*American Economic Review 2026, 116(8): 2883–2927*
*https://doi.org/10.1257/aer.20241007*

# The Productivity of Professions: Evidence from the Emergency Department[†]

*By* David Chan and Yiqun Chen*

*This paper studies the productivity of nurse practitioners (NPs) and physicians, two professions performing overlapping tasks but with starkly different backgrounds, training, and pay. Using quasi-experimental variation in patient assignment to NPs versus physicians in Veterans Health Administration emergency departments, we find that, on average, NPs use more resources and exhibit a higher 30-day preventable hospitalization rate than physicians. However, the NP-physician performance difference varies by case complexity and severity. Importantly, even larger productivity variation exists within each profession, leading to substantial overlap between the productivity distributions of the two professions; NPs outperform physicians in 38 percent of random pairs.* (*JEL* I11, I18, J24, J44, M53)

Professions play a key role in determining the division of labor and the returns to skilled work (Abbott 2014). In selecting and training future members, professional groups may both restrict the supply of professionals and impact their quality. Professional groups may also act as special interest groups for their members, lobbying policymakers and negotiating with payers. Large differences in pay between professional groups may, therefore, reflect differences in worker productivity or rents from restricted supply or privileged arrangements (Freidson 1974; Shapiro 1986).

Evidence comparing the productivity of distinct classes of professionals performing overlapping tasks remains scant. Professional groups, by nature, act to exclude outsiders from their "jurisdictions" (Abbott 2014). While the medical profession provides a well-documented case study of historical exclusion, it now provides an opportunity for study. Recent decades have witnessed an increased demand for health care outstripping the supply of physicians and the rise of NPs seeking to

*Chan: US Department of Veterans Affairs, UC Berkeley, and NBER (email: david.c.chan@berkeley.edu); Chen: University of Illinois Chicago and NBER (email: yqchen@uic.edu). Authors are listed alphabetically in this paper's byline. Chinhui Juhn was the coeditor for this article. We are grateful to Diane Alexander, Ricardo Alonso, Benjamin Arold, Sandy Black, Kate Bundorf, Marika Cabral, David Card, Stuart Craig, Janet Currie, Shooshan Danagoulian, Qing Gong, Joshua Gottlieb, Mitch Hoffman, Tom Hubbard, Bapu Jena, Amanda Kowalski, Brad Larsen, Darren Lubotsky, Bentley MacLeod, Neale Mahoney, Benjamin McMichael, David Molitor, Jessica Monnet, Ciaran Phibbs, Maria Polyakova, Julian Reif, Michael Richards, Steve Rivkin, Evan Rose, Susan Schmitt, Molly Schnell, Brad Shapiro, Isaac Sorkin, Chris Walters, and many seminar and conference participants for helpful comments and suggestions. Sam Bock, Noah Boden-Gologorsky, Akriti Dureja, Jesse Kozler, Matthew Merrigan, Francis Peng, Jonatas Prates, Aadit Shah, Kemin Wang, Justine Weng, and Saam Zahedian provided excellent research assistance. IRB Approval was obtained from the University of Illinois Chicago, UC Berkeley, and Stanford University.

[†]Go to https://doi.org/10.1257/aer.20241007 to visit the article page for additional materials and author disclosure statement(s).

perform some of the same tasks that physicians do. The number of NPs has reached about one-third of the number of physicians in the United States (Bureau of Labor Statistics 2021a,b), and various states have responded to the shortfall in physicians by liberalizing NP "scope of practice" to perform tasks traditionally performed by physicians (e.g., McMichael and Markowitz 2023). While policy discussions of NPs have largely highlighted their role in primary care, NPs practice in a wide range of areas. Around one-third of NPs work in primary care; the remaining work in various specialty settings.[1]

In this paper, we exploit a quasi-experiment in the Veterans Health Administration (VHA) to study the productivity of NPs versus physicians. In December 2016, the VHA granted full practice authority to NPs, allowing them to practice without physician supervision. We leverage quasi-experimental variation in the availability of physician and NP providers in the emergency department (ED). Nationwide, the share of ED visits seen by NPs reached 13 percent in 2019 (Cairns and Kang 2022); 4 percent of NPs work in emergency care, close to the share of physicians specializing in emergency medicine.[2] In a sample of 1.1 million ED visits, our approach compares patients arriving at the same ED and during similar times (i.e., the year, month, day of the week, and hour of the day) that differ in the number of NPs on duty. We show that the number of on-duty physicians declines with the number of on-duty NPs, and the number of NPs on duty strongly predicts whether an arriving patient will be assigned to an NP versus a physician. Under the plausible assumption that patients arrive quasi-randomly within cells of ED stations and time categories, this instrumental variables (IV) design allows us to study the effect of NPs on patient resource utilization and health outcomes.

The results show that, on average in the ED setting, NPs use more resources during the visit and exhibit a higher 30-day preventable hospitalization rate than physicians. On average, NPs increase patient length of stay by 11 percent and raise the cost of ED care by 7 percent. We do not find NP effects on inpatient admissions in the overall sample, though NPs raise admissions for high-risk cases. For downstream patient outcomes, NPs on average raise 30-day preventable hospitalizations, a frequently used quality-of-care measure, by 20 percent relative to the mean.[3] We do not find statistically significant NP effects on 30-day mortality in the overall sample (95 percent confidence interval: $-0.34$ to 0.11 percentage points). In contrast to IV estimates, ordinary least squares (OLS) estimates for resource use and 30-day preventable hospitalizations are negative in sign, consistent with the descriptive evidence that NPs treat healthier patients than physicians do.

Under various analytical lenses, we uncover mechanisms behind and responses to the average performance difference between NPs and physicians in the ED. First, we find suggestive evidence that experience matters: The NP-physician gap in some

[1] See Measure 3.2 of Milbank Memorial Fund (2024) and Table 5-2 of Health Resources and Services Administration (2010).

[2] According to the National Sample Survey of Registered Nurses (2022), 4 percent of NPs work in emergency care; according to the Association of American Medical Colleges (2021), 5 percent of physicians specialize in emergency medicine.

[3] Preventable hospitalizations are defined using the standardized algorithm developed by the Agency for Healthcare Research and Quality and relied upon by researchers and policymakers to measure quality (e.g., New Jersey Department of Health 2011; Nyweide et al. 2013; California Health and Human Services Agency 2025; Currie and Zhang 2025).

outcomes narrows among providers who have seen more prior patients, both in general and for the diagnosis in question. This indicates that differences in training may play some role in the performance differences between NPs and physicians. Second, we document differences in clinical decision-making. Compared to physicians, NPs are likelier to obtain external information from radiology tests and formal consults. NPs also exhibit prescription patterns consistent with responses to lower skill (Chan, Gentzkow, and Yu 2022): Relative to physicians, NPs are *less* likely to prescribe opioids, which have higher health risks if incorrectly prescribed, but they are *more* likely to prescribe antibiotics, which have higher health risks if incorrectly not prescribed. Despite lower opioid prescription rates, NPs obtain similar downstream rates of opioid use disorder, and despite higher antibiotic prescription rates, NPs exhibit higher rates of subsequent ED visits with infection. Third, we examine heterogeneity in the NP-physician gap by patient condition complexity and severity. The NP-physician gap in patient length of stay, ED cost, and the probability of inpatient admission is smaller for patients with fewer comorbidities and lower severity. Finally, case assignment seems to respond to NPs' relative disadvantage in treating complex cases and absolute disadvantage: On average, NPs are assigned healthier patients of those available for assignment; NPs are also assigned a (modestly) smaller share of patients when the ED is less busy.

The heterogeneity of our results by case complexity and severity complements the earlier literature on the productivity of NPs relative to physicians. Most of the previous studies focus on the primary care setting. With the caveat that those studies feature small samples with possibly underpowered estimates or empirical strategies with limited causal interpretations, our results of smaller NP-physician performance differences in the ED for less complex and less severe cases may align with the earlier literature that indicates greater parity between NPs and physicians in primary care, providing a policy-relevant message that suggests roles for both the NP and physician professions in health care delivery.[4]

We undertake several analyses to assess the validity of our IV quasi-experiment. We show that, conditional on the baseline controls, a broad range of patient characteristics that predict outcomes is well balanced across values of our instrument. Relatedly, our IV estimates are remarkably stable, regardless of the inclusion of a wide set of patient covariates; in contrast, OLS estimates under the full set of covariates are less than half the magnitude (but still opposite-signed relative to IV estimates) of those when only baseline covariates are included. We also assess the

[4] An older medical literature has raised the question of the performance of NPs relative to physicians, but the generally small numbers of providers and other features of these earlier studies limit inference on systematic differences between the two classes of providers. See Laurant et al. (2005) for a systematic review of this literature. The literature features small randomized trials with null results, sometimes comparing a single-digit number of physicians with a single-digit number of NPs. The studies tend to focus on primary care settings and usually have short follow-up times that may be insufficient to detect meaningful effects. According to Laurant et al. (2005), the null findings of this literature "should be viewed with caution given that only one study was powered to assess equivalence of care, many studies had methodological limitations, and patient follow-up was generally 12 months or less." Some recent studies leverage larger-scale data, yet patient selection between NPs and physicians limits the causal inference (e.g., Perloff, DesRoches, and Buerhaus 2016; Lutfiyya et al. 2017; Liu et al. 2020). A notable exception is Currie and Zhang (2025), who find that nonphysician providers—physician assistants and NPs combined—may not perform worse than physicians in primary care in the VHA. Yet the paper's study period, January 2005 to February 2017, is before physician assistants were legally allowed to practice independently at the VHA and mostly before the VHA implementation of NP full practice authority. Thus, nonphysician providers in the paper's setting were largely mandated to practice under physician supervision or oversight.

validity of the exclusion restriction, that the number of NPs on duty is not correlated with other factors that could drive care delivery or patient outcomes. Specifically, we show that our instrument is conditionally unrelated to the characteristics of on-duty physicians and NPs. We further examine potential spillovers between NPs and physicians and find no evidence suggesting such spillovers.[5] Finally, we show that our results are robust to controlling for a series of other factors that may vary with the instrument, including patient volume, total staffing, and patient wait times.

We perform counterfactual analyses to understand the magnitude of the average NP-physician productivity difference, relative to differences between physician and NP wages. We find that compared to the counterfactual scenario of no NPs, allocating one-quarter of ED patients to NPs increases nonwage spending by \$197 million per year to the VHA. Accounting for lower NP wages that are on average half those of physicians, it again implies a net cost of \$129 million per year. However, the cost implications of NPs decline with case complexity. Based on IV estimates, using NPs to staff the least-complex quarter of cases (based on Elixhauser comorbidities) results in net spending that is only about one-fifth of the amount for assigning a quarter of cases at the average complexity.

Finally, we examine variation in productivity across providers within the professional classes of NPs and physicians. To arrive at provider-specific measures of productivity, we estimate a just-identified IV model, in which we instrument patient assignment to specific providers by indicators for on-duty providers. Using a method developed by Efron (2016) and adapted by Kline, Rose, and Walters (2022), we deconvolve the estimates of provider-specific productivity into underlying productivity distributions for each of the two professional classes. We find wide variation in productivity within professions and substantial overlap between the productivity distributions of NPs and physicians. The probability that a randomly chosen NP is more productive than a randomly chosen physician can be as large as 38 percent. Within each professional class, the productivity distribution implies greater medical spending of around \$900,000 per year under a provider at the twenty-fifth percentile of productivity than under a provider at the seventy-fifth percentile—about three times the mean annual spending difference between NPs and physicians. That is, despite the stark differences in training and selection, there is a sizable productivity overlap between NPs and physicians. While training and selection are relatively uniform within professions, productivity variation within professions is even larger than that between professions.

We extend our distributional analysis of productivity to examine the complexity of cases assigned to and the wages paid to individual providers. While productivity differences are much larger within each profession than the average difference between NPs and physicians, within each profession, a provider's productivity shows a limited relationship with her wages or the complexity of her assigned cases. This starkly contrasts with differences in case assignment and wages between professions: EDs assign noticeably less complex cases to NPs—enough to reverse OLS estimates

[5] Specifically, we examine whether NP presence may affect physician performance. E.g., NPs asking physicians for assistance could slow down physicians. We thus examine whether physicians' treatment outcomes change with the presence of NPs; we find no such evidence. We also examine whether on-duty physician quality impacts NP performance; we find that the outcomes of patients treated by NPs are unrelated to the value-added of the physicians on duty.

for key outcomes relative to the corresponding IV estimates—and pay NPs half the wages of physicians. In other words, a provider's productivity is far from a sufficient statistic explaining her pay or determining her case assignment; professional class is a much stronger predictor of both wages and assigned case complexity.

Our findings contribute to several strands of literature. First, given the dramatic rise in the supply of NPs to meet the growing demand for health care, policy debates have arisen around the NP provision of care. A recent body of research has examined the effects of liberalizing state "scope of practice laws" for NPs.[6] By design, these papers study the general-equilibrium impacts of both allowing NPs greater scope to practice and increasing the supply of providers; results will depend on how labor is reallocated between professions. Our study sheds light on the effect of assigning patients to NPs versus physicians. The widely heterogeneous complexity and severity of cases in our setting also provide a broader picture of the productivity of NPs and physicians. The heterogeneity of our results by patient type suggests returns to optimal case assignment for NPs. Also, the substantial overlap between the productivity distributions of NPs and physicians, even in the ED, may generate important policy implications supporting the use of NPs in less complex settings.

Strikingly, we find that productivity variation within the NP and physician professions is much larger than the average productivity difference between the two professions. This finding puts the prior literature on practice variation (e.g., Epstein and Nicholson 2009; Gowrisankaran, Joiner, and Léger 2023) in context: In our setting, the interquartile range in productivity within professions is severalfold larger than the overall productivity difference between two professions with starkly different selection and training processes. In contrast to the assignment of cases between NPs and physicians, in which less complex cases are, on average, assigned to NPs, we find essentially no matching between case complexity and provider productivity within professions. This suggests substantial informational and organizational barriers to such "skill-task matching" within professions, though professional boundaries may provide a mechanism for this matching between professions (Acemoglu and Autor 2011).

