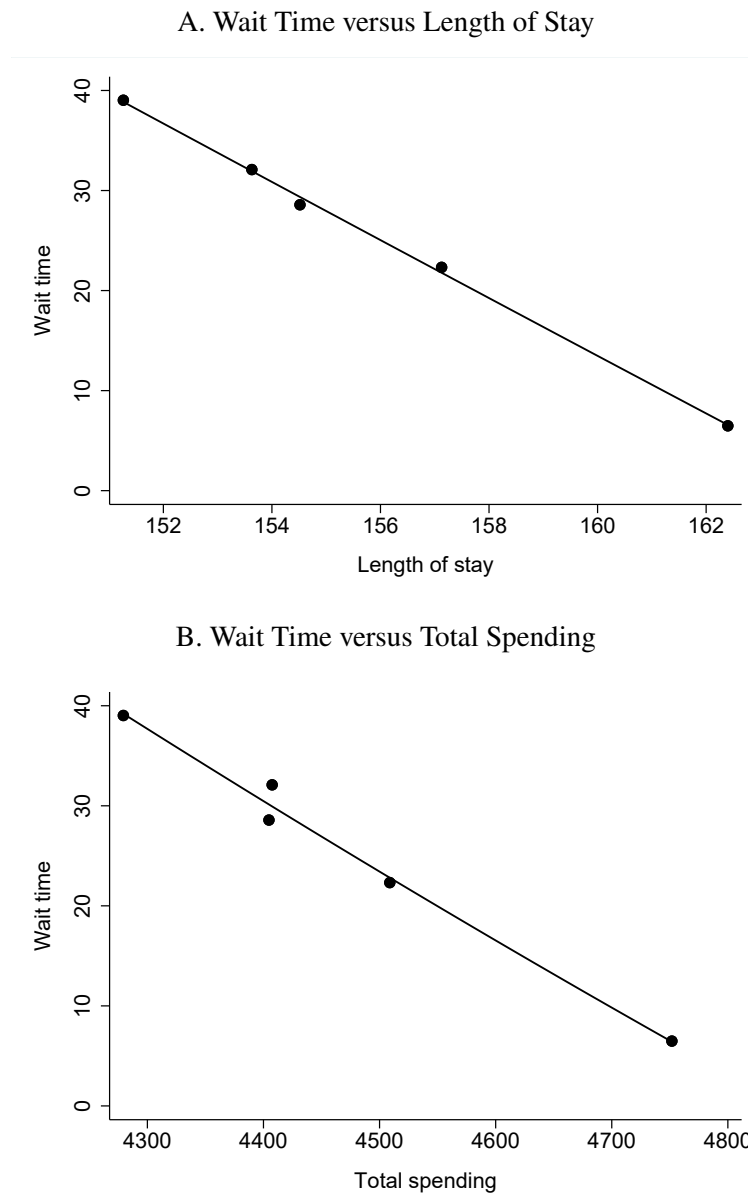
Notes: Panel A plots the distribution of the t -statistics on whether NPs and physicians are significantly different in diagnosis coding from 836 separate regressions that use each three-digit diagnosis indicator as the outcome variable. The distribution is estimated using an Epanechnikov kernel with the optimal bandwidth and shown in the solid line. For comparison, the standard normal density is plotted in the dashed line. Panel B shows binned scatter plots of the t -statistics against the prevalence of the diagnosis (measured as the share of cases with the diagnosis among cases treated by physicians on days without any NPs, to restrict the influence of patient sorting between NPs and physicians). The coefficient from the regression of the t -statistics on prevalence is reported at the bottom of the panel, along with its standard error (shown in parentheses) and p -value.

Figure A.15: Stability of Provider Value-Added and Practice Style with Varying Patient Covariates



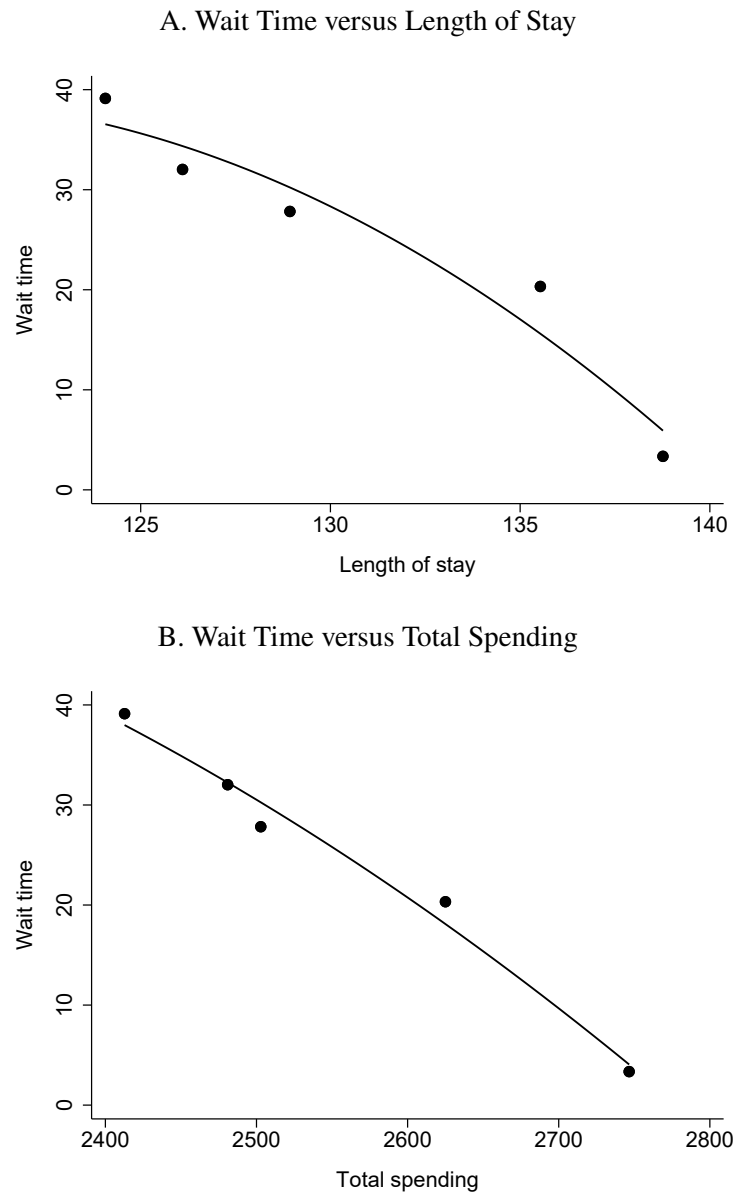
Notes: This figure shows the stability of provider value-added and practice style estimated using alternative patient covariates. Appendix A.3 provides details. The x -axis reports provider value-added/practice style constructed using the full set of patient covariates, including demographics (five-year age-bin indicators, marital status, gender, and race indicators), indicators for 31 Elixhauser comorbidities, prior health care use (the number of outpatient visits and the number of inpatient stays in VHA facilities in the prior 365 days), vital signs, and indicators for the three-digit primary diagnosis of the visit. The y -axis reports provider value-added/practice style constructed using alternative sets of patient covariates: parsimonious set 1 that includes demographics, three-digit primary diagnosis indicators, and 31 Elixhauser comorbidities; parsimonious set 2 that includes demographics and three-digit primary diagnosis indicators; and parsimonious set 3 that includes five-year age-bin and three-digit primary diagnosis indicators. Value-added and practice style constructed using parsimonious sets 1-3 are shown in squares, circles, and “+”, respectively, which largely overlap. Correlations 1-3 report correlations of value-added/practice style estimated using the full set of patient covariates with those using parsimonious sets 1-3, respectively. The solid lines show the 45-degree line. Panels A, C, and E report results for physicians. Panels B, D, and F report results for NPs.

