Online Appendix

A.1 Diagnosis Coding: NPs versus Physicians

In this appendix, we explore whether NPs and physicians are significantly different in reporting three-digit ICD-10 diagnoses. All diagnoses in our data are coded in ICD-10 within 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 diagnoses 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 the diagnosis:

$$y_i = \delta \mathrm{NP}_i + \mathbf{T}_i \eta + \varepsilon_i, \tag{A.1}$$

$$\mathrm{NP}_i = \lambda Z_i + \mathbf{T}_i \zeta + v_i, \tag{A.2}$$

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). $\mathbf{T}_i$ are indicators for ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day. The coefficient of interest is δ. As with the main specification, we cluster standard errors by provider.

Panel A of Appendix Figure A.10 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., share 0.05 of t-statistics indicating a p-value below 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% level. Panel B of Appendix Figure A.10 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]

The pattern of similar three-digit diagnosis coding between NPs and physicians could arise for the relatively straightforward cases who are compliers. Additional consults and diagnostics (Section 5.2) may also aid NPs to reach the same three-digit diagnosis as physicians. Perhaps also worth noting, VHA ED providers' reimbursements are independent of patient diagnoses, and NPs and physicians are unlikely to have differential financial incentives in diagnosis coding.

[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 potential influences of patient sorting between NPs and physicians.

A.2 Characterizing Compliers and Never-Takers

This appendix describes 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 $x_i \times \mathrm{NP}_i$, i.e., the interaction between each patient characteristic x_i and the indicator for being treated by an NP. Results are discussed in Section 4.3 and shown in Columns 2-3 of Appendix Table A.3.

To estimate 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 indicators for ED-by-year, ED-by-month, and ED-by-day-of-the-week. 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-days 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 without any NPs on duty.

Columns 4 of Appendix Table A.3 report 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.3 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 are, respectively, 4.0, 3.6, and 3.3; the average predicted 30-day mortality among these three types of cases are, 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 4.4. We consider physician value-added as a measure of risk-adjusted mortality outcomes and form these measures using leave-out data. Specifically, for physician p on day d, we measure

$$A_{p,d} = \frac{\sum_{i \in \mathcal{I}_p} \mathbf{1}(d(i) \neq d, Z_i = 0)\tilde{Y}_i}{\sum_{i \in \mathcal{I}_p} \mathbf{1}(d(i) \neq d, Z_i = 0)}, \tag{A.3}$$

where $\tilde{Y}_i$ is risk-adjusted 30-day mortality, or the difference between patient actual and predicted 30-day mortality. To deal with possible finite-sample bias, we leave out cases visiting on day d.[2] We also leave out cases visiting on days with any NPs on duty, to mitigate the concern on 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

[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.

estimates to patient predicted mortality constructed on the basis of 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.11 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 patient 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.11.

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.11 shows that NP valued-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 prior health care use and vital signs upon arrival at the ED.

A.4 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.

A.4.1 Estimating Provider Effects

We generate a measure of total spending associated with the ED visit, as the sum of the three main components of costs that we find significant NP effects: ED costs, hospital admission, and 30-day preventable hospitalizations (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 effect of extreme values, we take the log of the medical spending. 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:

$$
\begin{aligned}
y_i &= \sum_j \theta_j \mathbf{1}_{\{j(i)=j\}} + \mathbf{T}_i\eta + \mathbf{X}_i\beta + \varepsilon_i, && \text{(A.4)}\\
\mathbf{1}_{\{j(i)=j\}} &= \sum_j \lambda_j \mathbf{1}_{\{j\in\mathcal{I}_i\}} + \mathbf{T}_i\zeta + \mathbf{X}_i\gamma + v_i. && \text{(A.5)}
\end{aligned}
$$

$\mathbf{1}_{\{j(i)=j\}}$ is an indicator for whether case i is treated by provider j, and $\mathcal{I}_i$ is the set of providers on duty in the ED-day cell of case i's visit. 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 0 within each ED, using linear constraints in the 2SLS estimation to yield valid standard errors.[3]

The F-statistics for the joint significance of on-duty provider indicators in the first-stage regressions, i.e., Equation (A.5), have shares of 0.99 and 0.68 above 10 and 100, respectively, suggesting that provider availability is strongly predictive of the treating provider.[4] Appendix Figure A.12 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, i.e., ED-by-time-category indicators.[5] 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 ED-by-time-category indicators, 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 make plausible the assumption that the set of on-duty providers is conditionally independent of the set of patients arriving, supporting the validity of our instruments.

A.4.2 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.4.1 yields a case-weighted variance of 0.054 for NPs, and 0.064 for physicians (see Appendix Table A.17).[6] However, these estimates are upward

[3] We also normalize provider effects to have a case-weighted mean of 0 within ED-provider types. We find the results are very similar: For the (split-sample) variance of provider effects reported below in Appendix A.4.2, the standard deviation of NP effects with normalization by ED and by ED-provider type are 0.22 and 0.19, respectively; the standard deviation of physician effects remains stable at 0.21. For the probability that a randomly selected NP incurs a lower spending than a randomly selected physician (Appendix A.4.3), the number changes slightly from 38 percent to 35 percent.

[4] Since $\mathbf{1}_{\{j(i)=j\}}$ is always zero for patients outside of the ED a provider practices, we report F-statistics from the first stage regression in Equation (A.5) using observations in each ED separately.

[5] We compute case-weighted average characteristics of on-duty providers, with the index case left out. For practice style examined in this balance test, to deal with the concern on 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.4) and (A.5).

[6] 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.

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, resembling that employed in earlier studies (e.g., Chetty, Friedman, and Rockoff 2014; Silver 2021). Specifically, we randomly split a provider's patients within each day to two approximately equal-sized partitions. We then estimate the 2SLS model in Equations (A.4) and (A.5) 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 $\mathrm{Cov}(\theta, e_b) = \mathrm{Cov}(e_a, \theta) = 0$. The random split of patients for each provider-day makes plausible the assumption that e_a and e_b are uncorrelated, i.e., $\mathrm{Cov}(e_a, e_b) = 0$. We therefore can compute the variance of provider effects as the covariance of $\hat{\theta}_a$ and $\hat{\theta}_b$:

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

We perform this calculation for NPs and physicians separately.

Appendix Table A.17 reports the case-weighted variance of provider effects from the split-sample approach. The variance for physicians is estimated to be 0.045, which is 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 increases medical spending associated with the ED visit 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.4.3 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.4) and (A.5). 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) 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:

$$\hat{f}(\theta) = \frac{1}{J}\sum_{j=1}^{J}\frac{1}{s_j}\hat{g}_c(\theta/s_j). \tag{A.6}$$

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.18, 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.[7] 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.17. 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 9 displays the deconvolved density of provider effects for NPs and physicians.

Using the deconvolved density of NP and physician effects, we estimate the probability that a randomly drawn NP is costlier than a randomly drawn physician by

$$\Pr\left(\theta_j > \theta_{j'} \middle| j \in \mathcal{J}_{NP}, j' \in \mathcal{J}_{MD}\right) = \int_0^1 \hat{F}_{MD}(\theta)d\hat{F}_{NP}(\theta), \tag{A.7}$$

where $\hat{F}_{MD}(x)$ and $\hat{F}_{NP}(x)$ are the deconvolved cumulative density 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 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 assuming the average treatment effect is as large as that among the highest complexity quartile patients which is twice of the LATE estimate (0.136 versus 0.067), the probability that NPs are less costly remains large at 28 percent.

A.4.4 ROC Curve Representation

The probability in Equation (A.7) is equivalent to the c-statistic, or area under the curve (AUC), of a receiver operating characteristic (ROC) curve. The ROC curve displays the performance of a classification exercise in which one were to classify providers by a certain characteristic. In the case of provider effects, the c-statistic of 0.62 indicates relatively poor performance in classifying providers as NPs versus physicians depending

[7] To restrict the inclusion of noisy δ_j, our deconvolution excludes providers with less than 150 cases. We set the support of provider effects to $[\delta^5 - SD, \delta^{95} + SD]$, where δ^5, δ^{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.

on their (true) provider effects from their respective deconvolved distribution.

We construct ROC curves, based on respective provider characteristics of productivity and wages, where we consider physicians as the "positive" class and NPs as the "negative" class. For each characteristic of productivity 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: $\mu_j = -\theta_j$. For a given characteristic x, we plot the ROC curve with $1 - \hat{F}^x_{MD}$ (i.e., the true positive rate) on the y-axis and $1 - \hat{F}^x_{NP}$ (i.e., the false positive rate) on the x-axis, where $\hat{F}^x_{MD}$ and $\hat{F}^x_{NP}$ are the empirical cumulative distribution functions of x among NPs and physicians, respectively. For productivity, we use the deconvolved distributions previously described in Appendix A.4.3, noting that $\hat{F}^\mu_{MD} = 1 - \hat{F}^\theta_{MD}$ and $\hat{F}^\mu_{NP} = 1 - \hat{F}^\theta_{NP}$. For wages, we use the empirical cumulative distribution function based on the annualized full-time-equivalent ("yearly") wage of each provider j, inflation-adjusted to 2020 dollars.