Second, our research relates to the widespread practice of occupational licensing, affecting about a third of all jobs in the United States (Kleiner and Krueger 2013). The existing literature suggests that occupational licensing increases the earnings of licensed workers (Kleiner and Krueger 2013; Kleiner, Marier, Park, and Wing 2016; Farronato, Fradkin, Larsen, and Brynjolfsson 2020) but provides little evidence on whether higher earnings arise from restricting the supply of workers or from improving the quality of their work in modern settings (see Farronato, Fradkin, Larsen, and Brynjolfsson 2020 for a review).[7] Studies in this literature compare

[6] The findings of this literature are varied and somewhat mixed. Perry (2009), Kleiner, Marier, Park, and Wing (2016), and Dillender, Sasso, Phelan, and Richards (2022) find that these laws affect NP earnings. Stange (2014) finds a minimal impact of greater NP supply on utilization, access, or prices but perhaps a moderate impact on primary care utilization. Yet Traczynski and Udalova (2018) find increases in utilization and some evidence of increased quality, and Alexander and Schnell (2019) find evidence of better access and outcomes in mental health.

[7] Two studies of an earlier, unregulated environment of midwifery, near the beginning of the twentieth century, demonstrate meaningful reductions in maternal and infant mortality with the initial implementation of occupational licensing (Lazuka 2018; Anderson, Brown, Charles, and Rees 2020). In the education literature, Carrell and West (2010) find that (mostly tenure-track) professors with doctorate degrees produce better long-term student outcomes than (mostly adjunct) professors without such degrees.

workers within professions (along the margin of occupational licensing), while ours compares two competing professions with important differences beyond licensing.

A third related literature is concerned with worker human capital and productivity. These issues have received growing attention in health care (e.g., Doyle, Ewer, and Wagner 2010; Currie and MacLeod 2017, 2020; Chen 2021; Chan, Gentzkow, and Yu 2022) and more broadly (e.g., Gennaioli, La Porta, Lopez-de-Silanes, and Shleifer 2013; Chetty, Friedman, and Rockoff 2014). Our study contributes to this literature by comparing the productivity of distinct professional classes with starkly different selection and training. In the rich ED setting, we also uncover key mechanisms connecting productivity to human capital along dimensions of experience, information gathering, decision-making, and case complexity. Our results suggest that professional selection and training may give rise to important productivity implications, though these implications may be small relative to the variation within professions.

Fourth, a broad set of questions concerns the distribution of wealth in society across occupations and strata of educational attainment. In recent decades, societies have witnessed an increased concentration of wealth in occupations associated with high human capital (Smith, Yagan, Zidar, and Zwick 2019). Training to reach the highest levels of income has become increasingly competitive among the upper class, while the middle and lower classes are increasingly left behind, characterizing a "meritocracy trap" (Markovits 2020). Interestingly, our results suggest a productivity difference between professions at least as large as their wage difference, at least in our resource-intensive and information-dependent setting within health care. Yet the lack of relationship between worker-specific productivity and wages within professions suggests frictions in observing or contracting by actual productivity across similar workers (e.g., Acemoglu and Pischke 1998). Entering a profession may represent a costly and imperfect way for workers to distinguish themselves.

The remainder of this paper is organized as follows. Section I describes the institutional setting and the data. Section II describes our empirical approach and provides evidence for its validity. Section III presents the average performance difference between NPs and physicians and evidence on mechanisms and responses. Section IV presents analyses on policy-relevant counterfactual scenarios. Section V reports distributions of productivity, case assignments, and wages within professions and examines how productivity correlates with case assignments and wages. Section VI concludes.

## I. Background and Data

### A. *NPs and Physicians in the United States*

To understand the physician and NP professions in the United States, it is instructive to consider their distinct origins in the American context. According to the landmark work by Starr (1982, p.28), "among the professions, medicine is both the paradigmatic and the exceptional case: paradigmatic in the sense that other professions emulate its example; exceptional in that none have been able to achieve its singular degree of economic power and cultural authority." Aided by scientific advances and demographic shifts, the US medical profession in the early twentieth century captured authority by standardizing education and licensing toward a scientific orientation, in the process excluding a large swath of practitioners (Brown 1979; Larson 1979).[8]

The "professionalization" of American nursing began in the 1870s with the establishment of the first three nursing schools in the United States. In contrast to the scientific orientation of the medical profession, the driving force behind the nursing profession was to install (female) staff in hospitals to improve hygiene and cleanliness (Ashley 1976). NPs emerged from the nursing tradition in the 1960s, in the setting of increased specialization in medicine, worsening access to care, and new federal funding from Medicare and Medicaid to increase the training of providers (Fairman 2009; Hallett 2016). Over the next few decades, pressures to contain costs and increase throughput further expanded the boundaries of NP practice in a variety of settings (e.g., primary care, emergency care) (Fairman 2009; Kleinpell, Cook, and Padden 2018). While NPs have historically focused on the provision of primary care, which is still the largest area employing NPs, NPs now work in various specialty settings. Only around one-third of NPs worked in primary care between 2016 and 2021, approximately the share of physicians working in primary care (Milbank Memorial Fund 2024). Anecdotal evidence suggests that some of the same pressures that have lured physicians into specialty care have also begun to take hold among NPs (Andrews and Beard 2024).

Present-day NPs and physicians remain starkly different regarding training and income, despite performing overlapping tasks in many settings. Admission rates to medical schools are only about one-tenth of admission rates to graduate nursing programs,[9] and the number of years of training for physicians is around twice that for NPs.[10] Undergoing a highly selective process and long training periods, physicians constitute the most common profession in the top percentile of the income distribution (Gottlieb et al. 2020); the income of NPs is roughly half that of physicians (Bureau of Labor Statistics 2021a,b).

### B. *VHA and ED Setting*

In December 2016, the VHA granted full practice authority to NPs. The policy enables NPs to practice without requiring physician supervision at VHA facilities.

[8] Much of this transformation centered around the Flexner Report in 1910, which strongly advocated for a scientific orientation of medical education and the exclusion of alternative practices (Flexner 1910; Beck 2004). Following the report, more than half of medical schools closed or consolidated (Patel and Rushefsky 2004).

[9] We searched https://www.petersons.com/graduate-schools.aspx for admission rates to graduate nursing programs and to medical schools in the following universities with both graduate nursing programs and medical schools: Columbia University, Duke University, Emory University, Johns Hopkins University, the University of North Carolina, the University of Pennsylvania, and the University of Washington. Admission rates ranged from 25 to 63 percent for graduate nursing programs and from 3 to 7 percent for medical schools.

[10] According to the Association of American Medical Colleges (2020), physicians must complete a four-year undergraduate degree, a four-year Doctor of Medicine degree, and three to seven years of residency training. Some physicians further undergo one to three years of fellowship training after residency. According to the American Association of Nurse Practitioners (2020), NPs must complete a four-year Bachelor of Science in Nursing degree and may choose between one to two years in a Master of Science in Nursing degree or three to four years in a Doctor of Nursing Practice degree. There is no residency or fellowship training requirement for NPs.

NPs can treat patients independently, regardless of state restrictions that would otherwise limit NPs' practice authority.[11]

Several features make the ED a setting well suited to studying provider productivity. First, each patient visit is generally assigned to a single ED provider. Such independence allows us to attribute patient outcomes to individual ED providers. Second, patient flow in the ED is highly unpredictable, while provider schedules are typically set well in advance. Variation in NP availability is thus unrelated to the set of patients arriving. Third, the ED is an important setting that uses the NP workforce. Across the nation in 2019, 13 percent of ED visits were seen by NPs (Cairns and Kang 2022).[12] In the VHA, the share of ED visits seen by NPs was 11 percent in 2019—about half the share of visits seen by NPs in primary care at around 20 percent in the VHA (Morgan, Abbott, McNeil, and Fisher 2012).[13] Nationwide, around 4 percent of NPs practice in emergency care, close to the share of physicians specializing in emergency medicine at 5 percent (Association of American Medical Colleges 2021; National Sample Survey of Registered Nurses 2022). Fourth, patients present at the ED with a wide spectrum of conditions, ranging in complexity and severity, which provides an opportunity to investigate productivity across a range of tasks.

In Supplemental Appendix Table A.1, we report the characteristics of NPs and physicians working at VHA and non-VHA EDs. NPs in VHA EDs are representative of their non-VHA counterparts in female share (about 80 percent), while they appear to be more experienced (as indicated by age, 51 versus 43 years old). Among ED physicians, those practicing at the VHA are slightly more likely to be female than those outside of the VHA (34 versus 27 percent), but the two groups have a similar average age (48 versus 46 years old). Supplemental Appendix Table A.1 also compares NPs practicing at the ED and the overall NP workforce, showing that ED NPs are representative of the NP workforce in age and gender.

## C. *Data*

Our main data are administrative health records from the VHA, the largest integrated health care system in the United States, serving more than 9 million veterans. For each ED visit, the data record the type of provider treating the patient (i.e., NP or physician), resource utilization, and patient outcomes. The data also contain detailed information on patient characteristics (e.g., demographics, comorbidities, vital signs) and provider characteristics (e.g., birth date, gender). The large sample size and the availability of detailed information (e.g., time stamps for measuring ED length of stay, provider orders) enable a comprehensive analysis of NP- and physician-provided care. Finally, we access detailed payroll records for providers to impute wages.

*Sample Construction.*—We restrict our analysis sample in the following ways. First, we restrict the sample to ED visits between January 2017 and January 2020,

[11] As of 2021, about half of the states in the United States had not granted NPs full practice authority (American Association of Nurse Practitioners 2021).

[12] About 6 percent of ED visits nationwide were seen by NPs without physicians; 7 percent were seen by NPs with physicians.

[13] We estimate the share of ED visits seen by NPs from the VHA data, which is described in the following section.

that is, after full practice authority was granted to NPs at the VHA and before the onset of the COVID-19 pandemic in the United States. Second, we include only cases arriving during the daytime (8 AM to 6 PM) because the data show that few NPs take evening or night shifts.[14] Third, we focus on visits to VHA EDs in months when the ED used NPs to treat patients. Local VHA facilities varied in whether they used NPs in the ED.[15] Fourth, to examine a single margin between NPs and physicians, we exclude EDs that used nonphysician providers other than NPs (mainly physician assistants).[16] Finally, we drop a small number of cases with missing age or gender or aged under 20 or above 99 years. Supplemental Appendix Table A.2 reports the sample size at each step. The final sample contains 1.1 million cases, over 44 EDs.[17]

*Outcome Variables.*—To measure medical resource use, we include two primary outcomes that are frequently used in the ED setting: (i) the patient length of stay (i.e., the time between patient assignment to the provider and patient discharge) and (ii) the cost of care during the ED visit (excluding costs due to a resulting hospital admission, measured separately next).[18] We also include hospital admission as a resource use measure, indicating the provider's decision to admit the patient for inpatient care during the ED visit. To measure the quality of care, we examine two prominent patient outcomes: We use linked death records to construct indicators of patient 30-day mortality and use linked inpatient data to construct indicators of 30-day preventable hospitalization as defined by the Agency for Healthcare Research and Quality (2021). We measure 30-day preventable hospitalization after ED discharge; thus, it is beyond the decision of the provider in the index ED visit.

In examining mechanisms behind the effect of NPs, we include the following sets of outcomes: (i) whether the provider orders consults (which are typically from specialists outside of the ED); (ii) whether the provider orders CT scans and X-rays, two primary diagnostics in the ED; (iii) whether the provider prescribes opioids and antibiotics, two major classes of drugs whose clinical indications for appropriate use are often unclear and can require skill to discern (e.g., Fleming-Dutra et al. 2016; Neuman, Bateman, and Wunsch 2019); and (iv) whether the patient has return visits with infections and subsequent opioid use disorder.

[14] One possible explanation is that, since patient volumes on average are much lower in the evening/night than in the daytime (e.g., the average number of cases arriving per hour is 3.6 between 8 AM and 6 PM versus only 0.9 outside of 8 AM–6 PM), EDs in our sample often staff only one provider per shift in the evening/night. Our interview with VHA ED leadership revealed that EDs are required to have at least one physician on duty at all times.

[15] We define a VHA ED as using NPs in a month if it has at least 15 cases treated by NPs in the month. The sample size changes only slightly when we use alternative thresholds: e.g., the sample size is 1.13 and 1.10 million when using a threshold of 10 and 20, respectively, compared to 1.12 million based on a threshold of 15.

[16] In addition, unlike NPs, physician assistants had not been allowed to practice independently at the VHA as of 2022 nor in any state as of mid-2019. In Section IIIF, we expand the sample to cases in all EDs that use NPs despite the use of physician assistants but exclude cases in ED-day cells with physician assistants to focus on the margin between NPs and physicians. This analysis expands the sample from 1.1 million cases across 44 EDs seen by 1,348 physicians and 156 NPs in our baseline analysis to 2.2 million cases across 110 EDs seen by 3,499 physicians and 491 NPs. We find similar results in this expanded sample.

[17] Supplemental Appendix Table A.3 shows that EDs included in our sample are a representative set of all VHA EDs, as measured by ED, provider, and patient characteristics. The EDs in our sample include those in large metropolitan areas (e.g., New York City, Los Angeles, the Bay Area) as well as those in smaller areas (e.g., Detroit, Michigan, and Madison, Wisconsin).

[18] For length of stay, we use detailed time-stamped data on patient assignment and discharge. For the cost of ED care, we use cost accounting by the VHA, which measures resource utilization.