Figure A.16: Wait Time versus Length of Stay and Total Spending



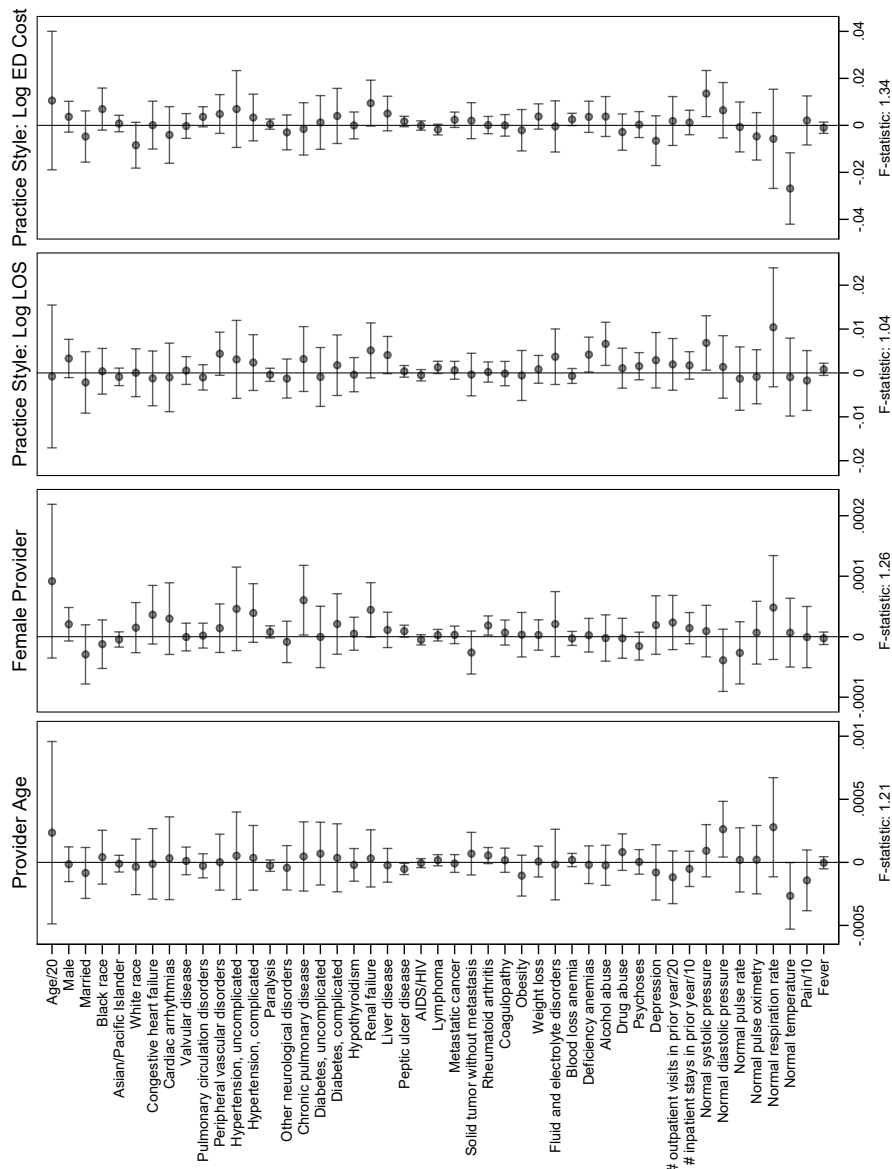
Notes: This figure shows changes in patient wait time, length of stay, and total spending with incremental numbers of NPs on duty, conditional on the number of physicians on duty, the number of cases arriving, and the controls included in the main specification (see details in Appendix A.4). Panel A presents the trade-off between wait time and length of stay. Panel B presents the trade-off between wait time and total spending of the ED visit, measured as the sum of the cost of ED care, hospital admission in the ED visit, and 30-day preventable hospitalization. The solid lines show the quadratic fit estimated based on the plotted points.

Figure A.17: Wait Time versus Length of Stay and Total Spending: Lowest-Complexity Quartile Cases



Notes: This figure shows, for cases in the lowest-complexity quartile, changes in patient wait time, length of stay, and total spending with incremental numbers of NPs on duty, conditional on the number of physicians on duty, the number of cases arriving, and the controls included in the main specification (see details in Appendix A.4). Panel A presents the trade-off between wait time and length of stay. Panel B presents the trade-off between wait time and total spending of the ED visit, measured as the sum of the cost of ED care, hospital admission in the ED visit, and 30-day preventable hospitalization. The solid lines show the quadratic fit estimated based on the plotted points.

Figure A.18: Balance of Patient Characteristics across On-Duty Provider Characteristics



Notes: This figure shows estimated coefficients and 95 percent confidence intervals from regressions of each patient characteristic listed on the y-axis on average characteristics of providers on duty in the ED-day cell of the patient's visit, controlling for baseline controls (ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day indicators). All average on-duty provider characteristics are case-weighted, with the index case left out. The average on-duty provider characteristics in each panel are, respectively, age, female share, practice style in terms of patient log length of stay, and practice style in terms of patient log cost of care at the ED. For readability, a few coefficients (and their confidence intervals) are scaled down by 10 and 20, as shown by "10" and "20" on the y-axis, respectively. The bottom of each panel reports the F -statistic from the joint F -test for all patient characteristics in a reverse regression with the average on-duty provider characteristic as the dependent variable, conditioning on the baseline controls. Standard errors are clustered by provider.

Table A.1: Characteristics of NPs and Physicians at VHA and Non-VHA

	VHA (ED)	Non-VHA (ED)	Non-VHA (all)
Panel A: NPs			
Female (%)	81.4	79.1	90.0
Age	51.3	42.9	44.8
Panel B: Physicians			
Female (%)	34.0	27.3	31.0
Age	48.1	45.8	50.4

Notes: Panel A reports summary statistics for NPs; Panel B reports summary statistics for physicians. Column 1 reports summary statistics for NPs/physicians working at the ED in our analysis sample. Column 2 reports summary statistics for NPs/physicians working at the ED observed in the 20 percent Medicare data (with age and gender information obtained from the Medicare Data on Provider Practice and Specialty (MD-PPAS)). To provide a description of providers outside of the ED, Column 3 reports characteristics of all NPs/physicians (regardless of whether they work at the ED or not) observed in the Medicare data. VHA ED NPs' and physicians' mean age and female share reported in Column 1 are slightly different from those reported in Figures 6 and 7 because Figures 6 and 7 weight means by the number of ED-day cells a provider works and patient volume.