We show both ROC curves in Appendix Figure A.8. As mentioned above, the c-statistic based on productivity is 0.62. The c-statistic based on wages is 0.99.

A.4.5 Correlation Between Productivity and Wages

Separately for NPs and physicians, we estimate the correlation between provider wages and productivity, measured (as an additive inverse) by provider effects on log total spending associated with the ED visit (i.e., θ_j), with the following regression:

$$\text{wage}_j = \alpha\tilde{\theta}_j + \mathbf{L}_j\pi + \varepsilon_j. \tag{A.8}$$

The dependent variable wage_j is the 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 $\hat{\theta}_j$ is estimated with noise, we use empirical Bayes posterior means of each provider effects, $\tilde{\theta}_j$, which we calculate as

$$\tilde{\theta}_j = w_j\hat{\theta}_j + (1 - w_j)\hat{\theta}, \tag{A.9}$$

where $w_j = \frac{\hat{\psi}^2}{s_j^2 + \hat{\psi}^2}$ is the weight based on $\hat{\psi}^2$ and s_j^2, which are, respectively, the variance of the prior distribution of θ_j, estimated separately for NPs and physician in Appendix A.4.2, and the variance of the sampling error for each $\hat{\theta}_j$ calculated as the square of the standard error of $\hat{\theta}_j$. $\hat{\theta}$ is set to 0 for physicians, and 0.067 for NPs (i.e., the average NP effect estimated by the 2SLS model in Equations (1) and (2), using patient log total spending as the outcome). The shrinkage estimator in Equation (A.9) is often used (e.g., Chetty, Friedman, and Rockoff 2014; Chandra et al. 2016; Abaluck et al. 2021) and equivalent to empirical Bayes posterior means when assuming the prior distribution is normal.

We also estimate empirical Bayes posteriors of provider effects as Kline, Rose, and Walters (2022):

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

where φ denotes the standard normal density.

A.5 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:

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

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

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

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 overall IV estimate of the NP effect in Equations (1) and (2), which is reported in Section 4.

The gray bins in Appendix Figure A.7 plot the empirical Bayes posterior mean $\hat{\delta}_\ell$ for each ED in our sample.[8] 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.

[8] The figure reports results for all EDs 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).

References

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

Abaluck, Jason, Mauricio Caceres Bravo, Peter Hull, and Amanda Starc. 2021. "Mortality Effects and Choice across Private Health Insurance Plans." *Quarterly Journal of Economics*, 136(3): 1557–1610.

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

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.

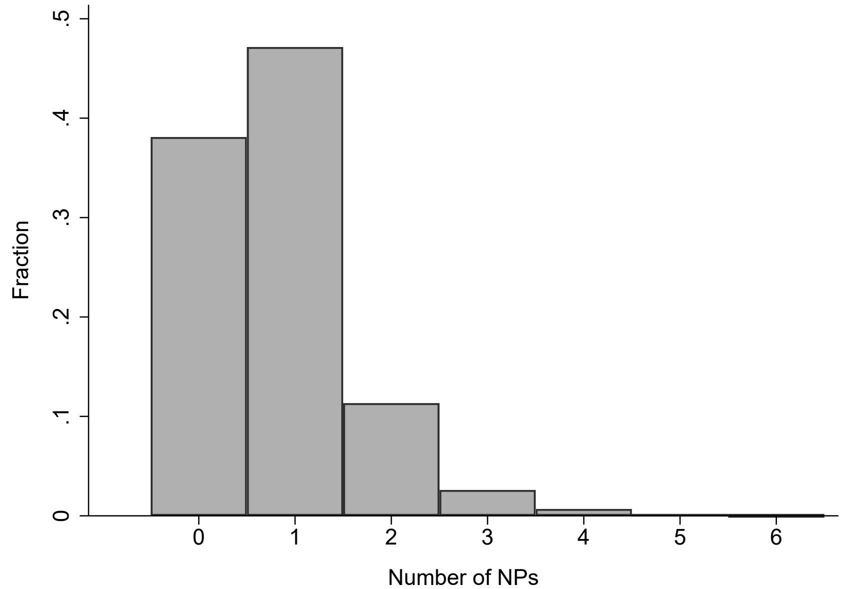
Dahl, Gordon B., Andreas Ravndal Kostøl, and Magne Mogstad. 2014. "Family Welfare Cultures." *Quarterly Journal of Economics*, 129(4): 1711–1752.

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.

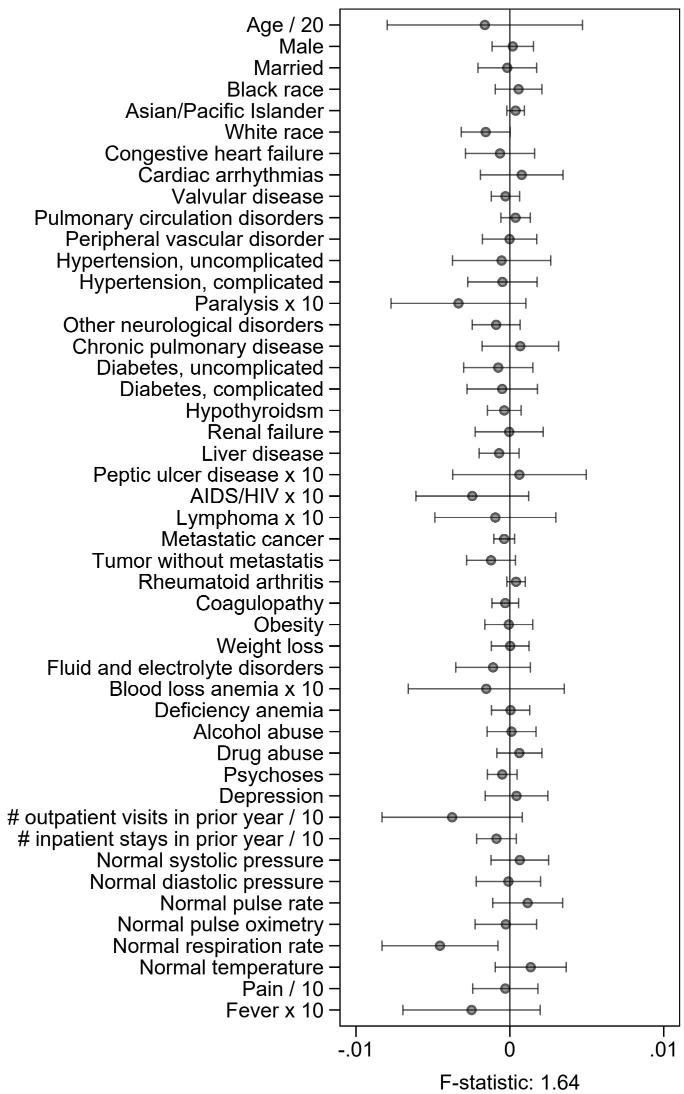
Silver, David. 2021. "Haste or Waste? Peer Pressure and Productivity in the Emergency Department." *Review of Economic Studies*, 88(3): 1385–1417.

Figure A.1: Number of NPs on Duty



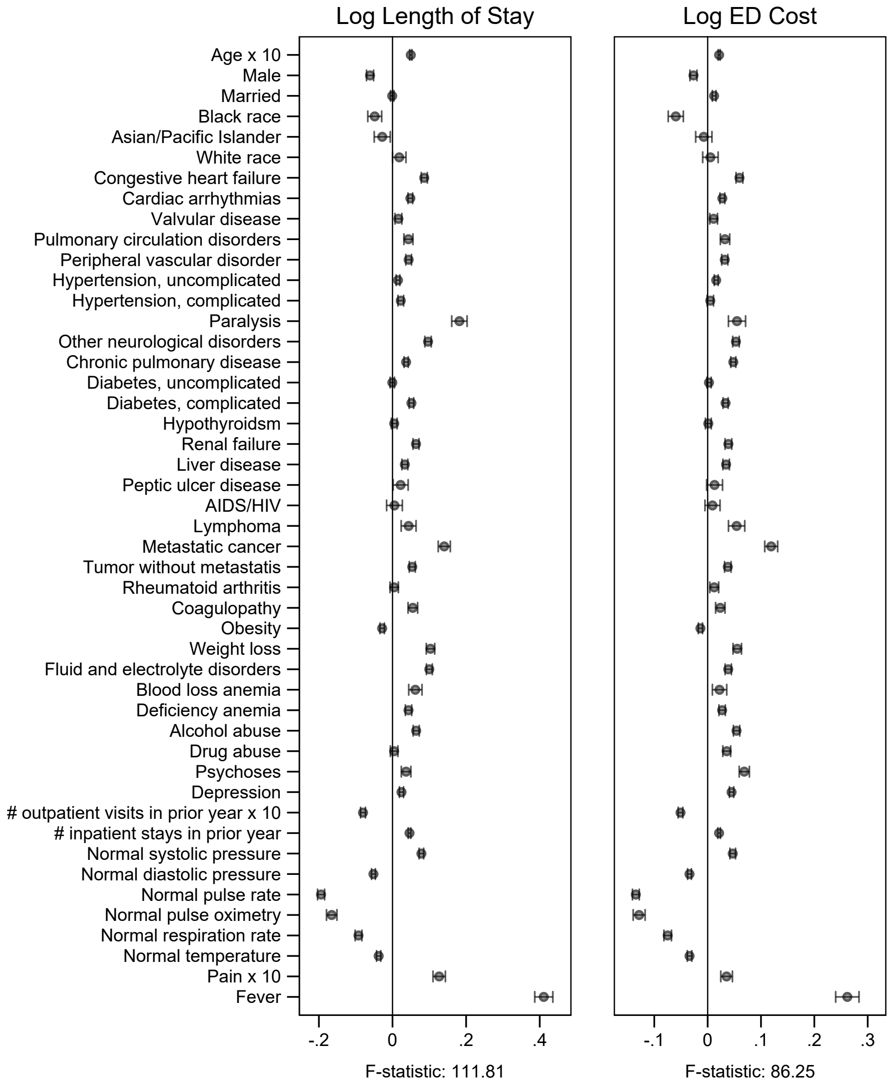
Notes: This figure shows the histogram of the number of NPs on duty in an ED-day cell. The unit of observation is at the ED-day level.