Table 1—Characteristics of Baseline Sample

| | All | Treated by NP | Treated by physician | $p$-value |
|---|---|---|---|---|
| Age | 62.05 | 60.72 | 62.46 | 0.00 |
| | [15.80] | [15.87] | [15.75] | |
| Married | 0.424 | 0.424 | 0.424 | 0.80 |
| | [0.494] | [0.494] | [0.494] | |
| Male | 0.905 | 0.904 | 0.906 | 0.00 |
| | [0.293] | [0.295] | [0.292] | |
| Black | 0.270 | 0.271 | 0.270 | 0.12 |
| | [0.444] | [0.445] | [0.444] | |
| White | 0.708 | 0.705 | 0.709 | 0.00 |
| | [0.455] | [0.456] | [0.454] | |
| Asian/Pacific Islander | 0.021 | 0.021 | 0.020 | 0.04 |
| | [0.142] | [0.144] | [0.142] | |
| Outpatient visits in prior year | 6.242 | 5.658 | 6.423 | 0.00 |
| | [7.284] | [6.361] | [7.538] | |
| Inpatient stays in prior year | 0.612 | 0.431 | 0.668 | 0.00 |
| | [1.543] | [1.249] | [1.620] | |
| Elixhauser comorbidity count | 3.599 | 3.190 | 3.726 | 0.00 |
| | [3.018] | [2.772] | [3.079] | |
| Length of stay (minutes) | 162.09 | 119.53 | 175.29 | 0.00 |
| | [172.48] | [131.28] | [181.38] | |
| ED cost ($, inflation-adjusted to 2020) | 939 | 813 | 978 | 0.00 |
| | [1,331] | [1,010] | [1,413] | |
| Inpatient admission (%) | 16.62 | 7.87 | 19.34 | 0.00 |
| | [37.23] | [26.92] | [39.50] | |
| 30-day preventable hospitalization (%) | 1.23 | 0.75 | 1.39 | 0.00 |
| | [11.04] | [8.60] | [11.69] | |
| 30-day mortality (%) | 1.25 | 0.63 | 1.44 | 0.00 |
| | [11.10] | [7.91] | [11.91] | |
| Observations | 1,118,836 | 264,789 | 854,047 | |

*Notes*: Column 1 shows average characteristics of all cases in our analysis sample. Columns 2 and 3 show average characteristics of cases treated by NPs and physicians in the sample, respectively. Standard deviations are reported in brackets; $p$-values of $t$-tests for the equivalence of means between cases treated by NPs and by physicians are shown in the last column.

*Descriptive Statistics.*—Table 1 summarizes the characteristics of the cases included in our analysis. In addition to demographics, we measure patient comorbidities as Elixhauser indices (Elixhauser, Steiner, Harris, and Coffey 1998), which are 31 indicators for comorbidities (e.g., cancer, diabetes) based on patient medical histories in the prior 365 days. We also report the average length of stay, average ED spending (inflation-adjusted to 2020 dollars), and 30-day preventable hospitalization rate. Column 1 shows the characteristics of the overall sample. Columns 2 and 3 compare cases treated by NPs with those treated by physicians. Along several dimensions, cases treated by NPs are healthier than those treated by physicians: Cases treated by NPs are younger (60.7 versus 62.5 years old), have fewer Elixhauser comorbidities (3.2 versus 3.7), and have fewer outpatient visits and fewer inpatient stays in the prior 365 days (5.7 versus 6.4 outpatient visits and 0.4 versus 0.7 inpatient stays). Consistent with selection, cases treated by NPs appear to have better outcomes: They have a shorter average length of stay (120 versus

175 minutes), a lower average ED cost (\$813 versus \$978), and a lower 30-day preventable hospitalization rate (0.8 versus 1.4 percentage points).

## II. Empirical Strategy

An ideal experiment to assess the effect of being treated by NPs would randomly assign cases to NPs and physicians. Lacking random assignment, we use a quasi-experimental approach: We leverage plausibly exogenous variation in the availability of NPs on duty to instrument for whether an NP or a physician treats a case. In this section, we begin by describing our instrumental variables approach. We then show evidence supporting the validity of our identification strategy.

### A. *Specification*

Our empirical specification is a two-stage least squares (2SLS) model that takes the following form:

$$(1) \qquad y_i = \delta NP_i + \mathbf{T}_i \eta + \mathbf{X}_i \beta + \varepsilon_i,$$

$$(2) \qquad NP_i = \lambda Z_i + \mathbf{T}_i \zeta + \mathbf{X}_i \gamma + v_i,$$

where $i$ denotes a case, $y_i$ is the outcome of interest, and $NP_i$ indicates whether case $i$ is treated by an NP. We use $Z_i$ to denote the instrument, that is, the number of NPs on duty between 8 AM and 6 PM (our analysis time window) at the ED on the day case $i$ visits.[19] The parameter of interest is $\delta$, which represents a local average treatment effect (LATE): the average causal effect among cases that would have been assigned to a different class of provider under a different number of NPs on duty. In Section IIIF, we include alternative instruments, including the share of on-duty providers who are NPs, the share of cases assigned to NPs (leaving out the index case), and an indicator for any NPs on duty in the ED-day cell.

The vector $\mathbf{T}_i$ encodes interactions between indicators for the ED and indicators for time categories of the patient's arrival. Specifically, $\mathbf{T}_i$ contains ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day indicators. We condition on $\mathbf{T}_i$ to allow for the sorting of NPs across shift types (e.g., weekdays versus weekends) and EDs, where patient characteristics and ED conditions may be systematically different. Controlling for ED-by-time-category indicators captures these potential systematic differences.[20]

[19] Since the data do not include direct information on provider scheduling, we measure $Z_i$ as the number of NPs treating cases arriving during the analysis time window in the ED-day cell of case $i$'s visit. We count an NP as on duty if she is observed treating at least two cases between 8 AM and 6 PM in the ED-day cell. The main analysis includes the index case in calculating $Z_i$. In Section IIIF, we show the robustness of our estimates to alternative measurements of $Z_i$, including (i) one that includes NPs with only one case in a shift and (ii) one that leaves out the index case in defining whether an NP is on duty.

[20] While fixed effects for ED-by-hour-of-the-day interactions are not necessary to condition on to yield quasi-random variation in the instrument (since the instrument varies at the day instead of hour level), we include them for statistical precision of estimates. Yet this inclusion does not meaningfully affect the statistical significance of our results.

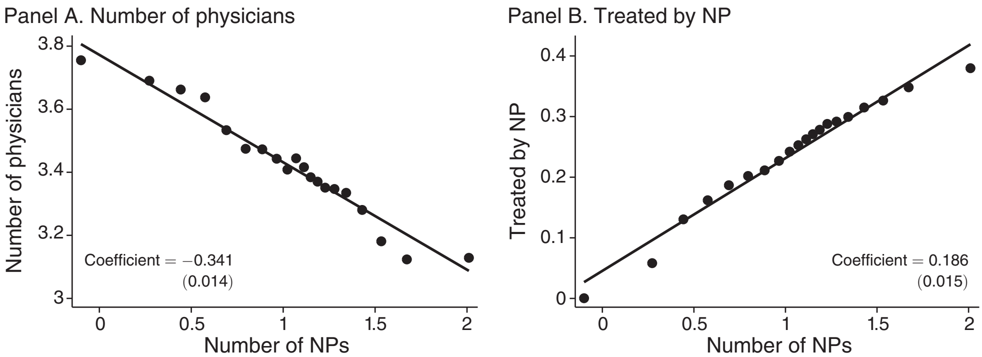


Figure 1. First Stage

*Notes*: This figure represents a graphical illustration of the first-stage estimation. Panel A shows a binned scatterplot of the number of physicians on duty versus the number of NPs on duty. Panel B shows a binned scatterplot of whether the case is treated by an NP versus the number of NPs on duty. To construct these binned scatterplots, we first 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) and then add means back for ease of interpretation. The coefficients report the estimated slope of the best-fit line between the $y$-axis and $x$-axis variables (conditional on the baseline control vector), with standard errors clustered by provider reported in parentheses.

As robustness checks, our specification also includes a vector of patient covariates $\mathbf{X}_i$, including indicators for five-year age bins, marital status, gender, and race (White, Black, and Asian/Pacific Islander, with other racial categories omitted as the reference group); 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 (pulse, respiratory rate, blood oxygen level, pain level, body temperature, an indicator for fever, systolic blood pressure, and diastolic blood pressure); and indicators for the three-digit ICD-10 code of the primary diagnosis of the visit.[21] For each patient covariate with missing values, we add an indicator for missing values and replace the missing values with zero. Finally, $\varepsilon_i$ and $v_i$ are error terms. We cluster standard errors by provider. We also consider alternative clustering approaches in Section IIIF.

## B. *Identification*

To interpret $\delta$ as the LATE of being treated by NPs, our IV approach requires four identifying assumptions: relevance, conditional independence, exclusion, and monotonicity. In this section, we summarize empirical evidence supporting the validity of the identifying assumptions.

*Relevance.*—Figure 1 shows the first stage of our IV model, controlling for the baseline controls $\mathbf{T}_i$. Panel A shows that the number of physicians on duty declines

[21] Since our study period is from January 2017 to January 2020, all disease diagnoses are coded in ICD-10 in the data. A potential question is whether the three-digit diagnoses are endogenous to being treated by NPs. Yet, as shown in Section III, our estimates are remarkably stable regardless of whether we control for three-digit diagnosis indicators. In Supplemental Appendix A.1, we also show that NPs and physicians are similar in their coding of three-digit diagnoses.

linearly with the number of NPs on duty. Consequently, panel B shows that a patient's probability of being treated by an NP increases with the number of NPs on duty: One more NP on duty increases her probability of being treated by an NP by 18.6 percent. The increase is highly significant (with an $F$-statistic of 149.2, even conditioning on $\mathbf{T}_i$) and is close to linear.

To provide context, Supplemental Appendix Figure A.1 presents a histogram of the number of NPs on duty across ED-day cells. The figure reveals a fair spread in the number of NPs on duty: 38.1, 47.2, and 11.3 percent of ED-day cells have zero, one, and two NP(s) on duty, respectively; 3.4 percent of ED-day cells have more than two NPs on duty. An important question is what drives the variation in the number of NPs on duty. As shown in Supplemental Appendix Figure A.2, NPs are less likely to work on weekends: 77 percent of weekday ED-day cells versus 24 percent of weekend ED-day cells have NPs on duty. However, conditional on the day of the week and other time categories (year and month), we still observe a large variation in the number of NPs on duty within EDs across days (standard deviation of 0.49).

Interviews and surveys with VHA ED providers and administrators point to sources of random variation in scheduling.[22] As EDs have long open hours (typically 24/7), EDs rotate providers on duty. Providers' on-duty schedules are generally set months in advance, restricting the possibility of arranging on-duty providers based on specific conditions in the ED on the day of occurrence, e.g., patient volumes or diseases of patients arriving. Each ED has a staffing model that prespecifies the staffing level for each type of shift defined by the time categories (matching $\mathbf{T}_i$) to meet patient care needs predicted by the time categories. In general, conditional on $\mathbf{T}_i$, there is no systematic variation in the characteristics of the staff scheduled to be on duty.[23] It is also extremely rare that EDs call in providers who are not scheduled to be on duty to care for patients.

In the empirical evidence below, we show that patient characteristics, as well as on-duty physicians' and on-duty NPs' characteristics, are well balanced across the number of NPs on duty (conditional on $\mathbf{T}_i$), consistent with quasi-random variation in the instrument. Otherwise, e.g., if the variation is driven by patient or staffing shocks, we would observe systematic changes in patient and/or on-duty physicians' and NPs' characteristics with the instrument.

In Supplemental Appendix Figure A.3, we decompose the variance of NP staffing. The figure shows that $\mathbf{T}_i$ explains around 70 percent of the variation in NP staffing, consistent with the notion that NP availability varies across EDs and time categories. Conditional on $\mathbf{T}_i$, a wide range of factors, including arriving patients' characteristics, on-duty physicians' characteristics, and a series of other factors that may vary across days (e.g., patient volumes), have virtually no explanatory power for the number of NPs on duty, further supporting quasi-random variation in NP staffing conditional on $\mathbf{T}_i$.

[22] We spoke with eight ED physicians and NPs about scheduling; we also distributed a survey to VHA ED administrators about scheduling and received responses from VHA ED administrators in seven states.

[23] Per our interviews and surveys with ED providers and administrators, while there may be revisions to on-duty schedules, revisions are infrequent. Further, revisions, if any, are mostly for providers' unexpected personal reasons (e.g., unanticipated family events) that are arguably unrelated to the patients arriving or other conditions at the ED and are typically done in advance of the shifts, making revisions plausibly exogenous.

*Conditional Independence.*—For our instrument to be valid, the number of NPs on duty must be uncorrelated with patient potential outcomes, conditional on the baseline controls $\mathbf{T}_i$. The institutional features described above support the conditional independence assumption. Further, two sets of empirical evidence provide strong support for this assumption. First, we show that patient observed characteristics are well balanced across the instrument, conditional on $\mathbf{T}_i$. As shown in Figure 2, patient average characteristics are remarkably stable across the instrument, conditioning on $\mathbf{T}_i$. The coefficients are far from being statistically significant and are negligible in magnitude relative to the sample mean. For completeness, Supplemental Appendix Figure A.4 reports similar balance tests using each of the various patient characteristics included in $\mathbf{X}_i$ as the dependent variable. Even though these characteristics are strong predictors of patient outcomes ($F$-statistics around 100 for joint significance, even controlling for $\mathbf{T}_i$; see Supplemental Appendix Figure A.5), there is little significant relationship between our instrument and the broad range of patient characteristics, conditional on $\mathbf{T}_i$.