Table A.2: Selection of Baseline Sample

Sample step	Description	Cases	Providers		
			NPs	Physicians	EDs
1. Build sample of ED cases from January 1, 2017, to January 31, 2020.	We restrict the sample to cases after the VHA directive granting NPs full practice authority in December 2016 and before COVID pandemic in the US.	7,886,164	547	5,749	146
2. Include only cases visiting during daytime.	Empirically NPs hardly work outside of the hours of 8 a.m. to 6 p.m. We drop cases visiting outside of the hours of 8 a.m. to 6 p.m. to focus on cases that could be assigned to an NP.	5,766,296	539	5,665	145
3. Include only cases in ED-month cells with NPs.	We restrict the sample to cases in ED-month cells that use NPs to treat patients.	3,597,347	521	3,781	111
4. Restrict EDs to those in which NPs and physicians are the only providers.	To focus attention on the margin between NPs and physicians and hold the population of cases seen by an NP or physician fixed, we drop EDs that use other provider types, mainly physician assistants.	1,119,396	156	1,348	44
5. Drop cases with missing demographics or extreme ages.	We drop cases with missing age or gender, or age above 99 or below 20.	1,118,836	156	1,348	44

Notes: This table reports changes in the sample size when applying each of the listed sample restrictions. Columns 3-6 report, respectively, the number of cases, NPs, physicians, and EDs remaining at each step.

Table A.3: EDs Included in Main Analysis versus All VHA EDs

	Included	All
Panel A: ED characteristics		
Daily census	30.69 [17.58]	36.06 [21.34]
Urban location	0.86 [0.35]	0.87 [0.34]
Panel B: Provider Characteristics		
Physician mean age	52.218 [11.288]	50.962 [10.979]
Share physicians female	0.286 [0.452]	0.326 [0.469]
NP mean age	51.245 [8.940]	50.915 [9.352]
Share NPs female	0.803 [0.398]	0.752 [0.432]
Panel C: Patient characteristics		
Age	61.987 [15.857]	61.770 [15.828]
Married	0.416 [0.493]	0.432 [0.495]
Male	0.905 [0.294]	0.901 [0.299]
Black	0.275 [0.447]	0.272 [0.445]
White	0.704 [0.456]	0.709 [0.454]
Asian/Pacific Islander	0.020 [0.140]	0.018 [0.133]
Outpatient visits in prior year	6.409 [7.697]	6.552 [7.717]
Inpatient stays in prior year	0.643 [1.612]	0.689 [1.684]
Elixhauser comorbidity count	3.626 [3.035]	3.705 [3.100]
Predicted 30-day mortality (%)	1.200 [2.753]	1.229 [2.782]

Notes: Columns 1 and 2 report the characteristics of EDs included in our main analysis sample and the characteristics of all VHA EDs, respectively. Panel A reports EDs' daily census in the analysis time window (i.e., 8 a.m. to 6 p.m.) and urban location. Panel B reports characteristics of the physicians and NPs in the EDs. Panel C reports characteristics of patients seen in the EDs. The standard deviation of each characteristic is reported in brackets. Patient and provider characteristics in the included EDs, i.e., Column 1, are slightly different from those reported in Table 1 and Figures 6 and 7 because Column 1 includes observations in the ED for all months between January 2017 and January 2020, regardless of whether the ED used NPs in the month.

Table A.4: Heterogeneous Effects by Provider Experience (Standardized, Cases in 2018-)

	Dependent variable				
	Log length of stay (1)	Log ED cost (2)	Admission (3)	30-day mortality (4)	30-day prevent. hosp. (5)
Panel A: Provider specific experience					
NP assignment	0.081 (0.046)	0.077 (0.032)	-0.268 (0.646)	-0.222 (0.142)	0.335 (0.149)
NP assignment $\times$ experience	-0.060 (0.025)	-0.050 (0.021)	-0.677 (0.308)	0.011 (0.044)	-0.012 (0.033)
Experience	-0.006 (0.005)	0.004 (0.009)	0.204 (0.303)	-0.008 (0.017)	-0.018 (0.012)
Panel B: Provider general experience					
NP assignment	0.104 (0.048)	0.089 (0.034)	-0.356 (0.700)	-0.238 (0.146)	0.347 (0.152)
NP assignment $\times$ experience	-0.130 (0.067)	-0.069 (0.040)	0.249 (1.337)	0.108 (0.114)	-0.029 (0.077)
Experience	-0.034 (0.015)	-0.009 (0.011)	-0.721 (0.230)	-0.017 (0.024)	-0.039 (0.025)
Full controls	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	4.637	6.529	16.304	1.251	1.226
S.D. dep. var.	1.133	0.887	36.940	11.114	11.005
Observations	742,968	741,027	747,510	747,510	747,510

Notes: This table reports heterogeneous effects of NPs by provider experience using cases visiting in 2018 or after. Panel A shows heterogeneity by provider specific experience in the case's condition, measured as the number of cases with the same three-digit primary diagnosis as the current case the provider has treated since the start of the study period to the day before the current case's visit. Panel B shows heterogeneity by provider general experience, measured as the number of cases (regardless of conditions) the provider has treated since the start of the study period to the day before the current case's visit. For ease of interpretation, both specific and general experience are standardized to have a mean of zero and a standard deviation of one for NPs and physicians separately. The outcome variables in Columns 1-5 are, respectively, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, and 30-day preventable hospitalization. All estimations include the full set of controls described in the notes to Table 2. Standard errors clustered by provider are shown in parentheses.

Table A.5: Heterogeneous Effects by Provider Experience (Standardized, Prior-Year)

	Dependent variable				
	Log length of stay (1)	Log ED cost (2)	Admission (3)	30-day mortality (4)	30-day prevent. hosp. (5)
Panel A: Provider specific experience					
NP assignment	0.078 (0.046)	0.074 (0.031)	-0.295 (0.641)	-0.221 (0.141)	0.333 (0.148)
NP assignment $\times$ experience	-0.053 (0.027)	-0.055 (0.023)	-0.646 (0.307)	0.016 (0.049)	-0.017 (0.035)
Experience	-0.009 (0.006)	0.009 (0.011)	0.260 (0.327)	-0.012 (0.018)	-0.012 (0.014)
Panel B: Provider general experience					
NP assignment	0.089 (0.046)	0.081 (0.033)	-0.331 (0.662)	-0.222 (0.143)	0.350 (0.150)
NP assignment $\times$ experience	-0.120 (0.088)	-0.063 (0.044)	-0.142 (1.162)	0.040 (0.091)	-0.117 (0.085)
Experience	-0.040 (0.017)	-0.002 (0.012)	-0.732 (0.235)	-0.008 (0.026)	-0.008 (0.026)
Full controls	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	4.637	6.529	16.304	1.251	1.226
S.D. dep. var.	1.133	0.887	36.940	11.114	11.005
Observations	742,968	741,027	747,510	747,510	747,510

Notes: This table reports heterogeneous effects of NPs by provider experience in the prior year. Panel A shows heterogeneity by provider specific experience in the case's condition, measured as the number of cases with the same three-digit primary diagnosis as the current case the provider has treated in the 365 days prior to the day of the current case's visit. Panel B shows heterogeneity by provider general experience, measured as the number of cases (regardless of conditions) the provider has treated in the 365 days prior to the day of the current case's visit. The sample is restricted to cases visiting in 2018 or after, to allow for at least a one-year look-back window for measuring experience in the prior 365 days. For ease of interpretation, both specific and general experience are standardized to have a mean of zero and a standard deviation of one for NPs and physicians separately. The outcome variables in Columns 1-5 are, respectively, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, and 30-day preventable hospitalization. All estimations include the full set of controls described in the notes to Table 2. Standard errors clustered by provider are shown in parentheses.