Figure A.2: Balance in Patient Characteristics



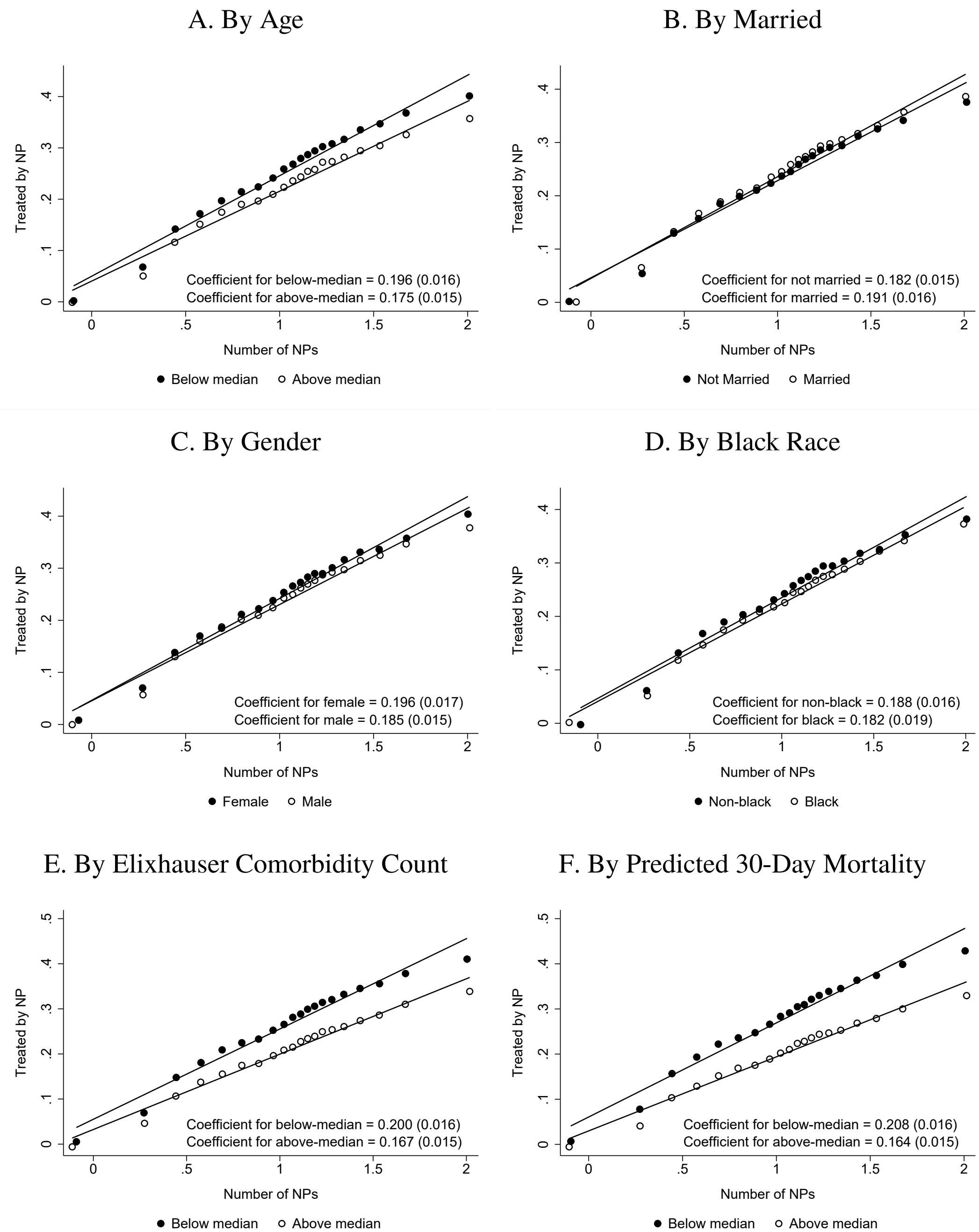
Notes: This figure shows estimated coefficients and 95% confidence intervals from regressions of each patient characteristic listed on the *y*-axis on the number of NPs on duty, controlling for the baseline control vector (i.e., indicators for ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day). For improved readability, a few coefficients (and their confidence intervals) are scaled up and down by, e.g., 10, as shown by "× 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.3: Predicting Log Length of Stay and Log ED Cost



Notes: This figure shows estimated coefficients and 95% confidence intervals from regressions of patient log length of stay (Panel A) and log cost of the ED visit (Panel B) on patient characteristics, controlling for the baseline control vector (i.e., indicators for ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day). For improved readability, a few coefficients (and their confidence intervals) are scaled up by 10, as shown by “× 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.

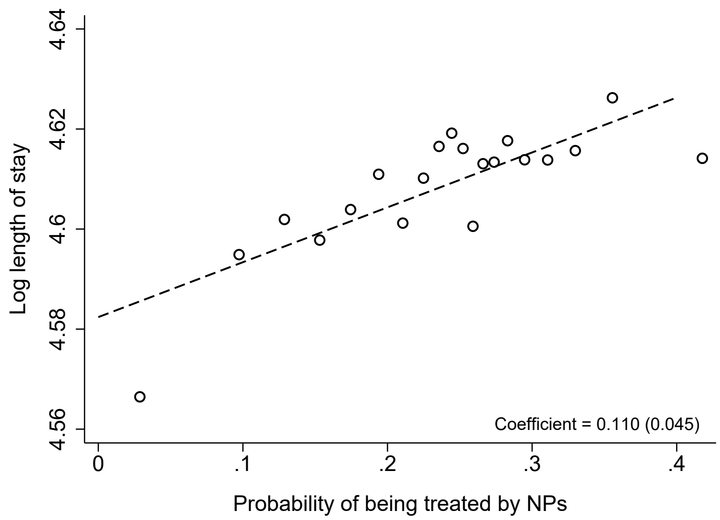
Figure A.4: Monotonicity Test



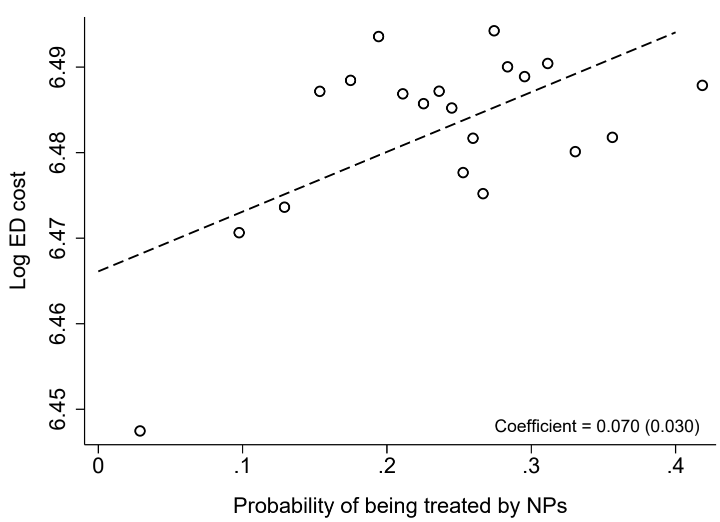
Notes: This figure shows the first-stage regression for cases of different characteristics. Panels A-F split the sample by, respectively, age (above versus below the median of the sample), marital status, gender, race (Black versus non-Black), total number of Elixhauser comorbidities (above versus below 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 diagnosis indicators. To construct these binned scatter plots, we residualize both the *y*-axis and *x*-axis variable with respect to the baseline control vector (i.e., indicators for ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day) 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.5: Visual IV