As a second set of empirical evidence, we examine the stability of our IV estimates under different sets of controls for patient covariates. Specifically, we divide observable patient characteristics into eight groups and estimate separate regressions that control for each of the $2^8 = 256$ different combinations of patient covariates. We show in the empirical results below that controlling for any combination of patient covariates results in virtually no change in our IV estimates, despite these patient covariates being strong predictors of patient outcomes given the substantial sensitivity of the OLS estimates to their inclusion. Following the logic of Altonji, Elder, and Taber (2005), this evidence implies limited selection bias due to either observed or unobserved patient characteristics that predict patient outcomes. In sum, conditional on $\mathbf{T}_i$, there appears to be little relationship between NP availability and patient characteristics.

*Exclusion.*—While conditional independence supports a causal interpretation of the reduced-form effect, interpreting the IV estimates as identifying the causal effect of being treated by NPs requires an exclusion restriction. That is, the number of NPs on duty impacts patient outcomes only through the patient's probability of being treated by an NP, not through any other channels. After discussing our main results, we present a series of empirical evidence supporting the validity of the exclusion restriction. We first show that the characteristics of both on-duty physicians and on-duty NPs are well balanced across the number of NPs on duty, conditional on $\mathbf{T}_i$, consistent with quasi-random variation in the instrument as well as our IV estimates being unlikely to be driven by different sets of physicians or NPs available across days. Second, we find no evidence that NP presence influences physician performance. Third, we investigate a series of alternative explanations, finding no evidence indicating a violation of the exclusion restriction.

*Monotonicity.*—In the presence of heterogeneous treatment effects, we need to assume monotonicity to interpret IV estimates as a LATE, that is, the average causal effect among cases induced by the instrument into being treated by NPs. In our setting, monotonicity requires that cases treated by NPs on days with fewer NPs on duty would also be treated by NPs on days with more NPs, and vice versa.

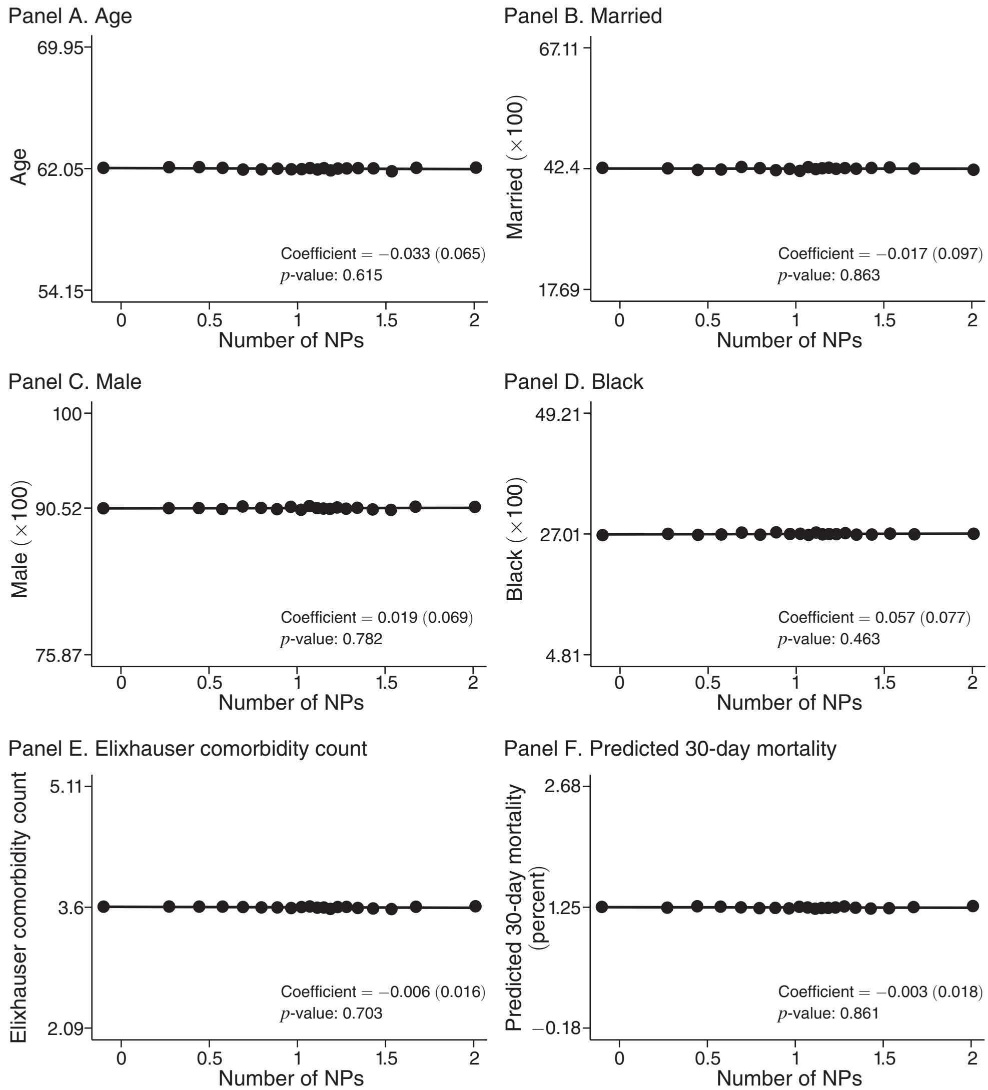


Figure 2. Balance in Patient Characteristics

*Notes*: This figure shows balance in patient characteristics across the number of NPs on duty, conditional on the baseline controls (i.e., ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day indicators). To construct these binned scatterplots, we first residualize both the $y$-axis and $x$-axis variables with respect to the baseline controls and then add means back for ease of interpretation. The middle number on the $y$-axis of each panel reports the mean of the sample; the top and bottom numbers report the mean plus and minus a half standard deviation, respectively (except for panel C, which caps the top number at 100 since the mean plus a half standard deviation is beyond the maximum possible). The coefficients report the estimated slope of the best-fit line between the $y$-axis and $x$-axis variables (conditional on the baseline controls), with standard errors clustered by provider reported in parentheses. Each panel also reports $p$-values for the coefficient estimates. For readability of the coefficients, panels B, C, and D scale up the dependent variable by 100. Predicted 30-day mortality is generated from a linear regression of actual 30-day mortality on patient characteristics $\mathbf{X}_i$ in equations (1) and (2), including demographics, comorbidities, prior health care use, vital signs, and three-digit primary diagnosis indicators.

We examine a testable implication of the monotonicity assumption: The instrument and the probability of being treated by NPs should be positively correlated for any subsample defined by patient characteristics. We test this implication in Supplemental Appendix Figure A.6, where we split the sample by patient characteristics and estimate the first-stage effect separately for each subsample. In particular, we divide the sample by patient age, marital status, gender, race, number of Elixhauser comorbidities, and predicted 30-day mortality.[24] Supplemental Appendix Figure A.6 shows that for all subsamples, the first-stage estimates are positive and statistically different from zero, consistent with the validity of the monotonicity assumption.[25]

## III. Empirical Results

In this section, we present our empirical results. We start by showing the effect of NPs on resource use and downstream patient outcomes. We find that relative to physicians, on average, NPs use more medical resources in the ED visit (longer lengths of stay and higher spending) and exhibit a higher 30-day preventable hospitalization rate. We do not find statistically significant NP effects on 30-day mortality in the overall sample. Next, we examine mechanisms and responses related to the performance difference between NPs and physicians. First, we show that experience may play a role in the NP-physician gap in some outcomes. Second, we find NP responses to lower skill in clinical decision-making—in calling on external resources and setting prescription thresholds. Third, we find that the NP effects on lengths of stay and spending during the ED visit are smaller for less complex and less severe cases. Fourth, we show that patient allocation between NPs and physicians responds toward optimality in the sense of skill-task matching in organizations. Finally, we characterize compliers, present evidence supporting the exclusion restriction, and show a series of additional robustness checks.

### A. *Resource Use*

As summary measures of resource use, we start by examining the effect of NPs on patient length of stay and cost of care during the ED visit. Panels A–B of Figure 3 show the reduced-form estimates of patient log length of stay and log cost of the ED visit versus the instrument (i.e., the number of NPs on duty), conditional on the baseline controls $\mathbf{T}_i$. Log length of stay and log cost increase significantly with the instrument. As a comparison, we also plot in the panels patients' predicted log length of stay and predicted log cost, both of which are well balanced across the instrument.[26]

Columns 1 and 2 of Table 2 report the OLS and IV estimates of the effect of NPs on patient log length of stay and log cost of the ED visit, along with the reduced-form coefficients. All regressions control for the full set of controls described in

[24] Predicted 30-day mortality is generated from a linear regression of actual 30-day mortality on the full set of patient characteristics $\mathbf{X}_i$ in equations (1) and (2).

[25] An interesting pattern in Supplemental Appendix Figure A.6 is that the first-stage estimates appear to be larger for healthier patients: younger patients, those with fewer Elixhauser comorbidities, and those with lower predicted 30-day mortality. This pattern reflects that compliers are more heavily concentrated among healthy patients. In Section IIID, we compute the characteristics of compliers and never-takers. We find consistent evidence that compliers are healthier than the overall sample, while never-takers are riskier.

[26] We form these predictions using linear regressions of actual outcomes on the full set of patient covariates $\mathbf{X}_i$.

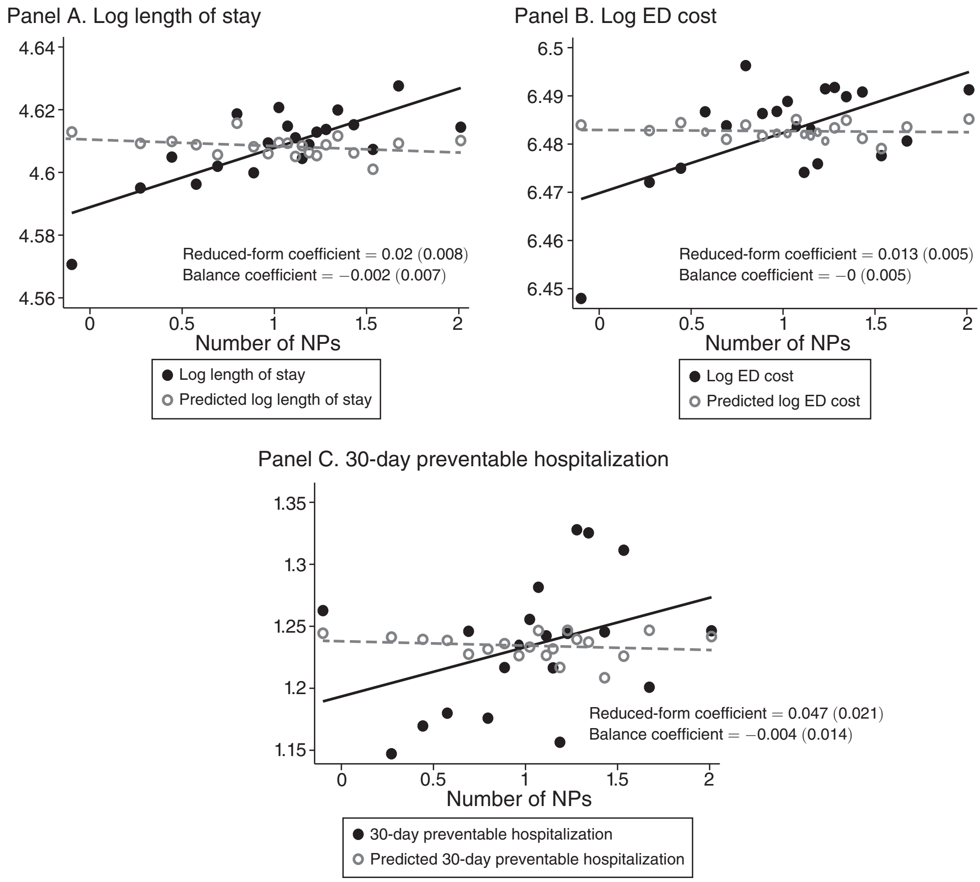


FIGURE 3. REDUCED FORM AND BALANCE

*Notes*: This figure shows binned scatterplots 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). Panel A reports results for log length of stay; panel B reports results for log cost of the ED visit; panel C reports results for 30-day preventable hospitalization. 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.

Section IIA. The OLS estimates (panel A) show that cases treated by NPs have significantly shorter lengths of stay and lower costs, which could reflect NPs treating healthier cases than physicians, at least in terms of observable characteristics in Table 1. Exploiting quasi-random variation in patients' probability of being treated by NPs, the IV estimates (panel C) show that NPs raise patients' medical resource use in the ED visit: On average, cases quasi-randomly assigned to NPs have lengths of stay that are 11 percent longer and ED costs that are 7 percent higher. Given sample means, the NP effects equal an 18-minute increase in the length of stay and a \$66 increase in the cost per ED visit.

Table 2—Effect of NPs on Resource Use and Patient Outcomes

| | Dependent variable | | | | |
|---|---|---|---|---|---|
| | Log length of stay (1) | Log ED cost (2) | Admission (3) | 30-day mortality (4) | 30-day preventable hospitalization (5) |
| *Panel A. OLS* | | | | | |
| NP assignment | −0.168 | −0.101 | −2.982 | −0.211 | −0.067 |
| | (0.034) | (0.023) | (0.460) | (0.032) | (0.030) |
| *Panel B. Reduced form* | | | | | |
| Number of NPs | 0.020 | 0.013 | 0.019 | −0.021 | 0.047 |
| | (0.008) | (0.005) | (0.108) | (0.021) | (0.021) |
| *Panel C. IV* | | | | | |
| NP assignment | 0.110 | 0.070 | 0.103 | −0.116 | 0.252 |
| | (0.045) | (0.030) | (0.585) | (0.115) | (0.120) |
| Full controls | Yes | Yes | Yes | Yes | Yes |
| Mean dep. var. | 4.608 | 6.483 | 16.625 | 1.247 | 1.234 |
| SD 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 estimates of the NP effect on patient log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, and 30-day preventable hospitalization in columns 1–5, respectively. Panel A reports the OLS estimates; panel B reports the reduced-form estimates; panel C reports the IV estimates. Sample sizes in columns 1 and 2 are smaller than that reported in column 1 of Table 1 due to missing outcomes for a small number of cases. The set of full controls includes ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day indicators and patient characteristics that include five-year age-bin indicators, marital status, gender, race indicators (white, Black, and Asian/Pacific Islanders, with other racial categories omitted as the reference group), 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 patient primary diagnosis of the visit. Standard errors clustered by provider are shown in parentheses.