Table A.6: Heterogeneous Effects by Provider Experience (Standardized, Days)

	Dependent variable				
	Log length of stay (1)	Log ED cost (2)	Admission (3)	30-day mortality (4)	30-day prevent. hosp. (5)
NP assignment	0.105 (0.045)	0.065 (0.029)	0.225 (0.623)	-0.100 (0.117)	0.263 (0.123)
NP assignment $\times$ experience	-0.031 (0.068)	-0.036 (0.037)	1.098 (1.157)	0.126 (0.111)	0.101 (0.080)
Experience	-0.015 (0.014)	-0.002 (0.009)	-0.403 (0.227)	-0.022 (0.024)	-0.053 (0.026)
Full controls	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	4.608	6.483	16.625	1.247	1.234
S.D. dep. var.	1.161	0.878	37.230	11.099	11.041
Observations	1,110,798	1,108,961	1,118,836	1,118,836	1,118,836

Notes: This table reports heterogeneous effects of NPs by provider general experience measured by the number of days the provider has worked since the start of the study period to the day before the current case's visit. For ease of interpretation, the experience measure is standardized to have a mean of zero and a standard deviation of one for NPs and physicians separately. The outcome variables in Columns 1-5 are, respectively, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, and 30-day preventable hospitalization. All estimations include the full set of controls described in the notes to Table 2. Standard errors clustered by provider are shown in parentheses.

Table A.7: Opioid Use Disorder and Return Visit with Infection

	Dependent variable			
	Opioid use disorder		Infection return visit	
	30-day (1)	180-day (2)	30-day (3)	180-day (4)
NP assignment	-0.0005 (0.0015)	-0.0000 (0.0019)	0.0142 (0.0068)	0.0225 (0.0100)
Full controls	Yes	Yes	Yes	Yes
Mean dep. var.	0.022	0.037	0.068	0.187
S.D. dep. var.	0.146	0.190	0.251	0.390
Observations	1,118,836	1,118,836	123,395	123,395

Notes: Columns 1 and 2 report the IV estimates of the effect of NPs on the probability of opioid use disorder in 30 days and 180 days after the index ED visit, respectively. Columns 3 and 4 report the IV estimates of the effect of NPs on the probability of return visits with infection in 30 days and 180 days after the index ED visit, respectively. Columns 3 and 4 restrict the sample to patients with respiratory or genitourinary system infections, following the sample restriction for the antibiotic prescription analysis in Table 4. All estimations include the full set of controls described in the notes to Table 2. Standard errors clustered by provider are shown in parentheses.

Table A.8: Ten Most Common High-Mortality Diagnoses

ICD code	Description	30-day mortality (%)	Cases	Share
I50	Heart failure	5.56	12,637	0.221
N17	Acute kidney failure	6.47	4,278	0.075
R41	Other symptoms and signs involving cognitive functions and awareness	7.59	3,872	0.068
D64	Other anemias	5.01	3,634	0.064
I21	Acute myocardial infarction	7.50	3,162	0.055
A41	Other sepsis	11.51	2,754	0.048
J15	Bacterial pneumonia, not elsewhere classified	5.12	2,715	0.047
J96	Respiratory failure, not elsewhere classified	12.99	2,548	0.045
F03	Unspecified dementia	5.40	1,427	0.025
R62	Lack of expected normal physiological development in childhood and adults	16.55	1,033	0.018

Notes: This table summarizes the 10 most common three-digit diagnosis codes in the group of diagnoses with a 30-day mortality rate equal to or above the 95th percentile of the sample. The columns report, from the leftmost to the rightmost, the three-digit ICD-10 code, description of the code, 30-day mortality rate of cases with the diagnosis code, number of cases in the analysis sample with the diagnosis code, and share of cases with the diagnosis code among all cases with a three-digit diagnosis whose 30-day mortality is equal to or above the 95th percentile of the sample.

Table A.9: Outcomes of Patient Subgroups

	Log length of stay	Log ED cost	Admission	30-day mortality	30-day prevent. hosp.
Panel A: Elixhauser comorbidity count					
1st quartile	4.393 [1.145]	6.321 [0.871]	8.297 [27.583]	0.406 [6.359]	0.249 [4.981]
2nd quartile	4.517 [1.155]	6.424 [0.879]	12.526 [33.101]	0.705 [8.364]	0.538 [7.317]
3rd quartile	4.664 [1.158]	6.531 [0.875]	18.448 [38.787]	1.188 [10.833]	1.155 [10.683]
4th quartile	4.944 [1.110]	6.720 [0.832]	30.717 [46.132]	3.062 [17.227]	3.448 [18.245]
Panel B: Diagnosis predicted 30-day mortality					
< 95th percentile	4.563 [1.159]	6.452 [0.876]	14.157 [34.861]	0.853 [9.195]	1.038 [10.138]
≥ 95th percentile	5.452 [0.818]	7.059 [0.707]	62.423 [48.433]	8.572 [27.995]	4.869 [21.522]
Panel C: Diagnosis category					
Sepsis	5.724 [0.715]	7.260 [0.632]	89.587 [30.548]	11.526 [31.939]	2.442 [15.437]
Stroke	5.365 [1.027]	7.209 [0.707]	55.926 [49.658]	2.638 [16.031]	0.606 [7.760]
AMI	5.181 [0.968]	7.275 [0.929]	57.527 [49.438]	7.495 [26.336]	3.479 [18.327]
Heart failure	5.501 [0.708]	7.050 [0.566]	66.072 [47.348]	5.534 [22.866]	13.382 [34.047]
Other	4.592 [1.160]	6.470 [0.877]	15.659 [36.341]	1.150 [10.662]	1.084 [10.353]

Notes: This table shows the mean outcomes of each group of patients examined in Table 5. Standard deviations are reported in brackets. Panel A divides cases into quartiles by their total number of Elixhauser comorbidities, with higher quartiles indicating more complex cases. Panel B divides cases by whether the condition severity measured by the 30-day mortality of cases with the same three-digit ICD-10 primary diagnosis is equal to or above the 95th percentile of the sample. Panel C divides cases by their condition. The outcomes in Columns 1-5 are, respectively, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, and 30-day preventable hospitalization.

Table A.10: Robustness to Controlling for Average Risks of Patients Assigned to Physicians

	Dependent variable				
	Log length of stay (1)	Log ED cost (2)	Admission (3)	30-day mortality (4)	30-day prevent. hosp. (5)
Panel A: Baseline					
NP assignment	0.110 (0.045)	0.070 (0.030)	0.103 (0.585)	-0.116 (0.115)	0.252 (0.120)
Panel B: Control for average age					
NP assignment	0.107 (0.045)	0.070 (0.030)	0.101 (0.590)	-0.125 (0.116)	0.254 (0.122)
Panel C: Control for average Elixhauser comorbidity count					
NP assignment	0.106 (0.045)	0.069 (0.030)	0.117 (0.592)	-0.136 (0.116)	0.238 (0.122)
Panel D: Control for average predicted 30-day mortality					
NP assignment	0.100 (0.044)	0.068 (0.030)	0.191 (0.596)	-0.114 (0.117)	0.258 (0.124)
Full controls	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	4.608	6.483	16.625	1.247	1.234
S.D. dep. var.	1.161	0.878	37.230	11.099	11.041
Observations	1,110,798	1,108,961	1,118,836	1,118,836	1,118,836

Notes: This table shows the robustness of the estimates to controlling for the average health risks of cases (age, total number of Elixhauser comorbidities, and predicted 30-day mortality in Panels B, C, and D, respectively) assigned to physicians in the ED-day cell (leaving out the index case). For comparison, Panel A repeats our baseline IV estimates reported in Panel C of Table 2. The outcomes in Columns 1-5 are, respectively, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, and 30-day preventable hospitalization. All estimations include the full set of controls described in the notes to Table 2. Standard errors clustered by provider are reported in parentheses.