A. Log Length of Stay

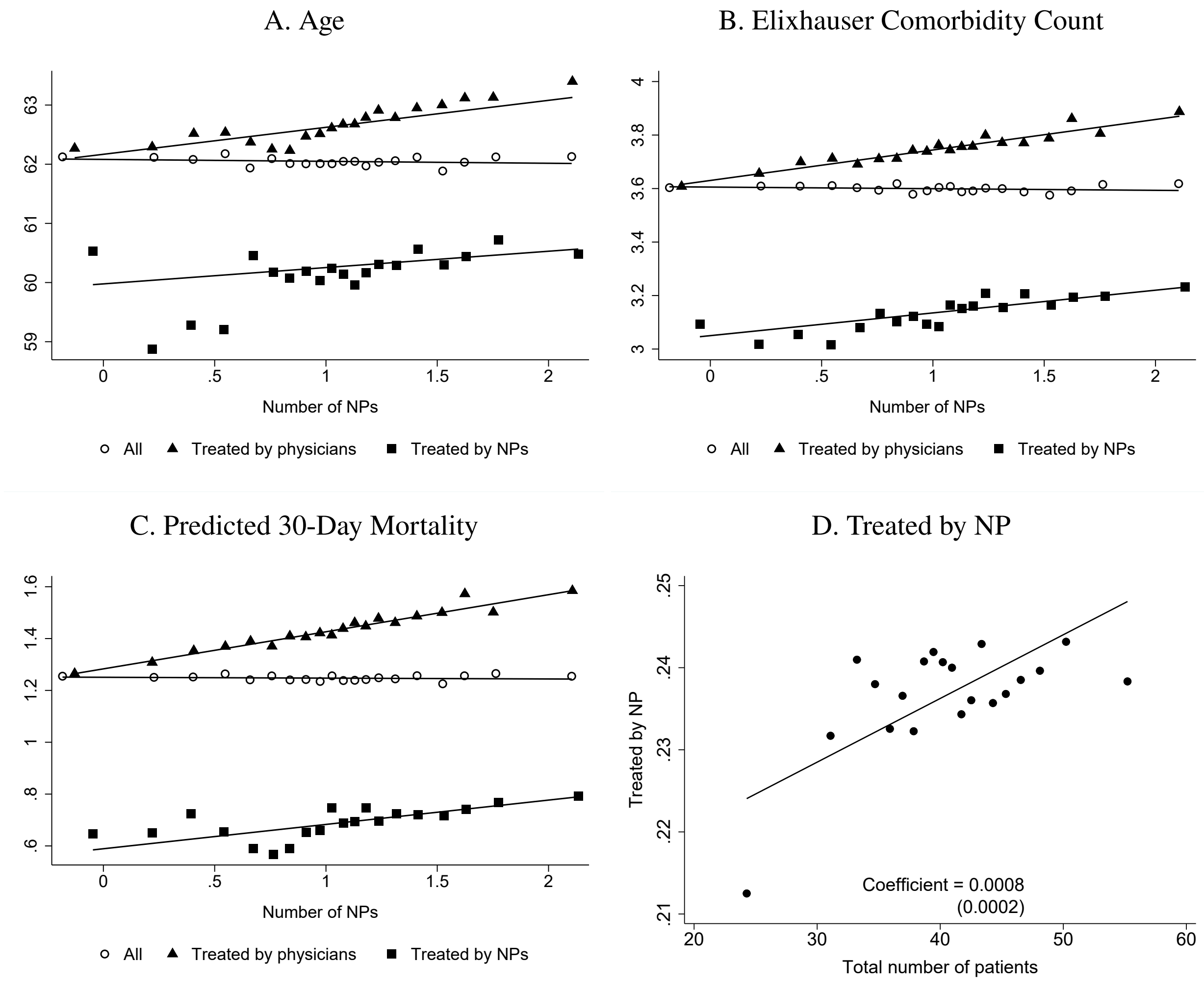


B. Log ED Cost



Notes: This figure shows the visual IV plot of the effect of NPs on patient log length of stay (Panel A) and log cost of the ED visit (Panel B). In each panel, we plot the mean outcome (log length of stay or log ED cost) on the *y*-axis versus patient probability of being treated by an NP on the *x*-axis. Patient probability of being treated by an NP is generated using the first-stage regression in Equation (2). Patient outcomes on the *y*-axis are generated using the corresponding reduced-form regression with a dependent variable of log length of stay in Panel A and log ED cost in Panel B. The coefficients correspond to the IV estimates, with standard errors clustered by provider reported in parentheses.

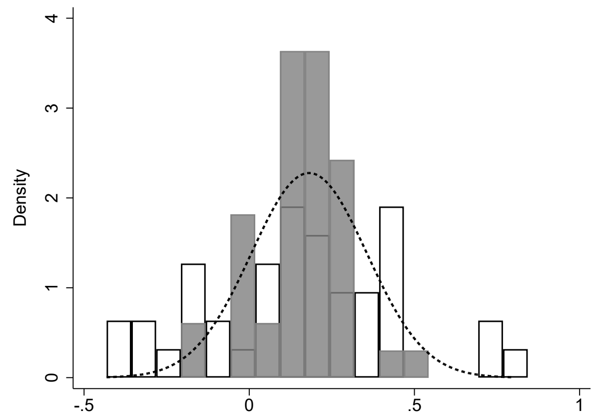
Figure A.6: Patient Assignment



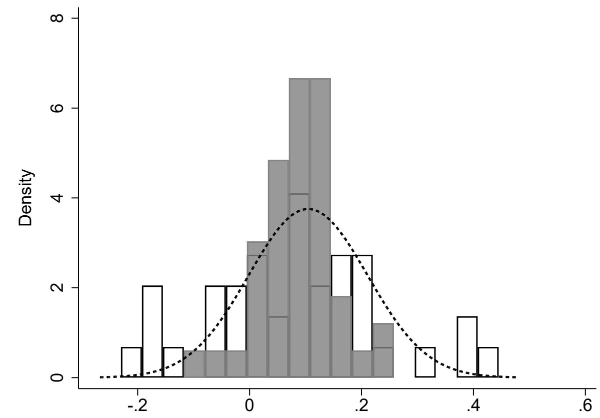
Notes: Panels A-C report patient characteristics (age, total number of Elixhauser comorbidities, and predicted 30-day mortality, respectively) on the *y*-axis against the number of NPs on duty on the *x*-axis. The unit of observation is at the case level. Both the *y*-axis and *x*-axis variables are 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 patient is treated by NPs against the number of patients arriving in the analysis time window (i.e., 8 a.m. to 6 p.m.) of the ED-day cell of the patient'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.7: ED-Specific Estimates of NP Effect

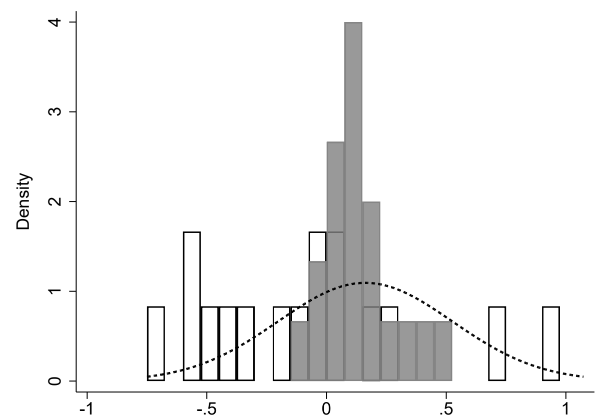
A. Log Length of Stay



B. Log ED Cost

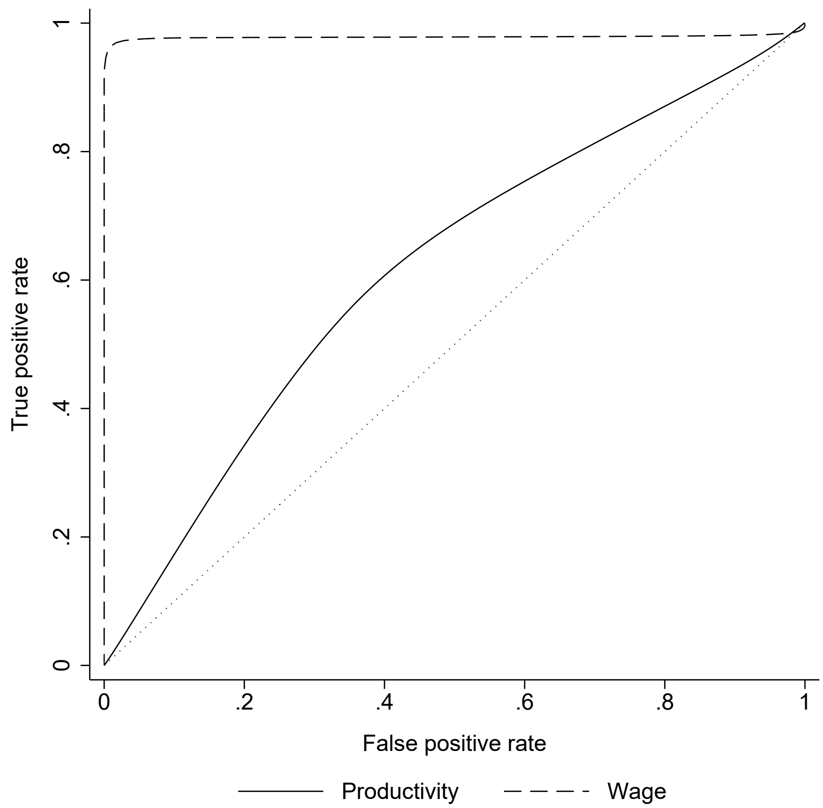


C. 30-day Preventable Hospitalization



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.5). The dashed lines show the standard normal density with a variance of the prior distribution of ED-specific IV estimates for each outcome. Panels A and B display estimates for all 44 EDs in our sample. As 30-day preventable hospitalization is not common (occurs in less than 2 percent of the sample), the estimates are relatively imprecise when using observations from a single ED, Panel C thus includes only EDs with at least 25,000 cases in the analysis sample (in total 20 such EDs).