Figure 4 examines the robustness of our OLS and IV estimates to the inclusion of different combinations of patient controls. Specifically, we divide patient covariates into eight subsets: (i) five-year age-bin indicators, (ii) marital status, (iii) gender, (iv) race indicators, (v) dummies for 31 Elixhauser comorbidities, (vi) vital signs, (vii) prior health care use, and (viii) indicators for the three-digit primary diagnosis of the visit. We then estimate separate regressions that control for each of the $2^8 = 256$ different combinations of patient covariates for each outcome for both OLS and IV estimations. Figure 4 shows the range of the coefficients across specifications with different patient controls. Each $n$ on the $x$-axis reports 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.

Figure 4 shows a stark divergence between the OLS and IV estimates. The OLS estimates are negative and decline sizably in magnitude with the addition of patient controls. For example, in the OLS results, conditioning only on baseline controls $\mathbf{T}_i$, we find that cases treated by NPs have 30 percent lower costs than cases treated by physicians. When we add the full set of patient controls, the difference

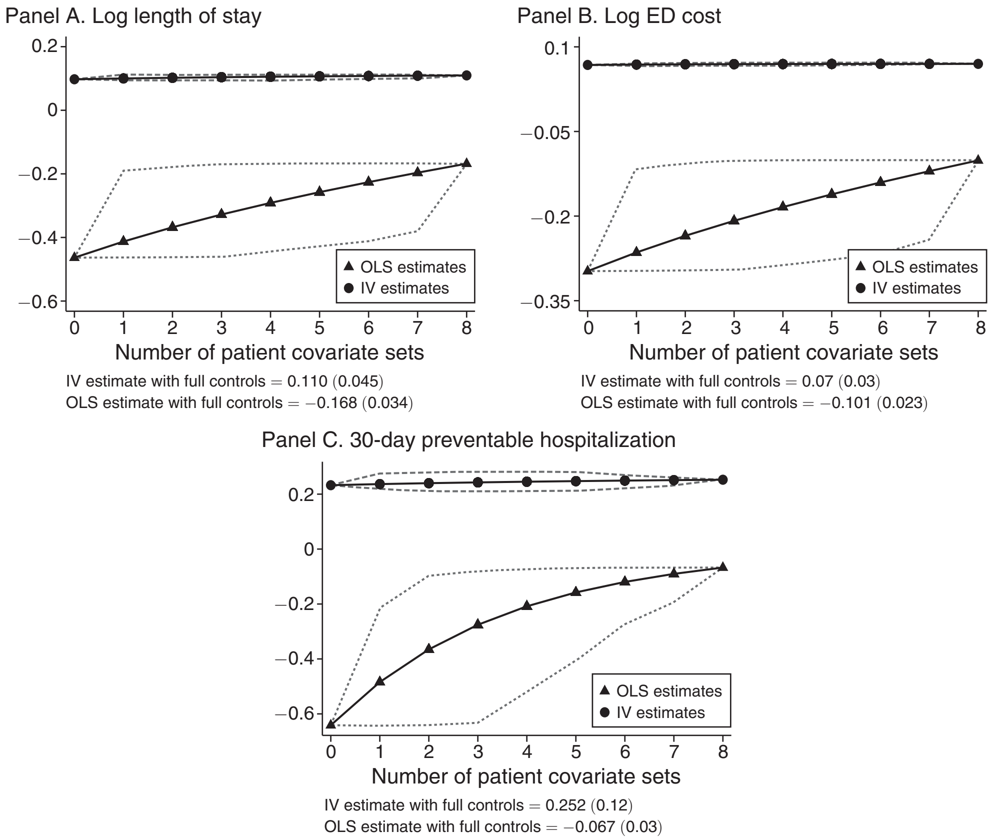


Figure 4. Stability of OLS and IV Estimates

*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 clustered by provider reported in parentheses. Panel A reports results for log length of stay. Panel B reports results for log cost of the ED visit. Panel C reports results for 30-day preventable hospitalization.

attenuates to 10 percent, a decline of two-thirds. The lower health risks of cases treated by NPs (Table 1) and the sensitivity of the OLS estimates to patient controls suggest the presence of selection bias due to unobservable patient characteristics. In contrast, the IV estimates are remarkably robust to controlling for any combination of patient covariates: Despite any controls, the IV estimates for the NP effects on length of stay and cost remain stable at 11 and 7 percent, respectively. Following the logic of Altonji, Elder, and Taber (2005), the stability of the IV estimates implies limited scope for selection on either observable or unobservable patient characteristics that predict potential outcomes, further supporting the validity of our instrument.[27]

In column 3 of Table 2, we assess NP effects on admissions. In the overall sample, we do not find significant NP effects on admissions, while interestingly, Section IIIC shows that NPs significantly increase admissions for severe cases. Supplemental Appendix Figures A.7 and A.8 show actual and predicted outcomes against the instrument for admissions and the robustness of our estimates for admissions to the inclusion of $2^8 = 256$ different sets of patient controls, respectively.

## B. *Patient Outcomes*

We next assess NP effects on downstream patient outcomes, in columns 4 and 5 of Table 2. The IV estimate suggests a higher 30-day preventable hospitalization rate under NPs. Compared to physicians, NPs, on average, raise the 30-day preventable hospitalization rate by 0.25 percentage points, equivalent to a 20 percent increase relative to the mean. Preventable hospitalization is frequently used as a quality-of-care outcome by both researchers and policymakers (e.g., New Jersey Department of Health 2011; Nyweide et al. 2013; California Health and Human Services Agency 2025; Currie and Zhang 2025). Differences in health risk detection and thus timely preventive treatment and differences in treatment quality conditional on risk detection during the ED visit can lead to differences in postdischarge preventable hospitalization rates. The IV estimate for 30-day mortality is less precise. The 95 percent confidence interval for the estimate shown in Table 2 is $-0.34$ to 0.11 percentage points.

Panel C of Figure 3 and panel B of Supplemental Appendix Figure A.7 plot actual and predicted outcomes against the instrument for patient outcomes in this subsection; panel C of Figure 4 and panel B of Supplemental Appendix Figure A.8 show that the IV estimates for patient outcomes are remarkably robust to the inclusion of $2^8 = 256$ different sets of patient controls, whereas the OLS estimates are highly sensitive to including patient controls.

The results suggest that NPs and physicians, on average, operate on different production functions: NPs, on average, use more inputs in the ED visit (longer lengths of stay and higher costs) but exhibit a higher 30-day preventable hospitalization rate. The higher 30-day preventable hospitalization rate also suggests that the greater resource use among NPs is unlikely to be simply because NPs have greater bandwidth to spend time with patients, which we would expect to lead to better patient outcomes under NPs. These results indicate that, when comparing NPs and physicians as two professional classes, physicians, on average, exhibit higher productivity in the ED. However, we note that the estimated average productivity difference between NPs and physicians may be specific to the ED setting, which features,

[27] Another possible explanation for the divergence between the OLS and IV estimates is heterogeneity in treatment effects since OLS reports the average effect among the analysis sample, while IV reports the average effect among compliers (i.e., cases on the margin of being treated by NPs). To explore this possibility, we follow the procedure by Bhuller, Dahl, Løken, and Mogstad (2020) and reweight the analysis sample to match the sample of compliers using predicted 30-day mortality, that is, a composite index of all patient observables. The OLS estimates with the reweighting still differ in sign from the IV estimates, suggesting that the difference between the OLS and IV estimates cannot be accounted for by heterogeneity in NP effects, at least not by heterogeneous effects across observables.

on average, more complex cases than settings such as primary care. Importantly, despite the relative complexity of cases in the ED and the stark differences in training between NPs and physicians, we find substantial overlap between the productivity distributions of NPs and physicians, as shown below.

### C. *Mechanisms and Responses*

This section examines mechanisms and responses related to the average performance difference between NPs and physicians. First, we show that experience may play a role in the NP-physician gap in some (though not all) outcomes. Second, we find NP-physician differences in clinical decision-making (calling on external resources and setting prescription thresholds). Third, we show that the NP effect on resource use is smaller for less complex and less severe cases. Finally, we find that patient assignment moves toward optimality in the sense of skill-task matching.

*Provider Experience.*—First, we ask whether experience affects the magnitude of the performance difference between NPs and physicians. The professions of NPs and physicians entail stark differences in the training that new members undergo and in the selectivity of choosing new members. Whether physicians are more productive than NPs due to training or innate ability has important policy implications. While it is difficult to disentangle these two mechanisms, the extent to which the NP-physician performance gap varies with experience may shed suggestive light on this question. If NPs could be made more productive with more extensive training, the gap may narrow with experience. On the other hand, if the gap derives from innate ability, the gap may persist or even widen with experience.

We form measures of both general and specific experience. We measure general experience as the number of cases the provider has treated from the start of the study period to the day before the current case's visit. We measure specific experience as the number of cases with a condition the same as the current case (measured by the three-digit primary diagnosis) the provider has treated from the start of the study period to the day before the current case's visit. For ease of interpretation, we standardize both general and specific experience to have a mean of zero and a standard deviation of one for NPs and physicians separately.

Our empirical model takes the following form:

$$(3) \quad y_i = \delta_1 NP_i \times Experience_i + \delta_2 NP_i + \delta_3 Experience_i + \mathbf{T}_i \eta + \mathbf{X}_i \beta + \varepsilon_i .$$

$Experience_i$ denotes standardized experience within each provider type.[28] We instrument for $NP_i$ and $NP_i \times Experience_i$ using $Z_i$ (i.e., the number of NPs on duty) and $Z_i \times Experience_i$.

We find that, for some outcomes, experience predicts a smaller NP-physician performance gap. Panel A of Table 3 examines the role of specific experience. For length of stay, column 1 indicates that a one standard deviation increase in specific

[28] A one standard deviation increase in specific experience equals 108 and 107 cases for NPs and physicians, respectively. A one standard deviation increase in general experience equals 1,855 and 1,258 cases for NPs and physicians, respectively.

Table 3—Heterogeneous Effects by Provider Experience

| | Dependent variable | | | | |
|---|---|---|---|---|---|
| | Log length of stay (1) | Log ED cost (2) | Admission (3) | 30-day mortality (4) | 30-day preventable hospitalization (5) |
| *Panel A. Provider specific experience* | | | | | |
| NP assignment | 0.101 | 0.072 | 0.092 | −0.115 | 0.255 |
| | (0.044) | (0.030) | (0.579) | (0.116) | (0.121) |
| NP assignment × experience | −0.058 | −0.042 | −0.504 | −0.001 | 0.016 |
| | (0.025) | (0.019) | (0.331) | (0.041) | (0.030) |
| Experience | −0.001 | 0.006 | 0.238 | −0.001 | −0.016 |
| | (0.006) | (0.010) | (0.314) | (0.018) | (0.014) |
| *Panel B. Provider general experience* | | | | | |
| NP assignment | 0.086 | 0.062 | 0.089 | −0.100 | 0.255 |
| | (0.043) | (0.029) | (0.608) | (0.116) | (0.121) |
| NP assignment × experience | −0.103 | −0.035 | 0.340 | 0.088 | 0.043 |
| | (0.056) | (0.033) | (1.048) | (0.093) | (0.068) |
| Experience | −0.036 | −0.013 | −0.719 | −0.012 | −0.048 |
| | (0.015) | (0.011) | (0.221) | (0.023) | (0.025) |
| Full controls | Yes | Yes | Yes | Yes | Yes |
| Mean dep. var. | 4.608 | 6.483 | 16.625 | 1.247 | 1.234 |
| SD 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 shows heterogeneous effects of NPs 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 from the start of the study period to the day before the current case's visit. Panel B shows heterogeneous effects of NPs by provider general experience, measured as the number of cases (regardless of their diagnoses) the provider has treated from the start of the study period to the day before the current case's visit. For ease of interpretation, both specific and general measures of 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.

experience among NPs and physicians is associated with a 5.8 percent decline in the performance gap, reducing the gap at the mean experience levels by about half. The NP-physician cost gap similarly declines with specific experience. Panel B shows that increasing general experience by one standard deviation in both NPs and physicians is associated with a 10 percent decline in the NP-physician gap in length of stay. Unlike length of stay and cost, the NP-physician gap in the 30-day preventable hospitalization rate remains stable across levels of general or specific experience, suggesting that experience may not fully eliminate the NP-physician performance gap.