Table A.11: Complier and Never-Taker Characteristics

	All	Compliers		Never-takers	
	Mean	Mean	Ratio	Mean	Ratio
Age	62.05 (0.15)	61.11 (0.31)	0.98 [0.98 - 0.99]	63.69 (0.17)	1.03 [1.02 - 1.03]
Married	0.424 (0.004)	0.424 (0.008)	1.00 [0.97 - 1.04]	0.436 (0.007)	1.03 [1.00 - 1.06]
Male	0.905 (0.002)	0.905 (0.003)	1.00 [0.99 - 1.01]	0.917 (0.002)	1.01 [1.01 - 1.02]
Black	0.270 (0.011)	0.262 (0.019)	0.97 [0.83 - 1.11]	0.228 (0.015)	0.84 [0.73 - 0.95]
White	0.708 (0.011)	0.716 (0.019)	1.01 [0.96 - 1.06]	0.756 (0.015)	1.07 [1.03 - 1.11]
Asian/Pacific Islander	0.021 (0.001)	0.020 (0.002)	0.95 [0.74 - 1.15]	0.013 (0.001)	0.65 [0.55 - 0.76]
Outpatient visits in prior year	6.242 (0.080)	5.824 (0.129)	0.93 [0.89 - 0.97]	6.537 (0.110)	1.05 [1.01 - 1.08]
Inpatient stays in prior year	0.612 (0.014)	0.490 (0.029)	0.80 [0.71 - 0.89]	0.695 (0.026)	1.14 [1.05 - 1.22]
Elixhauser comorbidity count	3.599 (0.030)	3.324 (0.066)	0.92 [0.89 - 0.96]	3.965 (0.041)	1.10 [1.08 - 1.12]
Predicted 30-day mortality (%)	1.247 (0.032)	0.902 (0.067)	0.72 [0.62 - 0.83]	1.697 (0.049)	1.36 [1.28 - 1.44]

Notes: This table reports average characteristics for the overall sample, compliers, and never-takers. Complier characteristics are estimated by 2SLS regressions replacing the outcome variable y_i with $x_i \times \text{NP}_i$, i.e., the interaction between patient characteristic and the indicator for being treated by an NP, controlling for the baseline control vector (i.e., ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day indicators). Standard errors clustered by provider are reported in parentheses. Never-takers are defined as cases treated by physicians in ED-day cells with the residual share of cases treated by NPs at least as high as the 90th percentile of ED-day cells with at least one case treated by NPs. Residual shares are constructed by first collapsing the data to ED-day cells and then residualizing the share of cases treated by NPs by ED-by-year, ED-by-month, and ED-by-day-of-the-week indicators. Standard errors for the overall sample and never-takers are estimated by bootstrap, using 500 replications and blocking observations by provider. For each characteristic, the table reports the mean as well as the ratio between this mean and the overall sample mean. The 95 percent confidence interval for each ratio is shown in brackets. Predicted 30-day mortality is generated from a linear regression of actual 30-day mortality on patient characteristics $\mathbf{X}_i$ included in Equations (1) and (2), including patient demographics, comorbidities, prior health care use, vital signs, and three-digit primary diagnosis indicators.

Table A.12: Robustness to Controlling for On-duty Provider Practice Style

	Dependent variable				
	Log length of stay (1)	Log ED cost (2)	Admission (3)	30-day mortality (4)	30-day prevent. hosp. (5)
Panel A: Control for average practice styles of on-duty physicians					
NP assignment	0.091 (0.044)	0.057 (0.029)	0.013 (0.588)	-0.116 (0.116)	0.256 (0.122)
Panel B: Control for average practice styles of on-duty physicians and NPs					
NP assignment	0.091 (0.044)	0.059 (0.030)	0.021 (0.589)	-0.116 (0.116)	0.255 (0.122)
Full controls	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	4.608	6.483	16.625	1.247	1.234
S.D. dep. var.	1.161	0.878	37.230	11.099	11.041
Observations	1,110,798	1,108,961	1,118,836	1,118,836	1,118,836

Notes: Panel A reports the robustness of our IV estimates to controlling for on-duty physicians' average practice styles in terms of both log length of stay and log cost of the ED visit. Panel B reports the robustness of our IV estimates to controlling for on-duty physicians' and on-duty NPs' average practice styles in terms of both log length of stay and log cost of the ED visit. In Panel B, for ED-day cells without NPs, we replace the missing average NP practice styles with the mean practice styles of NPs in the analysis sample. The outcomes in Columns 1-5 are, respectively, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, and 30-day preventable hospitalization. All estimations include the full set of controls described in the notes to Table 2. Standard errors clustered by provider are reported in parentheses.

Table A.13: On-Duty Physician Value-Added and Outcomes of Patients Treated by NPs

	Dependent variable					
	Elixhauser comorbidity count (1)	Predicted 30-day mortality (2)	Log length of stay (3)	Log ED cost (4)	Admission (5)	30-day prevent. hosp. (7)
Physician value-added	-0.006 (0.013)	-0.021 (0.014)	-0.010 (0.008)	-0.005 (0.005)	-0.245 (0.199)	-0.023 (0.044)
Controls	Baseline	Baseline	Full	Full	Full	Full
Mean dep. var.	3.128	0.728	4.302	6.298	7.726	0.633
S.D. dep. var.	2.711	2.115	1.083	0.870	26.700	7.929
Observations	147,936	147,936	146,948	146,935	147,936	147,936

Notes: This table shows the balance in outcomes for cases treated by NPs across the average value-added of physicians on duty. See Appendix A.3 for construction details of physician value-added. The outcomes in Columns 1-7 are, respectively, the total number of Elixhauser comorbidities, predicted 30-day mortality, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, and 30-day preventable hospitalization. The sample is restricted to patients treated by NPs on days with one NP on duty and at least one physician on duty. Since Columns 1-2 examine the balance in patient characteristics, the set of controls includes only the baseline control vector (i.e., ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day indicators). The set of full controls in Columns 3-7 is detailed in the notes to Table 2. Standard errors clustered by provider are reported in parentheses.