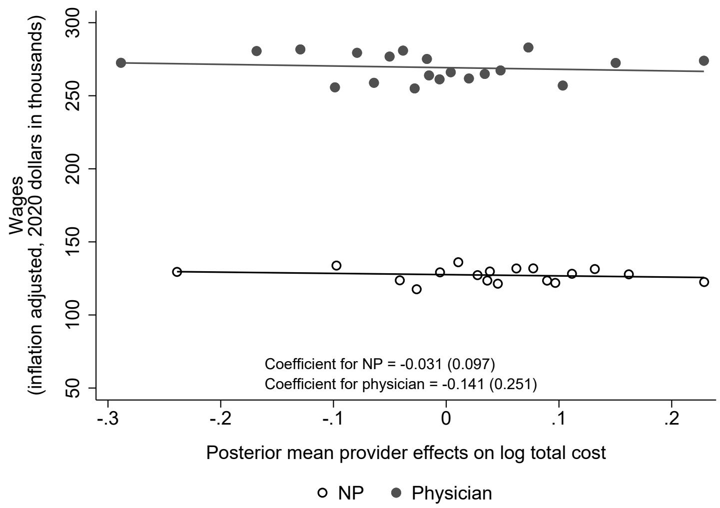
Figure A.8: Receiver Operating Characteristic (ROC) Curve



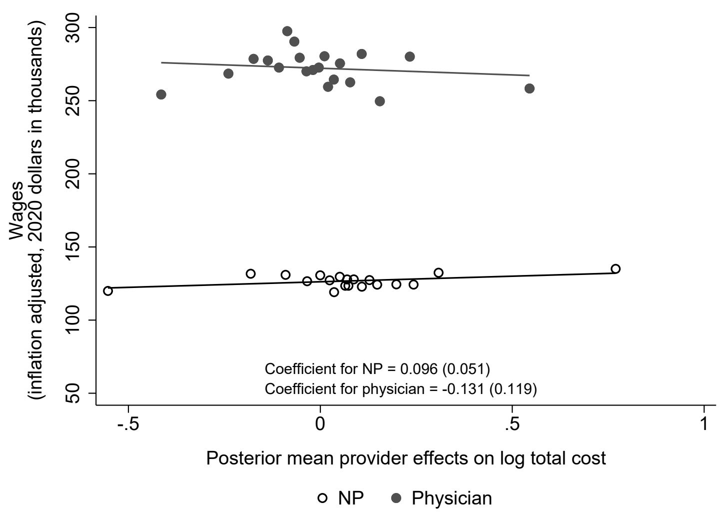
Notes: This figure displays the receiver operating characteristic curves for productivity (in the solid line) and wages (in the dashed line). The dotted line plots the 45-degree line. Productivity is defined as the additive inverse of provider-specific effects on log total spending associated with the ED visit estimated in Appendix A.4.1. Physicians are defined as the "positive" class and NPs are defined as the "negative" class. See Appendix A.4.4 for more details.

Figure A.9: Productivity versus Wages

A. Linear Shrinkage



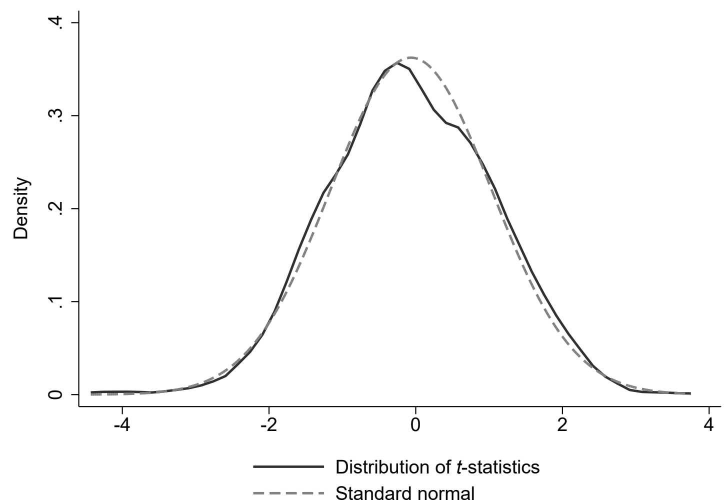
B. Deconvolution Shrinkage



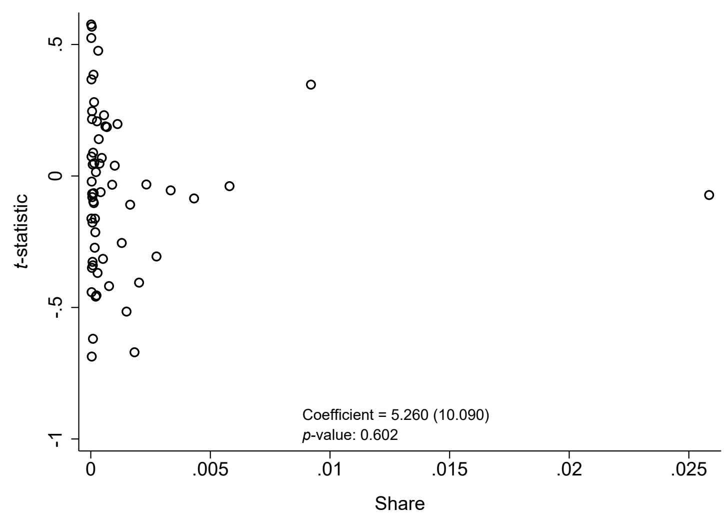
Notes: This figure shows binned scatter plots of provider yearly wage on the y-axis versus posterior mean provider effects on log total cost associated with the ED visit on the x-axis. Both the y-axis and x-axis variables are residualized with respect to ED indicators, with means added back for ease of interpretation. Wages are inflation adjusted to 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. Panel A shrinks provider effects linearly towards the grand mean with weights $w_j = \frac{\hat{\psi}^2}{s_j^2+\hat{\psi}^2}$, where $\hat{\psi}^2$ and s_j^2 are, respectively, the variance of the prior distribution of provider effects and the variance of the sampling error for each provider. Panel B constructs posterior mean provider effects using the deconvolved density. See more details in Appendix A.4.5.