We examine alternative measures of experience to address measurement concerns. One concern is that, since we may not observe all the cases providers saw from the start of their careers, our measures of experience are imperfect proxies for it. To mitigate this concern, we restrict the sample to cases visiting after 2017, so that our experience measures have at least a one-year look-back window. We also measure experience based on cases seen in the prior year (i.e., 365 days before the

TABLE 4—CLINICAL DECISION-MAKING

| | Dependent variable | | | | |
|---|---|---|---|---|---|
| | Consult (1) | CT (2) | X-ray (3) | Opioid (4) | Antibiotic (5) |
| NP assignment | 0.026 | 0.012 | 0.020 | −0.018 | 0.040 |
| | (0.009) | (0.007) | (0.009) | (0.006) | (0.022) |
| Full controls | Yes | Yes | Yes | Yes | Yes |
| Mean dep. var. | 0.226 | 0.145 | 0.291 | 0.088 | 0.639 |
| SD dep. var. | 0.418 | 0.352 | 0.454 | 0.283 | 0.480 |
| Observations | 1,118,836 | 1,118,836 | 1,118,836 | 1,118,836 | 123,395 |

*Notes*: This table shows IV estimates of the effect of NPs on various measures of clinical decision-making. The outcomes in columns 1–5 are whether the patient receives during the ED visit formal consults, CT scans, X-rays, opioid prescriptions, and antibiotic prescriptions, respectively. Since antibiotics generally only apply to patients with infections, column 5 restricts the sample to patients with respiratory or genitourinary system infections, two common types of infections. All estimations include the full set of controls described in the notes to Table 2. Standard errors clustered by provider are shown in parentheses.

index day), so that the estimates precisely represent heterogeneity by prior-year experience.[29] As the effect of experience may decay with time (e.g., Benkard 2000), recent experience could be more important than experience gained in the relatively distant past. A second concern is that the number of cases a provider has seen may capture speed, which may affect productivity independent of experience (e.g., faster providers accumulate more cases and discharge patients earlier). To examine this concern, we include an alternative (general) experience measure—the number of days a provider has worked since the start of our study period to the day before the current case's visit—which is independent of speed. Supplemental Appendix Tables A.4 to A.6 show that our estimates are qualitatively similar under these alternative measurements.[30] Overall, the results suggest that the NP-physician gap in some (though not all) outcomes may narrow with experience, instead of being persistent.

*Clinical Decision-Making.*—Next, we examine clinical decision-making that may respond to the differences in human capital, specifically, diagnostic skill. Providers with lower diagnostic skill may draw on more external resources, such as consults and diagnostic tests. They may also adjust treatment thresholds for decisions with asymmetric costs between type I and type II errors (Chan, Gentzkow, and Yu 2022). These could be optimal responses to skill differences; they could also increase the cost of care, manifesting in productivity differences.

Columns 1–3 of Table 4 report the effect of NPs on the use of informational resources, using the 2SLS estimation in equations (1) and (2). Column 1 shows that,

[29] Specifically, we measure general experience as the number of cases the provider treated in the preceding 365 days; we measure specific experience as the number of cases with the same three-digit primary diagnosis as the current case the provider treated in the preceding 365 days. We exclude from this estimation cases visiting in the first year of our analysis since we cannot fully observe their providers' experience in the preceding 365 days.

[30] We do not include age as a measure of general experience in this heterogeneity analysis because NPs often practice as nurses for varying lengths of time before becoming NPs, and thus, unlike for physicians, age is a noisy measure of experience for NPs.

relative to physicians, NPs are more likely to use consults: NPs increase consults by 2.6 percentage points, or 11 percent of the sample mean. Columns 2 and 3 show that, relative to physicians, NPs are more likely to order CT scans and X-rays, the two primary diagnostics in the ED setting. NPs increase CT scan and X-ray ordering by 1.2 and 2.0 percentage points, respectively, or 8.3 and 6.9 percent of the respective sample means.

These results suggest that NPs are more likely to collect resource-intensive information from external sources than physicians. This could directly increase lengths of stay and medical costs since consults and diagnostics take time and resources. On the other hand, consults and diagnostics allow lower-skilled providers to improve decision-making by incorporating information from other experts and, thus, may not be suboptimal. These findings may also imply that NPs seek inputs from other experts when needed, even when allowed to practice independently, which may be reassuring to policy discussions about granting NPs full practice authority.

Next, we evaluate skill from the lens of thresholds for prescriptions with asymmetric costs. Specifically, as shown by Chan, Gentzkow, and Yu (2022), provider skill may correlate with treatment thresholds when the costs of false-positive (type I) and false-negative (type II) errors are asymmetric. Compared to higher-skilled providers, lower-skilled providers may (*optimally*) adjust their treatment thresholds with less information. Specifically, providers with less information may more frequently opt for treatment when false negatives (not treating when a case should have been treated) are costlier than false positives (treating when a case should not have been treated); conversely, lower-skilled providers may less frequently opt for treatment when false positives are costlier than false negatives. Despite more frequent prescriptions, lower-skilled providers may nonetheless incur more false negatives; similarly, they may still incur more false positives with fewer prescriptions.

We choose two important prescriptions with different asymmetries in the costs of type I and type II errors—opioids and antibiotics—and examine prescription thresholds and downstream outcomes. For both of these prescriptions, the clinical indications for appropriate use are often unclear and require clinical judgment.[31] We estimate the NP-physician difference in prescriptions and downstream outcomes using the 2SLS model in equations (1) and (2). Column 4 of Table 4 shows that, for opioids, which have higher false-positive costs—e.g., addiction and overdose among patients who should not have received opioids compared to continued pain among patients who should have received them—NPs lower opioid prescriptions by 1.8 percentage points, or 20 percent of the mean. For antibiotics, which have higher false-negative costs—e.g., nontreatment of a potentially life-threatening infection compared to antibiotic resistance—NPs raise antibiotic prescriptions by 4.0 percentage points, 6.3 percent of the mean.[32] Yet, as Supplemental Appendix Table A.7 shows, compared to those treated by physicians, patients treated by NPs are likelier to incur return visits with infections despite higher antibiotic prescribing, and they obtain similar rates of subsequent opioid use disorder despite lower opioid prescribing.

[31] See, for example, Fleming-Dutra et al. (2016); Huang et al. (2018); Butler et al. (2019); and Neuman, Bateman, and Wunsch (2019).

[32] Since opioids apply to a wide range of conditions, we include all patients in examining opioid prescriptions. As antibiotics generally only apply to patients with infections, we restrict the sample to patients with respiratory or genitourinary system infections, two common types of infections.

*Case Complexity and Severity.*—In this analysis, we exploit the wide variety of cases that arrive at the ED to examine heterogeneity in the effects of NPs by case complexity and severity. Following Imbens and Rubin (1997), we estimate complier potential outcomes under NPs and under physicians. Specifically, we estimate complier potential outcomes under NPs using the following IV regression:

$$(4) \qquad y_i \cdot NP_i = \sum_{g=1}^{G} \mathbf{1}\left(Group_i = g\right)\left[\delta_g NP_i + \lambda_g\right] + \mathbf{T}_i \eta + \mathbf{X}_i \beta + \varepsilon_i,$$

where $y_i \cdot NP_i$ is the interaction between patient outcome and the indicator for being treated by an NP, and $\mathbf{1}(Group_i = g)$ is an indicator for case $i$ belonging to group $g \in \{1, \ldots, G\}$ characterizing complexity or severity. As a natural extension of our main IV model, we instrument for the interactions between $\left\{\mathbf{1}(Group_i = g)\right\}_{g=1}^{G}$ and $NP_i$ by interacting $\left\{\mathbf{1}(Group_i = g)\right\}_{g=1}^{G}$ with $Z_i$, where $Z_i$ is the number of NPs on duty. We estimate complier potential outcomes under physicians using an IV regression similar to equation (4) but with a dependent variable of $y_i \cdot (NP_i - 1)$.

We consider two partitions of cases. First, we divide cases into quartiles by their number of Elixhauser comorbidities and refer to higher quartiles as more complex cases. Second, we divide cases into two groups based on whether the condition severity, measured by the 30-day mortality for the three-digit primary diagnosis, is equal to or above the ninety-fifth percentile of the sample. The 30-day mortality for this top-severity group is 8.6 percentage points, compared with 1.2 percentage points for the whole sample. Included in this group are cases with severe conditions, such as acute kidney failure and acute myocardial infarction (AMI), which are potentially more challenging to manage.[33]

We find the NP effects on lengths of stay and ED costs are smaller when cases are less complex and less severe, as shown in Figure 5. For example, for cases in the second complexity quartile, NPs increase their ED costs by 7 percent; for cases in the highest-complexity quartile, NPs show a 12 percent increase in ED costs; and for cases with a condition severity at or above the ninety-fifth percentile, NPs show a 25 percent increase in ED costs. Table 5 summarizes the heterogeneous NP effects.[34] The table further examines four severe conditions: sepsis, stroke, AMI, and heart failure. We find smaller NP effects on length of stay and ED costs among cases without these severe conditions. Supplemental Appendix Table A.9 reports the outcome mean for each group of cases in Table 5.

[33] Supplemental Appendix Table A.8 summarizes the 10 most common three-digit primary diagnoses in this group. The largest diagnosis category is heart failure, followed by acute kidney failure. The top 10 diagnoses also include AMI, a form of sepsis, and a form of respiratory failure. The 30-day mortality rates range between 5 and 17 percentage points. A related question is why severe cases are assigned to NPs. While the data show a reduced probability of NP assignment among riskier patients, the probability remains positive (the share of patients assigned to NPs among patients whose condition severity is below and at least as high as the ninety-fifth percentile of the sample is, respectively, 0.24 and 0.11). Possible reasons that severe cases are assigned to NPs may include (i) triage providers may misevaluate patient severity (Chan and Gruber 2020) and assign severe cases to NPs, and (ii) when physicians are occupied with earlier cases, it may be more efficient to assign severe cases to available NPs than to delay care.

[34] Results in this table are estimated using equation (4) with patient outcome $y_i$ as the dependent variable. By construction, this dependent variable is the difference between the dependent variables used to estimate potential outcomes: $y_i = y_i \cdot NP_i - y_i \cdot (NP_i - 1)$.

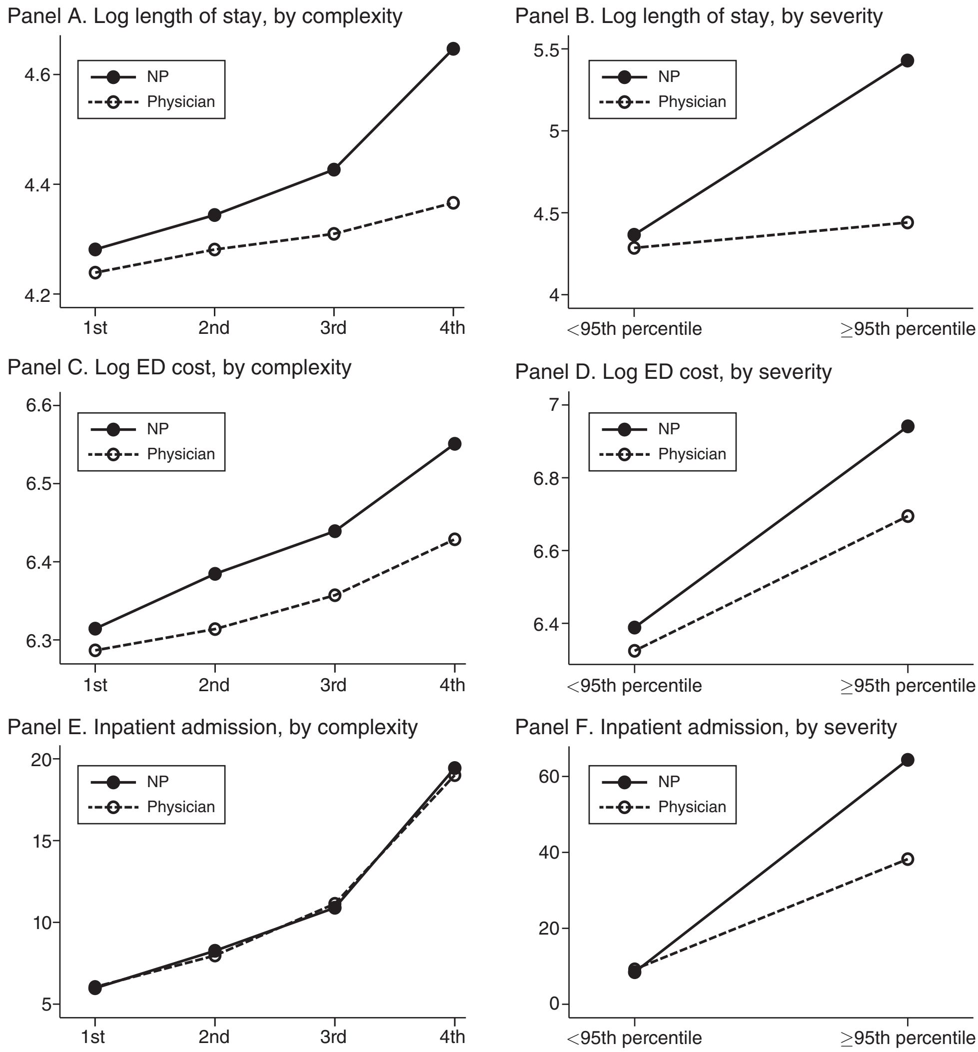


Figure 5. Heterogeneous Effects by Case Complexity and Severity

*Notes*: This figure shows heterogeneous effects of NPs by case complexity and severity. Panels A, C, and E divide cases into quartiles by their total number of Elixhauser comorbidities, with higher quartiles indicating more complex cases. Panels B, D, and F divide cases by whether the condition severity measured by the 30-day mortality of cases with the same three-digit primary diagnosis is equal to or above the ninety-fifth percentile of the sample. The solid and dashed lines show complier potential outcomes if they were treated by NPs and physicians, respectively. We estimate complier potential outcomes under NPs by the IV regression in equation (4). We estimate complier potential outcomes under physicians by an IV regression similar to equation (4) but with a dependent variable of $y_i \times (NP_i - 1)$, that is, the interaction between patient outcome and the indicator for being treated by an NP minus one. Panels A–B, C–D, and E–F report results for log length of stay, log ED cost, and inpatient admission in the ED visit, respectively. The IV estimates of the effects of NPs for each patient subgroup are reported in Table 5.