Table A.14: NP Presence and Outcomes of Patients Treated by Physicians

	Dependent variable						
	Elixhauser comorbidity count (1)	Predicted 30-day mortality (2)	Log length of stay (3)	Log ED cost (4)	Admission (5)	30-day mortality (6)	30-day prevent. hosp. (7)
Panel A: Baseline results							
NPs on duty	0.012 (0.032)	-0.013 (0.029)	-0.002 (0.012)	0.001 (0.007)	0.358 (0.303)	-0.022 (0.098)	-0.116 (0.117)
Panel B: By tercile of case count in ED-day cell							
NPs on duty							
× Bottom two terciles	0.027 (0.043)	0.017 (0.038)	0.027 (0.016)	0.009 (0.010)	0.386 (0.402)	0.023 (0.142)	-0.084 (0.162)
× Top tercile	0.002 (0.039)	-0.035 (0.034)	-0.017 (0.014)	-0.004 (0.009)	0.340 (0.383)	-0.053 (0.114)	-0.134 (0.146)
Controls	Baseline	Baseline	Full	Full	Full	Full	Full
Mean dep. var.	3.535	1.070	4.545	6.486	15.386	1.051	1.324
S.D. dep. var.	2.996	2.758	1.276	0.836	36.081	10.200	11.432
Observations	68,863	68,863	68,214	68,208	68,863	68,863	68,863

Notes: This table shows the balance in outcomes of patients treated by physicians against the presence of NPs. The sample is restricted to patients arriving between 5 and 8 a.m. in ED-day cells with all patients arriving between 5 and 8 a.m. being assigned to physicians. Panel A shows baseline results. The empirical specification takes the form $y_i = \gamma \mathbf{1}(Z_i > 0) + \mathbf{T}_i \eta + \mathbf{X}_i \beta + \varepsilon_i$, where $\mathbf{1}(Z_i > 0)$ is an indicator for whether there are NPs on duty during 8 a.m.-12 p.m. of the ED-day cell of the patient's visit. Panel B shows heterogeneous effects by whether the total number of cases in the ED-day cell is in the top tercile of all ED-day cells. The empirical specification takes the form $y_i = \sum_{g=1}^G \mathbf{1}(\text{Group}_i = g) [\gamma_g \mathbf{1}(Z_i > 0) + \lambda_g] + \mathbf{T}_i \eta + \mathbf{X}_i \beta + \varepsilon_i$, where $\mathbf{1}(\text{Group}_i = g)$ is an indicator for whether the ED-day cell has a number of cases arriving between 5 and 8 a.m. in the top tercile or bottom two terciles of all ED-day cells. The outcomes in Columns 1-7 are, respectively, the total number of Elixhauser comorbidities, predicted 30-day mortality, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, and 30-day preventable hospitalization. Since Columns 1-2 examine balance in patient characteristics, the set of controls includes only the baseline control vector (i.e., ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day indicators). The set of full controls in Columns 3-7 is detailed in the notes to Table 2. Standard errors clustered by provider are reported in parentheses.

Table A.15: Robustness to Additional Controls

	Dependent variable				
	Log length of stay (1)	Log ED cost (2)	Admission (3)	30-day mortality (4)	30-day prevent. hosp. (5)
Panel A: Baseline					
NP assignment	0.110 (0.045)	0.070 (0.030)	0.103 (0.585)	-0.116 (0.115)	0.252 (0.120)
Panel B: Control for patient volume (linear controls)					
NP assignment	0.097 (0.045)	0.077 (0.030)	0.376 (0.595)	-0.105 (0.115)	0.244 (0.118)
Panel C: Control for patient volume (linear splines)					
NP assignment	0.095 (0.045)	0.078 (0.030)	0.381 (0.595)	-0.101 (0.115)	0.245 (0.118)
Panel D: Control for patient volume (restricted cubic splines)					
NP assignment	0.096 (0.045)	0.078 (0.030)	0.377 (0.595)	-0.101 (0.115)	0.245 (0.118)
Panel E: Control for patient volume (fixed effects)					
NP assignment	0.097 (0.045)	0.078 (0.030)	0.383 (0.595)	-0.100 (0.115)	0.241 (0.118)
Panel F: Control for wait time					
NP assignment	0.109 (0.045)	0.069 (0.030)	0.317 (0.594)	-0.074 (0.124)	0.250 (0.121)
Full controls	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	4.608	6.483	16.625	1.247	1.234
S.D. dep. var.	1.161	0.878	37.230	11.099	11.041
Observations	1,110,798	1,108,961	1,118,836	1,118,836	1,118,836

Notes: Panel A repeats the baseline results in Panel C of Table 2. Panels B, C, D, and E add controls for patient volume in the analysis time window (8 a.m. to 6 p.m.) of the ED-day cell of the patient's visit, with patient volume controlled linearly, as linear splines, as restricted cubic splines, and as fixed effects for five-case bins, respectively. Panel F adds a control for patient wait time. As wait time is potentially endogenous to patient health (healthier cases may be assigned a lower priority and hence wait longer), we add an instrument for wait time: the average wait time of other cases in the ED-day cell. The outcomes in Columns 1-5 are, respectively, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, and 30-day preventable hospitalization. All estimations include the full set of controls described under Table 2. Standard errors clustered by provider are reported in parentheses.

Table A.16: Patient-Provider Gender Match

	Dependent variable					
	Elixhauser comorbidity count (1)	Predicted 30-day mortality (2)	Log length of stay (3)	Log ED cost (4)	Admission (5)	30-day prevent. hosp. (7)
Female NP $\times$ male patient	0.007 (0.103)	-0.065 (0.100)	0.027 (0.017)	-0.012 (0.015)	0.355 (0.593)	-0.058 (0.093)
Female NP	0.029 (0.153)	0.062 (0.153)	0.006 (0.133)	0.191 (0.081)	1.327 (1.886)	-0.015 (0.091)
Male patient	0.743 (0.092)	0.745 (0.092)	-0.051 (0.014)	-0.016 (0.013)	0.262 (0.508)	0.103 (0.078)
Controls	Baseline	Baseline	Full	Full	Full	Full
Mean dep. var.	3.190	0.743	4.304	6.341	7.866	0.745
S.D. dep. var.	2.772	2.145	1.137	0.856	26.921	8.598
Observations	264,772	264,772	262,960	263,045	264,772	264,772

Notes: This table examines whether NPs treat patients of the opposite gender differently compared to the same gender. We restrict the sample to patients treated by NPs, and regress each outcome on the interaction between the indicator for female NPs and the indicator for male patients, the indicator for female NPs, and the indicator for male patients. Columns 1-2 examine the balance in patient characteristics and add controls for the baseline control vector (i.e., ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day indicators). Columns 3-7 add the full set of controls described in the notes to Table 2. The outcomes in Columns 1-7 are, respectively, the total number of Elixhauser comorbidities, predicted 30-day mortality, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, and 30-day preventable hospitalization. Standard errors clustered by provider are reported in parentheses.

Table A.17: Alternative Standard Error Clustering

	Dependent variable				
	Log length of stay (1)	Log ED cost (2)	Admission (3)	30-day mortality (4)	30-day prevent. hosp. (5)
Panel A: Baseline					
NP assignment	0.110 (0.045)	0.070 (0.030)	0.103 (0.585)	-0.116 (0.115)	0.252 (0.120)
Panel B: Clustering by ED-day cell					
NP assignment	0.110 (0.015)	0.070 (0.010)	0.103 (0.348)	-0.116 (0.113)	0.252 (0.112)
Panel C: Two-way clustering by ED-day cell and provider					
NP assignment	0.110 (0.045)	0.070 (0.030)	0.103 (0.581)	-0.116 (0.113)	0.252 (0.119)
Full controls	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	4.608	6.483	16.625	1.247	1.234
S.D. dep. var.	1.161	0.878	37.230	11.099	11.041
Observations	1,110,798	1,108,961	1,118,836	1,118,836	1,118,836

Notes: This table reports the robustness of our estimates to alternative standard error clustering approaches. Panel A repeats our baseline estimates that cluster standard errors by provider. Panel B clusters standard errors by ED-day cell. Panel C clusters standard errors using two-way clustering by ED-day cell and provider. The outcomes in Columns 1-5 are, respectively, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, and 30-day preventable hospitalization. All estimations include the full set of controls described in the notes to Table 2.