Figure A.10: Diagnosis Coding: NPs versus Physicians

A. Distribution of *t*-Statistics

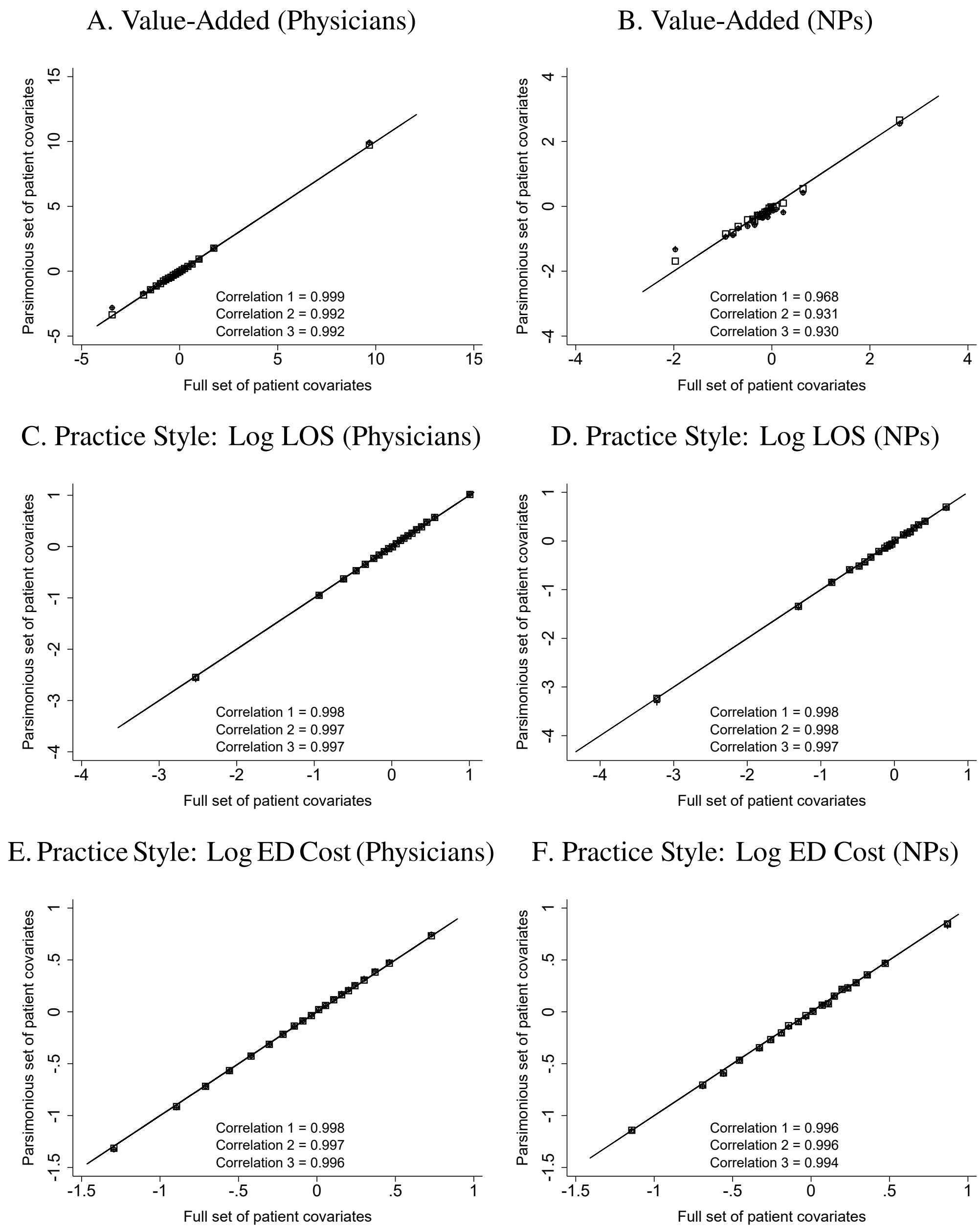


B. *t*-Statistics versus Diagnosis Prevalence



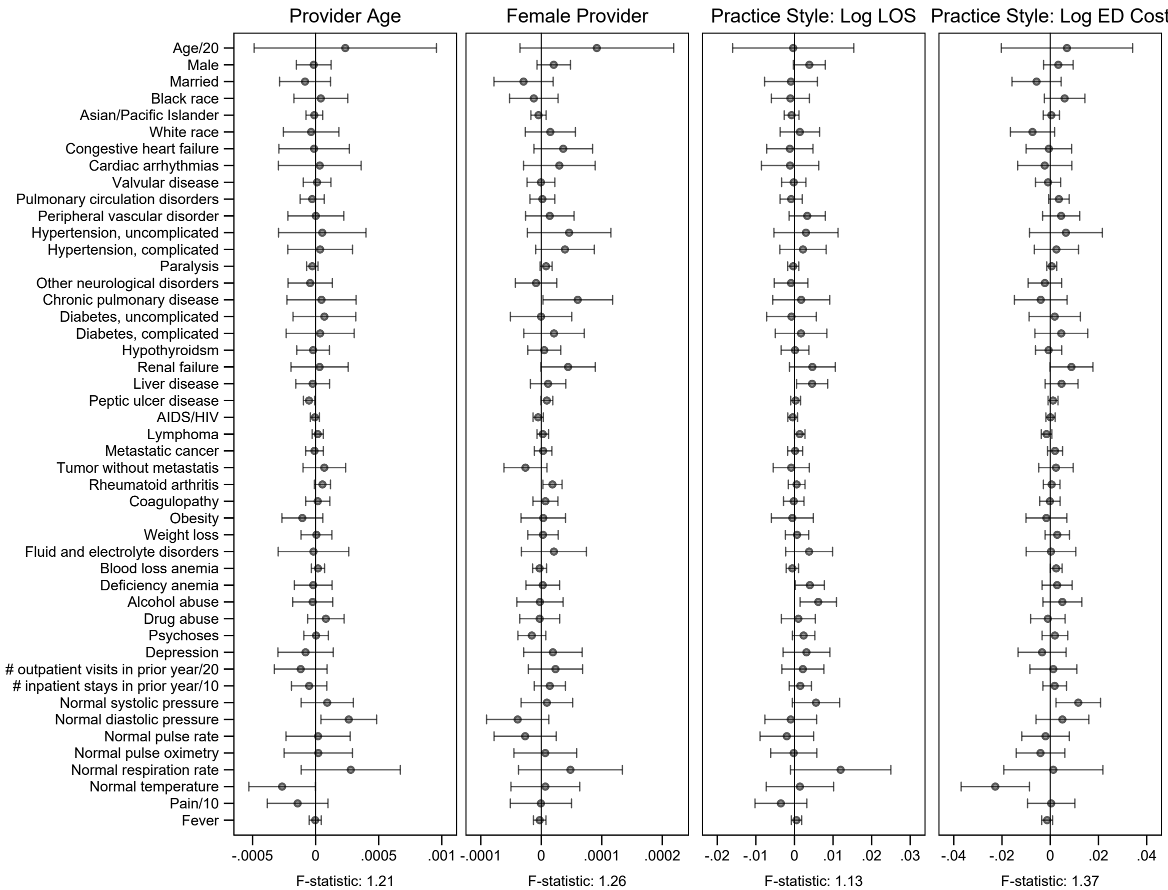
Notes: Panel A plots the distribution of the *t*-statistics on whether NPs and physicians are significantly different in diagnosis coding from 836 separate regressions that use each three-digit diagnosis indicator as the outcome variable. The distribution is estimated using an Epanechnikov kernel with 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 NP, to restrict influences of patient sorting between NPs and physicians). The coefficient from the regression of the *t*-statistics on prevalence is reported in the panel, along with its standard error (shown in parentheses) and *p*-value.

Figure A.11: 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. See Appendix A.3 for details. The *x*-axis in each panel 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 three-digit ICD-10 code of the 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 diagnosis indicators, and 31 Elixhauser comorbidities; parsimonious set 2 that includes demographics and three-digit diagnosis indicators; parsimonious set 3 that includes five-year age-bin and three-digit diagnosis indicators. Valued-added and practice style constructed using parsimonious sets 1-3 are shown in squares, circles, and "+", respectively. 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. Panel B, D and F report results for NPs.

Figure A.12: Balance of Patient Characteristics across On-Duty Provider Characteristics



Notes: This figure shows estimated coefficients and 95% 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 (i.e., indicators for ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day). All average on-duty provider characteristics are case-weighted, with the index case left out. The average on-duty provider characteristics in Panels A-D are, respectively, age, female, 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. At the bottom of each panel, we report 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 control vector. 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 working at the ED) observed in the 20 percent Medicare data. VHA ED NPs' and physicians' mean age and female share reported in Column 1 are slightly different from those reported in Figures 5 and 6 because the latter weight means by the number of ED-days a provider works and patient volume.

Table A.2: Selection of Baseline Sample

Sample step	Description	Cases	Providers		EDs
			NPs	Physicians	
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 do not work outside of the hours of 8 a.m. to 6 p.m. We drop outside of these cases to focus on cases that could be assigned to an NP.	5,766,296	539	5,665	145
3. Restrict EDs to those with NPs, in months with full practice authority.	We restrict the sample to EDs where NPs work. We restrict to months in which these EDs have granted NPs full practice authority.	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 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: Complier and Never-Taker Characteristics

	All	Compliers		Never-takers	
	Mean	Mean	Ratio	Mean	Ratio
Age	62.05	61.11	0.98	63.69	1.03
	(0.15)	(0.31)	[0.98 - 0.99]	(0.17)	[1.02 - 1.03]
Married	0.424	0.424	1.00	0.436	1.03
	(0.004)	(0.008)	[0.97 - 1.04]	(0.007)	[1.00 - 1.06]
Male	0.905	0.905	1.00	0.917	1.01
	(0.002)	(0.003)	[0.99 - 1.01]	(0.002)	[1.01 - 1.02]
Black	0.270	0.262	0.97	0.228	0.84
	(0.011)	(0.019)	[0.83 - 1.11]	(0.015)	[0.73 - 0.95]
White	0.708	0.716	1.01	0.756	1.07
	(0.011)	(0.019)	[0.96 - 1.06]	(0.015)	[1.03 - 1.11]
Asian/Pacific Islander	0.021	0.020	0.95	0.013	0.65
	(0.001)	(0.002)	[0.74 - 1.15]	(0.001)	[0.55 - 0.76]
Outpatient visits in prior year	6.242	5.824	0.93	6.537	1.05
	(0.080)	(0.129)	[0.89 - 0.97]	(0.110)	[1.01 - 1.08]
Inpatient stays in prior year	0.612	0.490	0.80	0.695	1.14
	(0.014)	(0.029)	[0.71 - 0.89]	(0.026)	[1.05 - 1.22]
Elixhauser comorbidity count	3.599	3.324	0.92	3.965	1.10
	(0.030)	(0.066)	[0.89 - 0.96]	(0.041)	[1.08 - 1.12]
Predicted 30-day mortality (%)	1.247	0.902	0.72	1.697	1.36
	(0.032)	(0.067)	[0.62 - 0.83]	(0.049)	[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., indicators for ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day). 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-days with at least one case treated by NPs. Residual shares are constructed by first collapsing the data to ED-days and then residualizing the share of cases treated by NPs by indicators for ED-by-year, ED-by-month and ED-by-day-of-the-week. 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. 95% confidence intervals of each ratio are 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 diagnosis indicators.