While NPs do not affect admissions for low-risk patients, they raise admissions among high-risk patients. For cases with severity at or above the ninety-fifth percentile, NPs show an increased admission rate of 26 percentage points, or 42 percent compared to the mean. For sepsis, stroke, AMI, and heart failure cases, NPs increase their admission rates by 45, 73, 123, and 20 percentage points, respectively,

Table 5—Heterogeneous Effects by Patient Characteristics

| | Dependent variable | | | | |
|---|---|---|---|---|---|
| | Log length of stay (1) | Log ED cost (2) | Admission (3) | 30-day mortality (4) | 30-day preventable hospitalization (5) |
| *Panel A. Elixhauser comorbidity count* | | | | | |
| First quartile | 0.042 | 0.028 | −0.077 | −0.147 | 0.555 |
| | (0.045) | (0.032) | (0.636) | (0.120) | (0.119) |
| Second quartile | 0.063 | 0.071 | 0.291 | −0.041 | 0.438 |
| | (0.044) | (0.030) | (0.642) | (0.126) | (0.132) |
| Third quartile | 0.117 | 0.082 | −0.245 | −0.099 | 0.250 |
| | (0.048) | (0.031) | (0.761) | (0.178) | (0.186) |
| Fourth 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) |
| $\geq$ 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* | | | | | |
| Sepsis | 1.480 | 0.095 | 44.880 | 24.533 | 11.117 |
| | (0.609) | (0.329) | (24.114) | (15.169) | (7.961) |
| 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) |
| 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 |
| SD 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. 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 primary diagnosis is equal to or above the ninety-fifth 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. Supplemental Appendix Table A.9 reports outcome means for each group of cases in this table.

or 50, 130, 214, and 31 percent relative to the mean. From the perspective of allocative efficiency, the greater use of admissions and other resources by NPs relative to physicians for more severe cases may not be suboptimal: It may improve NP performance on patient health outcomes relative to a counterfactual scenario that uniformly constrains NPs to the same admission rates as those of physicians.

We find interesting evidence of this trade-off between resource use and patient health outcomes when we examine heterogeneity in NP effects on 30-day preventable hospitalizations. In particular, we find decreasing NP effects on 30-day preventable hospitalization with case complexity or severity (column 5 of Table 5). If NPs are less skilled at treating more complex or more severe cases, they may worsen outcomes more for these cases. On the other hand, as NPs increase their intensity of care by a larger extent for these cases, the higher incremental care may in fact reduce preventable hospitalizations. As shown in panel A of Table 5, for cases in the highest-complexity quartile, NPs increase lengths of stay and ED costs by 28 percent and 12 percent, respectively, without a significant effect on the 30-day preventable hospitalization rate. For cases in the lower-complexity quartiles, NPs reveal a smaller-magnitude, and sometimes insignificant, effect on length of stay and medical costs, yet NPs significantly raise the 30-day preventable hospitalization rate for these cases. For the most severe cases (i.e., those with severity at or above the ninety-fifth percentile), NPs *reduce* 30-day preventable hospitalization by 3 percentage points, while increasing length of stay by 99 percent, ED costs by 25 percent, and admissions by 26 percentage points (panel B of Table 5).

*Patient Assignment.*—Finally, we examine patient assignment between NPs and physicians. The evidence on case heterogeneity above and, more directly, the counterfactual cost estimates by case complexity in Section IV (below) suggest a comparative disadvantage for NPs in treating complex and severe cases. We do not find a set of cases in which NPs outperform physicians, suggesting that NPs are also at an absolute disadvantage in the ED setting. These stylized facts imply a qualitatively optimal assignment of patients to NP versus physician provider classes, in the sense of skill-task matching in organizations (Acemoglu and Autor 2011): Of patients available, NPs should receive the healthier ones, and NPs should receive fewer patients when physicians are more available. Some recent studies suggest that optimal worker-job matching may not hold in practice (e.g., Adhvaryu, Kala, and Nyshadham 2022). Yet empirical patient allocation between NPs and physicians consistent with the direction of optimal assignments can augment the efficiency of using NPs.

Supplemental Appendix Figure A.9 provides descriptive insights into patient assignment by exploiting variation in NP staffing and patient arrivals, conditional on our baseline controls. Panels A–C show that NPs overall are assigned healthier cases, consistent with Table 1. The average complexity and severity of cases assigned to NPs, as measured by patient age, comorbidities, and predicted 30-day mortality, increase with NP staffing, while the average complexity and severity of all cases remain stable. The pattern suggests an assignment process in which the first cases assigned to NPs have the lowest health risks; when more NPs (and fewer physicians) are available, cases incrementally assigned to NPs are riskier compared to those initially assigned to NPs but are still relatively healthy among the remaining cases to be assigned.[35] Panel D of the figure assesses the probability that a patient is assigned to an NP as a function of ED busyness measured by the number of other

[35] As a result, cases left for physicians become riskier with more NPs. A potential question is whether the increased risk of cases assigned to physicians on days with more NPs affects physicians' performance, violating an exclusion restriction. To address this concern, Supplemental Appendix Table A.10 controls for multiple measures of average health risk (age, comorbidities, and predicted 30-day mortality) for cases assigned to physicians within the ED-day cell. The results show that our IV estimates are highly robust.

patients arriving in the analysis time window (i.e., 8 AM to 6 PM) of the ED-day cell. We find a small but clear trend in which patients are more likely to be assigned to NPs when the ED is busier, consistent with the efficiency of assigning fewer patients to NPs when physicians are less occupied.[36] Taken together, the empirical patient allocation between NPs and physicians is consistent with the direction of optimal skill-task matching in organizations.

### D. *Complier Characteristics*

Our IV estimates represent the LATE, that is, the average causal effect among complier cases quasi-randomly assigned to NPs versus physicians due to the instrument. To better understand this LATE, we examine complier characteristics following the approach developed by Abadie (2003), as described in Supplemental Appendix A.2. Supplemental Appendix Table A.11 reports the results. Consistent with NPs treating less severe cases, compliers are healthier than the average case. Compared with the average case, compliers are younger and have fewer Elixhauser comorbidities, fewer inpatient stays in the prior year, and lower predicted mortality. Supplemental Appendix Table A.11 also examines characteristics of never-takers (of NPs), following an approach from Dahl, Kostøl, and Mogstad (2014) that we detail in Supplemental Appendix A.2. In line with the notion that NPs treat healthier cases than physicians, never-takers are riskier than the average case, and both are riskier than compliers. (Our setting contains no always-takers since patients cannot be assigned to NPs on days with no NPs.)

### E. *Exclusion Restriction*

As discussed in Section IIB, interpreting the IV estimates as the causal effect of being treated by NPs requires the number of NPs on duty to affect patient outcomes only through patients' probability of being treated by NPs, not through any other channels. In this section, we present evidence supporting the exclusion restriction. We first show that the characteristics of both on-duty physicians and on-duty NPs are well balanced across the number of NPs on duty. Second, we assess and find little support for the possibility of productivity spillovers whereby NP presence influences physician performance in our setting. Third, we examine a range of factors that may vary across days, including patient volume, total staffing, and patient wait times, and consider patient–physician gender concordance, finding no evidence to suggest these factors are driving our IV estimates.

*Balance in Provider Characteristics.*—To start, we investigate whether on-duty physicians are similar across days with different numbers of NPs. If such a balance does not hold, our IV estimates could be driven by compositional changes

[36] This relationship is qualitatively the same when further conditioning on the number of NPs and the number of physicians staffing the ED on that day. A related question is whether the increasing NP assignment with ED busyness may affect the estimated NP effects. In contrast to OLS, a correlation between NP assignment and busyness per se would not affect estimates from our IV approach. In Section IIIE, we also show that our IV estimates are highly robust to controlling for the number of patients arriving.

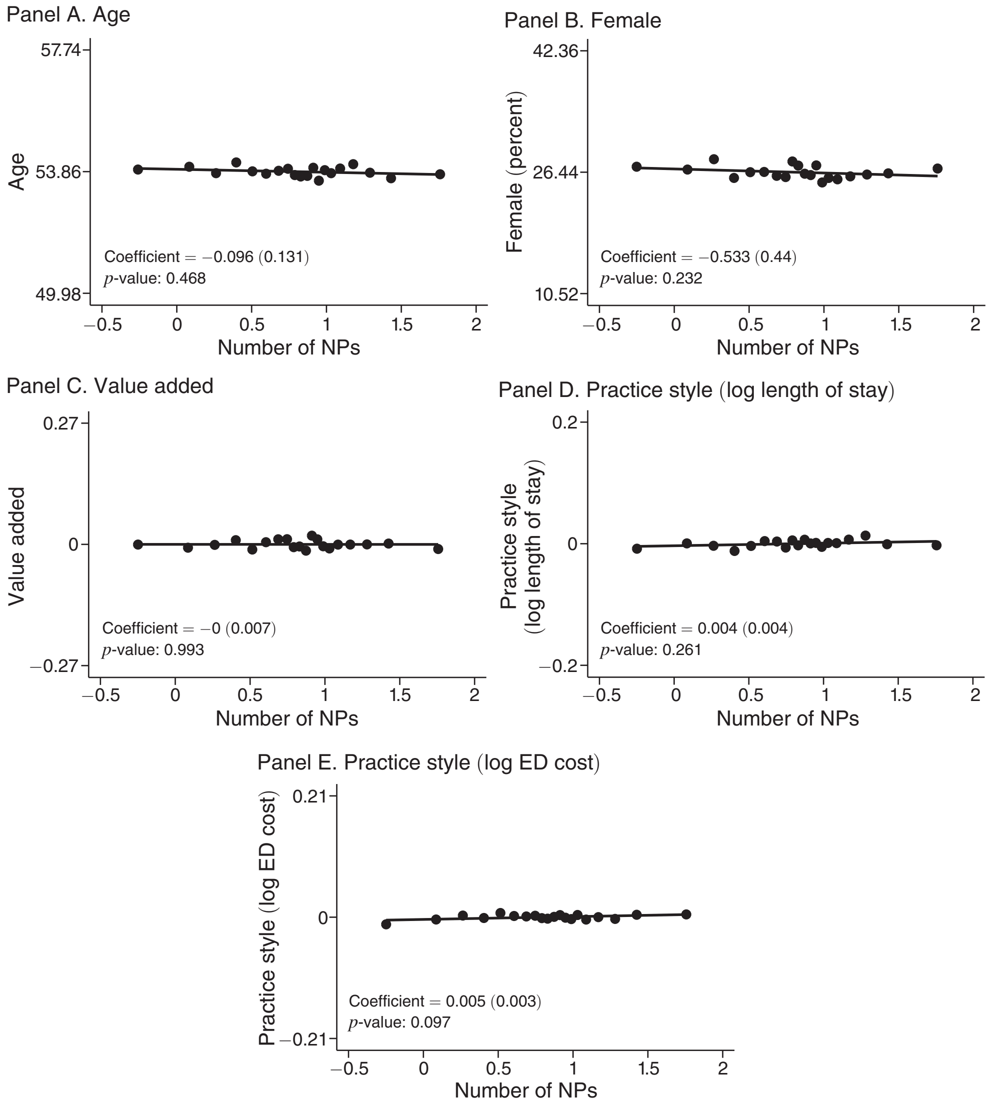


Figure 6. Balance in Physician Characteristics

*Notes*: These panels are graphical representations of the balance-test regression at the ED-day level of physician average characteristics (weighted by the number of cases treated by each physician) on the number of NPs on duty, conditional on ED-by-year, ED-by-month, and ED-by-day-of-the-week indicators. Coefficients from the regressions are reported in each panel, along with standard errors (shown in parentheses) and $p$-values. To construct the binned scatterplots, we first residualize both the $y$-axis variable (average characteristics of physicians on duty) and the $x$-axis variable (the number of NPs on duty) with respect to ED-by-year, ED-by-month, and ED-by-day-of-the-week indicators and then add means back to aid in interpretation. The middle number on the $y$-axis of each panel reports the mean of the sample; the top and bottom numbers report the mean plus and minus a half standard deviation, respectively. The physician characteristics reported in panels A–E are, respectively, age, gender, value added, practice style in terms of patient log length of stay, and practice style in terms of patient log cost of the ED visit. Construction details of value added and practice style are described in Supplemental Appendix A.3.

of physicians. Figure 6 reports the balance for various physician characteristics. Specifically, we consider an analysis at the ED-day level that asks whether on-duty physicians' average characteristics (weighted by the number of cases treated by each physician) are independent of the number of NPs on duty, conditional on ED-by-year, ED-by-month, and ED-by-day-of-the-week indicators.[37]

We examine three sets of physician characteristics: (i) demographics of age and gender; (ii) a measure of physician "value added," reflecting physician risk-adjusted impact on patient 30-day mortality; and (iii) measures of physician "practice style," reflecting a physician's risk-adjusted average input choices in terms of length of stay and ED costs (see Supplemental Appendix A.3 for construction details). Figure 6 shows that each of these physician characteristics is balanced across the instrument, conditional on the baseline controls.

We similarly examine whether NP characteristics are systematically different across days with differing numbers of NPs on duty. If such systematic variation exists, our baseline IV strategy may not be able to disentangle the effect of being treated by NPs from that due to the potentially different quality of NPs across days.[38] Following Figure 6, Figure 7 shows that an analogous set of NP characteristics is well balanced across days with different numbers of NPs, conditional on baseline controls.[39]

The balance in on-duty physicians' and on-duty NPs' characteristics across the instrument also supports quasi-random variation in the instrument, as with the balance in patient characteristics. Otherwise (e.g., if the variation is driven by patient or staffing shocks), we would observe systematic changes in these characteristics with NP staffing.

*Assessing Productivity Spillovers.*—We then consider the possibility of spillovers between NPs and physicians: whether NP presence may influence physician performance. For example, if NPs ask physicians in the ED for assistance, they may slow down physicians. Alternatively, a change in peers from days without any NPs to days with NPs may influence physician performance as, e.g., physicians may come under different degrees of peer pressure that motivate them to work differently (Chan 2016; Silver 2021).