Table A.18: Alternative Instruments

	Dependent variable				
	Log length of stay (1)	Log ED cost (2)	Admission (3)	30-day mortality (4)	30-day prevent. hosp. (5)
Panel A: Baseline					
NP assignment	0.110 (0.045)	0.070 (0.030)	0.103 (0.585)	-0.116 (0.115)	0.252 (0.120)
Panel B: Include NPs with only one case					
NP assignment	0.102 (0.045)	0.076 (0.030)	-0.011 (0.594)	-0.127 (0.116)	0.285 (0.125)
Panel C: Leave out the index case					
NP assignment	0.123 (0.046)	0.074 (0.031)	0.198 (0.605)	-0.110 (0.121)	0.260 (0.126)
Panel D: Share of on-duty providers who are NPs					
NP assignment	0.084 (0.039)	0.052 (0.025)	-0.086 (0.459)	-0.094 (0.098)	0.254 (0.099)
Panel E: Indicator for any NPs on duty					
NP assignment	0.108 (0.049)	0.080 (0.030)	0.185 (0.643)	-0.048 (0.121)	0.211 (0.130)
Panel F: Leave-out share of cases assigned to NPs					
NP assignment	0.114 (0.050)	0.058 (0.031)	0.678 (0.591)	-0.043 (0.119)	0.234 (0.118)
Full controls	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	4.608	6.483	16.625	1.247	1.234
S.D. dep. var.	1.161	0.878	37.230	11.099	11.041
Observations	1,110,798	1,108,961	1,118,836	1,118,836	1,118,836

Notes: Panel A repeats the baseline results in Panel C of Table 2. Panel B reports the results using an alternative measure of the number of NPs on duty as the instrument, which includes NPs with only one case in the analysis time window of the ED-day cell. Panel C reports the results using the measure of on-duty NPs that leaves out the index case in counting the number of on-duty NPs as the instrument. Panel D uses the share of on-duty providers who are NPs in the ED-day cell (the number of on-duty NPs divided by the total number of on-duty providers in the ED-day cell) as the instrument. As discussed in Appendix A.7, we measure on-duty physicians as those treating patients arriving at any time in the ED-day cell of the case's visit. Panel E uses an indicator for any NPs on duty in the ED-day cell as the instrument. Panel F uses the share of cases assigned to NPs in the ED-day cell, leaving out the index case, as the instrument. As with the share of on-duty providers who are NPs, we estimate the leave-out share of cases assigned to NPs using cases arriving at any time in the ED-day cell. The outcomes in Columns 1-5 are, respectively, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, and 30-day preventable hospitalization. All estimations include the full set of controls described in the notes to Table 2. Standard errors clustered by provider are shown in parentheses.

Table A.19: Sample Restricted to ED-Day Cells with Zero or One NP

	Dependent variable				
	Log length of stay (1)	Log ED cost (2)	Admission (3)	30-day mortality (4)	30-day prevent. hosp. (5)
NP assignment	0.109 (0.052)	0.084 (0.032)	0.146 (0.686)	-0.042 (0.128)	0.219 (0.140)
Full controls	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	4.594	6.445	16.301	1.241	1.235
S.D. dep. var.	1.147	0.887	36.937	11.069	11.045
Observations	862,416	860,798	868,930	868,930	868,930

Notes: This table shows results using only patients in ED-day cells with zero or one NP on duty. The outcomes in Columns 1-5 are, respectively, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, and 30-day preventable hospitalization. All estimations include the full set of controls described in the notes to Table 2. Standard errors clustered by provider are shown in parentheses.

Table A.20: Including Hospital Admissions and Preventable Hospitalizations Outside VHA

	Dependent variable			
	Admission		30-day prevent. hosp.	
	VHA only (1)	VHA+Medicare (2)	VHA only (3)	VHA+Medicare (4)
NP assignment	0.806 (0.798)	0.806 (0.794)	0.485 (0.205)	0.371 (0.222)
Full controls	Yes	Yes	Yes	Yes
Mean dep. var.	20.054	20.267	1.704	2.166
S.D. dep. var.	40.040	40.199	12.941	14.556
Observations	545,791	545,791	543,253	543,253

Notes: This table shows the robustness of our results to including hospital admissions and 30-day preventable hospitalizations outside of the VHA by examining patients who enroll in both the VHA and traditional Medicare. The VHA provides linked Medicare claims for beneficiaries who are traditional Medicare enrollees. Columns 1 and 2 show the robustness of results for hospital admissions during the ED visit. Column 1 measures only hospital admissions in the VHA; Column 2 adds hospital admissions in the Medicare claims. To obtain full observation of hospital admissions in non-VHA hospitals, Columns 1 and 2 restrict the sample to patients who enroll in traditional Medicare in the month of the ED visit. Columns 3 and 4 show the robustness of results for 30-day preventable hospitalizations. Column 3 measures 30-day preventable hospitalizations in the VHA; Column 4 adds 30-day preventable hospitalizations in the Medicare claims. To obtain full observation of 30-day preventable hospitalizations in non-VHA hospitals, Columns 3 and 4 restrict the sample to patients who enroll in traditional Medicare in both the month of the ED visit and the month that follows. All estimations include the full set of controls described in the notes to Table 2. Standard errors clustered by provider are shown in parentheses.

Table A.21: Including All EDs Using NPs

	Dependent variable				
	Log length of stay (1)	Log ED cost (2)	Admission (3)	30-day mortality (4)	30-day prevent. hosp. (5)
NP assignment	0.107 (0.030)	0.047 (0.018)	-0.266 (0.393)	-0.087 (0.076)	0.141 (0.083)
Full controls	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	4.600	6.522	17.127	1.216	1.268
S.D. dep. var.	1.151	0.869	37.674	10.958	11.190
Observations	2,167,104	2,166,520	2,184,032	2,184,032	2,184,032

Notes: This table shows the results that include cases in all VHA EDs that use NPs regardless of whether the ED uses physician assistants (but exclude cases in ED-day cells with physician assistants to focus on the margin between NPs and physicians). This table expands the analysis sample from 1.1 million cases across 44 EDs in our main analysis to 2.2 million cases across 110 EDs. The outcomes in Columns 1-5 are, respectively, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, and 30-day preventable hospitalization. All estimations include the full set of controls described in the notes to Table 2. Standard errors clustered by provider are shown in parentheses.