Table A.4: Physician Value-Added and Outcomes of Patients Treated by NPs

	Dependent variable						
	Elixhauser comorbidity count	Predicted 30-day mortality	Log length of stay	Log ED cost	Admission	30-day mortality	30-day prevent. hosp.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Physician value-added	-0.006	-0.021	-0.010	-0.005	-0.245	-0.005	-0.023
	(0.013)	(0.014)	(0.008)	(0.005)	(0.199)	(0.075)	(0.044)
Controls	Baseline	Baseline	Full	Full	Full	Full	Full
Mean dep. var.	3.128	0.728	4.302	6.298	7.726	0.633	0.719
S.D. dep. var.	2.711	2.115	1.083	0.870	26.700	7.929	8.446
Observations	147,936	147,936	146,948	146,935	147,936	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, 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.5: NP Presence and Outcomes of Patients Treated by Physicians

	Dependent variable						
	Elixhauser comorbidity count	Predicted 30-day mortality	Log length of stay	Log ED cost	Admission	30-day mortality	30-day prevent. hosp.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Baseline results							
NPs on duty	0.012	-0.013	-0.002	0.001	0.004	-0.022	-0.116
	(0.032)	(0.029)	(0.012)	(0.007)	(0.003)	(0.098)	(0.117)
Panel B: By tercile of case count in ED-day cell							
NPs on duty							
× Bottom two terciles	0.027	0.017	0.027	0.009	0.004	0.023	-0.084
	(0.043)	(0.038)	(0.016)	(0.010)	(0.004)	(0.142)	(0.162)
× Top tercile	0.002	-0.035	-0.017	-0.004	0.003	-0.053	-0.134
	(0.039)	(0.034)	(0.014)	(0.009)	(0.004)	(0.114)	(0.146)
Controls	Baseline	Baseline	Full	Full	Full	Full	Full
Mean dep. var.	3.535	1.070	4.545	6.486	0.154	1.051	1.324
S.D. dep. var.	2.996	2.758	1.276	0.836	0.361	10.200	11.432
Observations	68,863	68,863	68,214	68,208	68,863	68,863	68,863

Notes: This table shows 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-days. The empirical specification takes the form $y_i = \sum_{g=1}^{G} \mathbf{1}(\text{Group}_i = g) \left[\gamma_g \mathbf{1}(Z_i > 0) + \lambda_g\right] + \mathbf{T}_i\eta + \mathbf{X}_i\beta + \varepsilon_i$, 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 or bottom two tercile(s) of all ED-days. The outcomes in Columns 1-7 are, respectively, 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.6: Robustness to Additional Controls

	Dependent variable				
	Log length of stay	Log ED cost	Admission	30-day mortality	30-day prevent. hosp.
	(1)	(2)	(3)	(4)	(5)
Panel A: Baseline					
NP assignment	0.110	0.070	0.103	-0.116	0.252
	(0.045)	(0.030)	(0.585)	(0.115)	(0.120)
Panel B: Control for patient volume					
NP assignment	0.095	0.081	0.597	-0.082	0.237
	(0.045)	(0.030)	(0.606)	(0.116)	(0.118)
Panel C: Control for doctor equivalents (1 NP = 0.341 physicians)					
NP assignment	0.110	0.070	0.103	-0.116	0.252
	(0.044)	(0.029)	(0.584)	(0.115)	(0.120)
Panel D: Control for doctor equivalents (1 NP = 0.5 physicians)					
NP assignment	0.085	0.061	-0.019	-0.104	0.245
	(0.043)	(0.029)	(0.574)	(0.113)	(0.117)
Panel E: Control for wait time					
NP assignment	0.109	0.069	0.317	-0.074	0.250
	(0.045)	(0.030)	(0.594)	(0.124)	(0.121)
Panel F: Control for average risks of patients treated by physicians					
NP assignment	0.100	0.068	0.191	-0.114	0.258
	(0.044)	(0.030)	(0.596)	(0.117)	(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: Panel A repeats our main estimates reported in Tables 2 and 3. Panel B adds a control for patient volume in the analysis time window (i.e., 8 a.m. to 6 p.m.) of the ED-day cell of the patient's visit. Panels C and D add a control for the total number of doctor equivalents on duty at the ED on the day the patient visits. Panel C assumes a substitution rate of 0.341 between NPs and physicians; Panel D assumes a substitution rate of 0.5. Panel E adds a control for patient wait time. As wait time is potentially endogenous (healthier cases could be assigned a lower priority and hence wait longer), we add an instrument for wait time: the average wait time of cases visiting on the same day at the same ED as the index case. Panel F adds a control for the average predicted 30-day mortality risk of patients treated by physicians 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.7: Patient-Provider Gender Match

	Dependent variable						
	Elixhauser comorbidity count	Predicted 30-day mortality	Log length of stay	Log ED cost	Admission	30-day mortality	30-day prevent. hosp.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female NP × male patient	0.007	-0.065	0.027	-0.012	0.355	0.062	-0.058
	(0.103)	(0.100)	(0.017)	(0.015)	(0.593)	(0.085)	(0.093)
Female NP	0.029	0.062	0.006	0.191	1.327	0.027	-0.015
	(0.153)	(0.153)	(0.133)	(0.081)	(1.886)	(0.091)	(0.091)
Male patient	0.743	0.745	-0.051	-0.016	0.262	0.042	0.103
	(0.092)	(0.092)	(0.014)	(0.013)	(0.508)	(0.080)	(0.078)
Controls	Baseline	Baseline	Full	Full	Full	Full	Full
Mean dep. var.	3.190	0.743	4.304	6.341	7.866	0.630	0.745
S.D. dep. var.	2.772	2.145	1.137	0.856	26.921	7.910	8.598
Observations	264,772	264,772	262,960	263,045	264,772	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, 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.8: Alternative Standard Error Clustering

	Dependent variable				
	Log length of stay	Log ED cost	Admission	30-day mortality	30-day prevent. hosp.
	(1)	(2)	(3)	(4)	(5)
Panel A: Clustering by provider					
NP assignment	0.110	0.070	0.103	-0.116	0.252
	(0.045)	(0.030)	(0.585)	(0.115)	(0.120)
Panel B: Clustering by ED-day					
NP assignment	0.110	0.070	0.103	-0.116	0.252
	(0.015)	(0.010)	(0.348)	(0.113)	(0.112)
Panel C: Two-way clustering by ED-day and provider					
NP assignment	0.110	0.070	0.103	-0.116	0.252
	(0.045)	(0.030)	(0.581)	(0.113)	(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. Panel C clusters standard errors using two-way clustering by ED-day 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.9: 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: Include NPs with only one case					
NP assignment	0.102	0.076	-0.011	-0.127	0.285
	(0.045)	(0.030)	(0.594)	(0.116)	(0.125)
Panel B: Leave out the index case					
NP assignment	0.123	0.074	0.198	-0.110	0.260
	(0.046)	(0.031)	(0.605)	(0.121)	(0.126)
Panel C: Leave-out share of cases treated by NPs					
NP assignment	0.117	0.069	0.926	-0.033	0.208
	(0.052)	(0.032)	(0.628)	(0.118)	(0.121)
Panel D: Indicator for any NP on duty					
NP assignment	0.108	0.080	0.185	-0.048	0.211
	(0.049)	(0.030)	(0.643)	(0.121)	(0.130)
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 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 an ED-day cell. Panel B reports results leaving out the index case in measuring the number of NPs on duty. Panels C uses the share of cases treated by NPs in the ED-day cell (leaving out the index case in calculating the share) as the instrument. Panel D uses an indicator for any NP on duty as the instrument. 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.10: Sample Restricted to ED-Days with $Z_i \in \{0, 1\}$

	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	0.084	0.146	-0.042	0.219
	(0.052)	(0.032)	(0.686)	(0.128)	(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 when 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.11: Hospital Admissions and Preventable Hospitalizations Outside VHA