Table A.22: Variance of Provider Effects on Medical Spending

	NPs	Physicians
Basic estimates	0.0537	0.0643
Split-sample estimates	0.0476	0.0445

Notes: This table reports the variance of provider effects on log total spending of the ED visit. Total spending of the ED visit is computed as the sum of the three main cost components: the cost of care at the ED, hospital admission, and 30-day preventable hospitalization (we multiply the latter two components by the average cost of a hospital stay, \$19,220). Row 1 reports variance of provider effects $\hat{\theta}_j$ estimated using Equations (A.5) and (A.6). To account for biases due to estimation noise in $\hat{\theta}_j$, Row 2 reports variance estimated using a split-sample approach (details in Appendix A.5.2). Column 1 reports the variance for NPs. Column 2 reports the variance for physicians.

Table A.23: Relationship between z -scores and Standard Errors

	Dependent variable: provider z -score			
	(1)	(2)	(3)	(4)
Provider std. error	0.166 (0.434)	0.203 (0.348)	-0.456 (0.385)	0.085 (0.471)
Estimation sample	Full	Full	Split	Split
Provider group	NPs	Physicians	NPs	Physicians
Mean dep. var.	0.018	-0.137	0.071	-0.056
S.D. dep. var.	1.332	1.681	1.227	1.537
Mean std. error	0.307	0.226	0.283	0.202
S.D. std. error	0.515	0.206	0.365	0.124
Providers	75	644	64	474
Observations	75	644	128	948

Notes: This table reports coefficients from regressions of provider-specific z -scores on associated standard errors. Columns 1 and 2 report results using z -scores and standard errors estimated based on the full sample. Columns 3 and 4 randomly split cases for each provider into two approximately equal-sized partitions and regress z -scores from one partition on standard errors from the other partition, stacking the two partitions in the regressions. Columns 1 and 3 report results for NPs. Columns 2 and 4 report results for physicians. The numbers of unique providers in Columns 1 and 2 are smaller than those reported in Section I.C because our deconvolution includes only providers with at least 150 cases (to restrict the inclusion of noisy provider effect estimates). Columns 3 and 4 have smaller numbers of providers than those in Columns 1 and 2, respectively, because Columns 3 and 4 further drop providers with fewer than 150 cases in each split sample.

Table A.24: Balance across Different Measures of Number of On-Duty Physicians

	Raw measure		Exclude cases since 10 p.m.		24-hour measure	
	Elixhauser comorbidity count (1)	Predicted 30-day mortality (2)	Elixhauser comorbidity count (3)	Predicted 30-day mortality (4)	Elixhauser comorbidity count (5)	Predicted 30-day mortality (6)
Number of physicians	0.030 (0.007)	0.025 (0.008)	0.020 (0.007)	0.010 (0.007)	0.006 (0.006)	0.007 (0.007)
Controls	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline
Mean dep. var.	3.599	1.247	3.599	1.247	3.599	1.247
S.D. dep. var.	3.018	2.857	3.018	2.857	3.018	2.857
Observations	1,118,836	1,118,836	1,118,836	1,118,836	1,118,836	1,118,836

Notes: This table examines balance in patient health risks across different measures of the number of on-duty physicians. Columns 1-2 define the number of on-duty physicians as those treating patients arriving between 8 a.m. and 6 p.m. in the ED-day cell. Columns 3-4 define the number of on-duty physicians as those treating patients arriving between 8 a.m. and 6 p.m. but exclude physicians who also treated patients arriving at or after 10 p.m. in the ED-day cell. Columns 5-6 define the number of on-duty physicians as those treating patients arriving at any time in the ED-day cell. The odd-numbered columns examine balance in patients' number of Elixhauser comorbidities; the even-numbered columns examine balance in patients' predicted 30-day mortality. All estimations include the baseline control vector, i.e., ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day indicators. Standard errors clustered by provider are reported in parentheses.

Table A.25: Share of Providers Who are NPs as the Instrument: Different Measures of Number of On-Duty Physicians

	Dependent variable				
	Log length of stay (1)	Log ED cost (2)	Admission (3)	30-day mortality (4)	30-day prevent. hosp. (5)
Panel A: Number of physicians based on the raw measure					
NP assignment	0.041 (0.037)	0.047 (0.025)	-0.211 (0.450)	-0.041 (0.100)	0.218 (0.100)
Panel B: Number of physicians excluding physicians treating cases arriving ≥ 10 p.m.					
NP assignment	0.066 (0.038)	0.052 (0.025)	-0.159 (0.460)	-0.027 (0.103)	0.243 (0.102)
Panel C: Number of physicians based on the 24-hour measure					
NP assignment	0.084 (0.039)	0.052 (0.025)	-0.086 (0.459)	-0.094 (0.098)	0.254 (0.099)
Full controls	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	4.608	6.483	16.625	1.247	1.234
S.D. dep. var.	1.161	0.878	37.230	11.099	11.041
Observations	1,110,798	1,108,961	1,118,836	1,118,836	1,118,836

Notes: This table reports the IV estimates using the share of on-duty providers who are NPs, i.e., the number of on-duty NPs divided by the total number of on-duty NPs and physicians, as the instrument. Panel A measures the number of on-duty physicians as those treating patients arriving between 8 a.m. and 6 p.m. in the ED-day cell. Panel B measures the number of on-duty physicians as those treating patients arriving between 8 a.m. and 6 p.m. but excludes physicians who also treated patients arriving at or after 10 p.m. in the ED-day cell. Panel C measures the number of on-duty physicians as those treating patients arriving at any time in the ED-day cell. The outcomes in Columns 1-5 are, respectively, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, and 30-day preventable hospitalization. All estimations include the full set of controls described in the notes to Table 2. Standard errors clustered by provider are reported in parentheses.

Table A.26: Share of Physician Equivalents Who are NPs as the Instrument

	First stage	Reduced form				
	NP assignment (1)	Log length of stay (2)	Log ED cost (3)	Admission (4)	30-day mortality (5)	30-day prevent. hosp. (6)
Panel A: 1 NP = 0.25 physicians						
NP share	1.269 (0.106)	0.083 (0.048)	0.060 (0.031)	-0.477 (0.565)	-0.013 (0.146)	0.308 (0.138)
Panel B: 1 NP = 0.5 physicians						
NP share	1.064 (0.072)	0.067 (0.039)	0.051 (0.026)	-0.263 (0.477)	-0.024 (0.113)	0.263 (0.107)
Panel C: 1 NP = 0.75 physicians						
NP share	0.913 (0.064)	0.059 (0.033)	0.046 (0.022)	-0.175 (0.416)	-0.023 (0.095)	0.225 (0.090)
Panel D: 1 NP = 1 physician						
NP share	0.814 (0.059)	0.054 (0.030)	0.042 (0.020)	-0.129 (0.376)	-0.022 (0.084)	0.198 (0.080)
Full controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	0.237	4.608	6.483	16.625	1.247	1.234
S.D. dep. var.	0.425	1.161	0.878	37.230	11.099	11.041
Observations	1,118,836	1,110,798	1,108,961	1,118,836	1,118,836	1,118,836

Notes: This table reports the first-stage and reduced-form estimates using the share of on-duty physician equivalents who are NPs, i.e., the on-duty physician equivalents who are NPs divided by the total physician equivalents on duty, as the instrument. Panels A-D report results assuming that an NP substitutes for 0.25, 0.5, 0.75, and 1 physician, respectively. The outcomes in Columns 1-6 are, respectively, whether the patient is assigned to an NP, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, and 30-day preventable hospitalization. All estimations include the full set of controls described in the notes to Table 2. Standard errors clustered by provider are reported in parentheses.