	Dependent variable			
	Admission		30-day prevent. hosp.	
	(1)	(2)	(3)	(4)
	VHA only	VHA+Medicare	VHA only	VHA+Medicare
NP assignment	0.806	0.806	0.485	0.371
	(0.798)	(0.794)	(0.205)	(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.12: Heterogeneous Effects by Provider Experience (Cases in 2018-)

	Dependent variable							
	Log length of stay	Log ED cost	Admission	30-day mortality	30-day prevent. hosp.	Consult	CT	X-ray
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Provider specific experience								
NP assignment	0.081	0.077	-0.268	-0.222	0.335	0.026	0.013	0.021
	(0.046)	(0.032)	(0.646)	(0.142)	(0.149)	(0.011)	(0.007)	(0.011)
NP assignment × experience	-0.060	-0.050	-0.677	0.011	-0.012	-0.017	-0.012	0.006
	(0.025)	(0.021)	(0.308)	(0.044)	(0.033)	(0.007)	(0.004)	(0.009)
Experience	-0.006	0.004	0.204	-0.008	-0.018	-0.009	-0.003	0.007
	(0.005)	(0.009)	(0.303)	(0.017)	(0.012)	(0.005)	(0.001)	(0.002)
Panel B: Provider general experience								
NP assignment	0.104	0.089	-0.356	-0.238	0.347	0.030	0.015	0.021
	(0.048)	(0.034)	(0.700)	(0.146)	(0.152)	(0.011)	(0.008)	(0.011)
NP assignment × experience	-0.130	-0.069	0.249	0.108	-0.029	-0.026	-0.013	-0.005
	(0.067)	(0.040)	(1.337)	(0.114)	(0.077)	(0.013)	(0.013)	(0.008)
Experience	-0.034	-0.009	-0.721	-0.017	-0.039	-0.005	-0.005	0.002
	(0.015)	(0.011)	(0.230)	(0.024)	(0.025)	(0.006)	(0.003)	(0.004)
Full controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	4.637	6.529	16.304	1.251	1.226	0.227	0.150	0.368
S.D. dep. var.	1.133	0.887	36.940	11.114	11.005	0.419	0.357	0.482
Observations	742,968	741,027	747,510	747,510	747,510	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 (despite diagnoses) 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 1-8 are, respectively, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, 30-day preventable hospitalization, whether the patient receives formal consults in the ED visit, whether the patient receives CT scans in the ED visit, and whether the patient receives X-rays in the ED visit. 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.13: Heterogeneous Effects by Provider Experience (Prior-Year Experience)

	Dependent variable							
	Log length of stay	Log ED cost	Admission	30-day mortality	30-day prevent. hosp.	Consult	CT	X-ray
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Provider specific experience								
NP assignment	0.078	0.074	-0.295	-0.221	0.333	0.025	0.012	0.021
	(0.046)	(0.031)	(0.641)	(0.141)	(0.148)	(0.011)	(0.007)	(0.011)
NP assignment × experience	-0.053	-0.055	-0.646	0.016	-0.017	-0.018	-0.012	0.009
	(0.027)	(0.023)	(0.307)	(0.049)	(0.035)	(0.007)	(0.003)	(0.010)
Experience	-0.009	0.009	0.260	-0.012	-0.012	-0.009	-0.002	0.007
	(0.006)	(0.011)	(0.327)	(0.018)	(0.014)	(0.007)	(0.002)	(0.002)
Panel B: Provider general experience								
NP assignment	0.089	0.081	-0.331	-0.222	0.350	0.027	0.013	0.021
	(0.046)	(0.033)	(0.662)	(0.143)	(0.150)	(0.011)	(0.008)	(0.011)
NP assignment × experience	-0.120	-0.063	-0.142	0.040	-0.117	-0.025	-0.014	-0.008
	(0.088)	(0.044)	(1.162)	(0.091)	(0.085)	(0.014)	(0.011)	(0.008)
Experience	-0.040	-0.002	-0.732	-0.008	-0.008	-0.003	-0.004	0.002
	(0.017)	(0.012)	(0.235)	(0.026)	(0.026)	(0.007)	(0.003)	(0.004)
Full controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	4.637	6.529	16.304	1.251	1.226	0.227	0.150	0.368
S.D. dep. var.	1.133	0.887	36.940	11.114	11.005	0.419	0.357	0.482
Observations	742,968	741,027	747,510	747,510	747,510	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 (despite diagnoses) 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 1-8 are, respectively, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, 30-day preventable hospitalization, whether the patient receives formal consults in the ED visit, whether the patient receives CT scans in the ED visit, and whether the patient receives X-rays in the ED visit. 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.14: Heterogeneous Effects by Provider Experience (Measured in Days)

	Dependent variable							
	Log length of stay	Log ED cost	Admission	30-day mortality	30-day prevent. hosp.	Consult	CT	X-ray
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
NP assignment	0.105	0.065	0.225	-0.100	0.263	0.024	0.013	0.019
	(0.045)	(0.029)	(0.623)	(0.117)	(0.123)	(0.009)	(0.007)	(0.009)
NP assignment × experience	-0.031	-0.036	1.098	0.126	0.101	-0.015	0.006	-0.005
	(0.068)	(0.037)	(1.157)	(0.111)	(0.080)	(0.011)	(0.014)	(0.013)
Experience	-0.015	-0.002	-0.403	-0.022	-0.053	0.002	-0.002	0.003
	(0.014)	(0.009)	(0.227)	(0.024)	(0.026)	(0.004)	(0.003)	(0.004)
Full controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	4.608	6.483	16.625	1.247	1.234	0.226	0.145	0.291
S.D. dep. var.	1.161	0.878	37.230	11.099	11.041	0.418	0.352	0.454
Observations	1,110,798	1,108,961	1,118,836	1,118,836	1,118,836	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 1-8 are, respectively, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, 30-day preventable hospitalization, whether the patient receives formal consults in the ED visit, whether the patient receives CT scans in the ED visit, and whether the patient receives X-rays in the ED visit. 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.15: 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.16: Heterogeneous Effects by Patient Characteristics

	Dependent variable				
	Log length of stay	Log ED cost	Admission	30-day mortality	30-day prevent. hosp.
	(1)	(2)	(3)	(4)	(5)
Panel A: Elixhauser comorbidity count					
1st quartile	0.042	0.028	-0.077	-0.147	0.555
	(0.045)	(0.032)	(0.636)	(0.120)	(0.119)
2nd quartile	0.063	0.071	0.291	-0.041	0.438
	(0.044)	(0.030)	(0.642)	(0.126)	(0.132)
3rd quartile	0.117	0.082	-0.245	-0.099	0.250
	(0.048)	(0.031)	(0.761)	(0.178)	(0.186)
4th quartile	0.281	0.122	0.435	-0.203	-0.513
	(0.066)	(0.041)	(1.476)	(0.340)	(0.347)
Panel B: Diagnosis predicted 30-day mortality					
< 95th percentile	0.080	0.064	-0.768	-0.077	0.361
	(0.044)	(0.029)	(0.573)	(0.110)	(0.118)
≥ 95th percentile	0.988	0.247	26.140	-1.253	-2.989
	(0.239)	(0.115)	(7.829)	(2.127)	(1.492)
Panel C: Diagnosis category					
Stroke	1.863	0.651	72.609	3.373	-0.062
	(0.677)	(0.311)	(31.758)	(6.038)	(2.379)
AMI	0.806	1.780	123.007	-11.517	-3.219
	(0.562)	(0.655)	(62.695)	(9.684)	(7.593)
Sepsis	1.480	0.095	44.880	24.533	11.117
	(0.609)	(0.329)	(24.114)	(15.169)	(7.961)
Heart failure	1.125	0.088	20.263	1.262	-11.469
	(0.292)	(0.177)	(8.921)	(3.011)	(5.265)
Other	0.097	0.067	-0.343	-0.147	0.332
	(0.045)	(0.029)	(0.578)	(0.112)	(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: This table shows heterogeneous effects of NPs by patient characteristics described in Section 5.3. 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 condition severity measured by 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. 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.17: 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 variance of provider effects on log total spending associated with the ED visit. Total spending associated with the ED visit is computed as the sum of the three main components of costs that we find significant NP effects: the cost of care at the ED, hospital admission, and 30-day preventable hospitalizations (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.4) and (A.5). To account for biases due to estimation noise in $\hat{\theta}_j$, Row 2 reports variance using a split-sample approach (details are described in Appendix A.4.2). Column 1 reports variance for NPs. Column 2 reports variance for physicians.

Table A.18: Relationship Between z-scores and Standard Errors

	Dependent variable: provider z-score			
	(1)	(2)	(3)	(4)
Provider std. error	0.166	0.203	-0.456	0.085
	(0.434)	(0.348)	(0.385)	(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 in 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. Standard errors clustered by ED are reported in parentheses. The number of unique providers in Columns 1 and 2 are smaller than those reported in Section 2.3 because our deconvolution includes only providers with least 150 cases (to restrict the inclusion of nosiy 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 less than 150 cases in each split sample.