However, we find little empirical evidence to suggest meaningful spillovers between NPs and physicians. First, if NPs ask physicians for assistance, we may expect the outcomes of patients treated by NPs to depend on the quality of physicians on duty. Using the value added measure described in Supplemental Appendix A.3,

[37] The empirical specification takes the form $\bar{y}_{jd} = \tilde{\lambda} Z_{jd} + \tilde{\mathbf{T}}_{jd}\tilde{\eta} + \varepsilon_{jd}$, where $\bar{y}_{jd}$ is the average characteristics of physicians on duty at ED $j$ on day $d$ (weighted by the number of cases treated by each physician); $Z_{jd}$ is the number of NPs on duty; and $\tilde{\mathbf{T}}_{jd}$ includes ED-by-year, ED-by-month, and ED-by-day-of-the-week indicators. We cluster standard errors by ED. Since our main sample has only 44 EDs, a potential question is whether the estimated standard errors are biased, given the relatively small number of clusters. While there is currently no clear-cut definition of "small," if anything, such a potential issue would bias us toward rejecting the null hypothesis of physician balance across the instrument (Bertrand, Duflo, and Mullainathan 2004; Cameron and Miller 2015). Nonetheless, as a robustness check, we apply the correction for the small number of clusters using the wild cluster bootstrap as suggested by Cameron and Miller (2015); we find no meaningful change in the standard error estimates.

[38] This concern applies to our baseline IV strategy since it uses variation in the number of NPs in addition to whether there is an NP on duty. In Section IIIF, we include alternative estimations using variation in the extensive margin of whether there is any NP on duty and restricting the sample to days with zero or only one NP on duty; results are virtually unchanged.

[39] Though the balance-test coefficient for physician practice style in terms of ED costs has a $p$-value below 0.1 at 0.097, given multiple hypothesis testing, it can reflect a chance finding. To further assess the robustness, Supplemental Appendix Table A.12 reports our IV estimates controlling for on-duty providers' practice styles, showing that the results are stable.

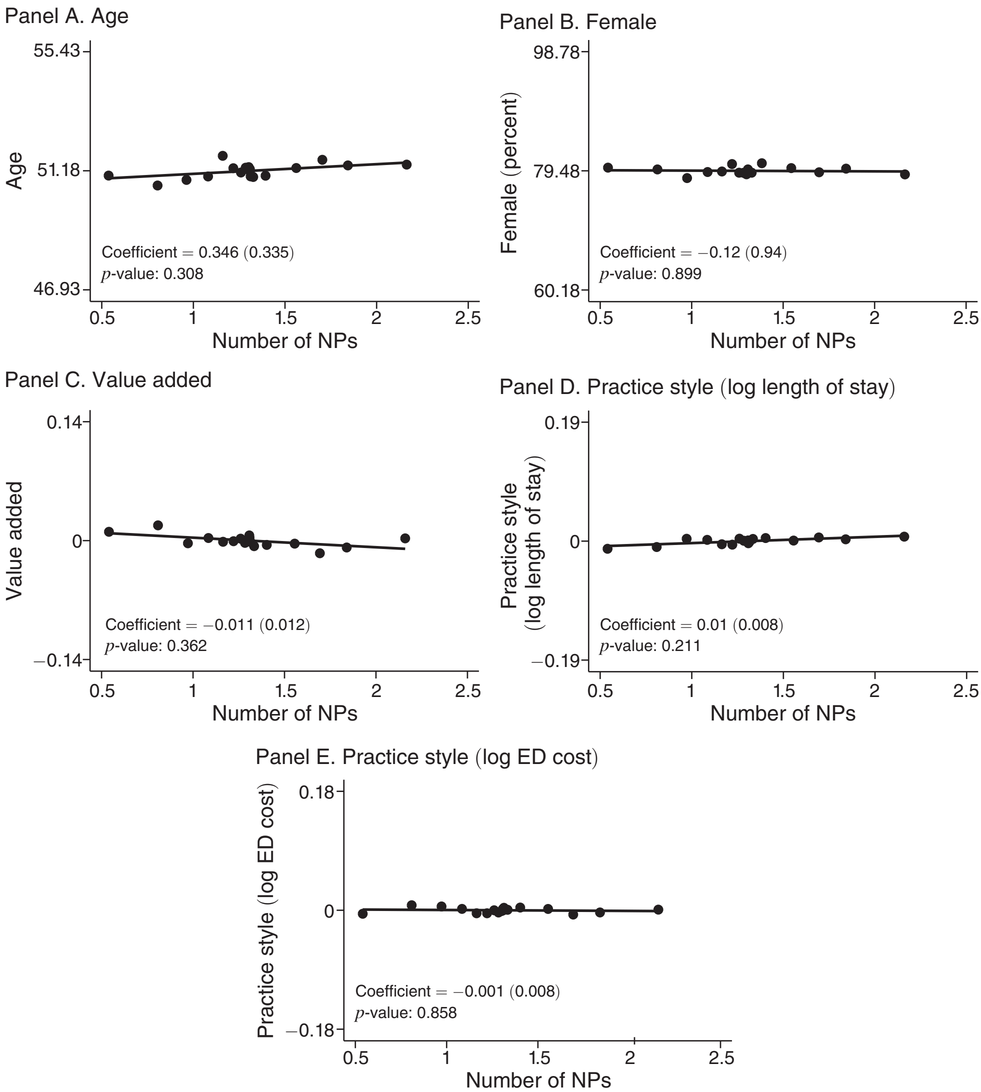


Figure 7. Balance in NP Characteristics

*Notes*: These panels are graphical representations of the balance-test regression at the ED-day level of NP average characteristics (weighted by the number of cases treated by each NP) on the number of NPs on duty, conditional on ED-by-year, ED-by-month, and ED-by-day-of-the-week indicators. Coefficients from the regressions are reported in each panel, along with standard errors (shown in parentheses) and $p$-values. To construct the binned scatterplots, we first residualize both the $y$-axis variable (average characteristics of NPs on duty) and the $x$-axis variable (the number of NPs on duty) with respect to ED-by-year, ED-by-month, and ED-by-day-of-the-week indicators and then add means back to aid in interpretation. The middle number on the $y$-axis of each panel reports the mean of the sample; the top and bottom numbers report the mean plus and minus a half standard deviation, respectively. The NP characteristics reported in panels A–E are, respectively, age, gender, value added, practice style in terms of patient log length of stay, and practice style in terms of patient log cost of the ED visit. Construction details of value added and practice style are described in Supplemental Appendix A.3.

we find no such relationship in Supplemental Appendix Table A.13. Second, with spillovers—either through assistance or peer pressure, or importantly, any other channels by which NP staffing may affect physician performance—outcomes for patients treated by physicians could change with the presence of NPs. Directly regressing outcomes of patients treated by physicians on the NP presence suffers from patient selection since physicians are allocated riskier cases on days with NPs (as healthier cases are assigned to NPs). To circumvent this issue, we examine cases arriving between 5 and 8 AM, that is, patients who arrive before the typical start of NP shifts so that they are unlikely to be assigned to NPs, but whose stay overlaps with NP shifts so that their physicians could be subject to spillover effects from NPs.[40] As shown in panel A of Supplemental Appendix Table A.14, we find no evidence of spillovers from NP presence on physicians' patients overall. In panel B, we examine days with high workloads, when NP spillovers may be more detectable, and again find no evidence of spillovers in this subsample.[41] In sum, there appears to be little evidence that NP presence affects physician performance in our setting.

*Robustness to Additional Factors.*—Finally, we show the robustness of our IV estimates to factors that may vary across days, including total physician equivalents on duty, the total number of cases arriving, and patient wait times (the time between arrival at the ED and assignment to a treating provider). We also show that our IV estimates are unlikely to be driven by patient–provider gender mismatch.

We control for total physician equivalents on duty to mitigate the concern that the effective level of staffing may vary across days with different numbers of NPs on duty. Conditional on $\mathbf{T}_i$, the ED plausibly maintains a consistent level of total effective staffing across days, lending credence to our baseline IV estimates. The coefficient in panel A of Figure 1 being less than one in magnitude does not necessarily indicate increased staffing with more NPs since the number of providers is not a sufficient statistic for staffing. For example, given that NPs take longer to discharge a patient but do not handle more or sicker patients simultaneously, the NP-physician substitution rate is below one, and maintaining balanced total effective staffing across days with differing NP availability would require a coefficient less than one in magnitude in panel A of Figure 1. Importantly, if an increased number of NPs is associated with higher total staffing, we would expect better patient outcomes, yet we find a higher 30-day preventable hospitalization rate on days with more NPs. For further robustness checks, we control for physician equivalents on duty, calculated as the sum of the number of on-duty physicians and the number of on-duty NPs multiplied by the substitution rate between NPs and physicians. Since we do not know how many physician equivalents an average NP represents, we assume a series of substitution rates—0.25, 0.5, 0.75, and 1, assuming that one NP substitutes for 0.25, 0.5, 0.75, and 1 physician, respectively. We cap the substitution rate at one since, as described above, one NP appears insufficient to

[40] To restrict further the possibility of patient selection between NPs and physicians, we exclude patients arriving between 5 and 8 AM in ED-day cells that contain any patients arriving during the window assigned to NPs. As shown in columns 1 and 2 of Supplemental Appendix Table A.14, patient health risks are balanced between days with and without NPs in this analysis sample.

[41] A potential question is whether the increased average health risk of patients assigned to physicians on days with NPs affects physicians' overall performance, violating an exclusion restriction. To assess this concern, as described in footnote 35, Supplemental Appendix Table A.10 controls for multiple measures of the average health risk (age, number of Elixhauser comorbidities, and predicted 30-day mortality) of patients assigned to physicians in the ED-day cell. The results show that our IV estimates are remarkably stable, suggesting that the different sets of patients assigned to physicians on days with NPs are unlikely to affect our IV estimates.

substitute for one physician. When the assumed substitution rate is higher than the actual rate, the IV estimates will be biased by lower staffing, leading to underestimated IV estimates for log length of stay and ED cost when lower staffing rushes providers to shorten patient visits and reduce treatments (Silver (2021) shows that when ED physicians are rushed, they discharge patients early and reduce treatments). Conversely, if the assumed substitution rate is lower than the actual rate, the IV estimates will be confounded by higher staffing, potentially leading to overestimated NP effects on log length of stay and ED cost. The more the assumed substitution rate is above (below) the actual rate, the smaller (larger) the estimates will be relative to the truth. Similarly, if lower (higher) staffing worsens (improves) patient outcomes, e.g., mortality, we would see decreased IV estimates for mortality with lower assumed substitution rates. Since the interval [0.25, 1] plausibly covers the true substitution rate, we consider this robustness test a bounding approach. Supplemental Appendix Figure A.10 reports the results. Though the IV estimates modestly change with the substitution rate as explained above, the estimates are qualitatively similar to our baseline IV estimates.

Turning to the total number of cases arriving and wait times, while potentially important, Supplemental Appendix Table A.15 shows that these two factors do not affect our estimates. The IV estimates are highly robust to controlling for these two factors. Since patient wait times are potentially endogenous (healthier cases could be assigned a lower priority and thus wait longer), we instrument for wait times using the average wait time of cases visiting on the same day at the same ED as the index case in Supplemental Appendix Table A.15.

We also ask if the estimated NP effect is driven by patient–provider gender mismatch since the vast majority of patients are male (91 percent), physicians are primarily male (74 percent), and NPs are mostly female (79 percent). Supplemental Appendix Table A.16 explores this possibility by asking whether the effect of NPs varies by whether the patient's and provider's genders match. The results show little heterogeneity.

### F. *Additional Robustness Checks*

Supplemental Appendix Tables A.17–A.21 report additional robustness checks. Supplemental Appendix Table A.17 shows that our findings are stable with alternative standard error clustering approaches: clustering by ED-day cell or two-way clustering by ED-day cell and provider. Panels B and C of Supplemental Appendix Table A.18 show the robustness of our estimates to, respectively, an alternative count of on-duty NPs that includes any NP with at least one case (instead of at least two cases as in our baseline IV) in the analysis time window of an ED-day cell and another count that leaves out the index case in measuring on-duty NPs. In panels D–F of Supplemental Appendix Table A.18, we construct three alternative instruments: the share of on-duty providers who are NPs, an indicator for any NPs on duty, and the share of cases assigned to NPs (leaving out the index case) in the ED-day cell; we find the results are stable. Supplemental Appendix Table A.19 shows that our results remain similar when looking at the margin between days with no versus only one NP on duty. Supplemental Appendix Table A.20 shows that our results for hospital admissions in the ED visit and 30-day preventable hospitalizations are robust to considering hospital stays outside of the VHA.[42] Supplemental Appendix Table A.21 expands the sample to all 110 VHA EDs that use NPs—which doubles the number of cases and triples the number of providers in our sample—and shows that the results are similar.[43]

## IV. Counterfactual Scenarios

In this section, we consider two counterfactual policy scenarios that quantify the average NP-physician performance difference in financial terms. First, we consider the cost implications of substituting physicians with NPs. Second, we consider augmenting the existing supply of physicians with NPs.

We first perform a simple calculation of the cost of assigning 25 percent of all cases in VHA EDs to NPs instead of physicians, approximately the share of cases treated by NPs in our sample, which consists of EDs that are early adopters of NPs. Using the 2SLS specification in Section IIA, we estimate the effect of NPs on log total spending per case, where total spending is the sum of the three main cost components investigated in Section III: the cost of ED care, hospital admission, and 30-day preventable hospitalization.[44] Assuming that the effects of NPs across all EDs are similar to those in our sample, we multiply the extra spending per case associated with NPs by the total number of cases in the 25 percent set. Based on point estimates, we find an extra cost of \$197 million per year for the VHA.[45] This figure is 2.6 times the yearly NP wage costs that the VHA would encounter to assign 25 percent of its ED cases to NPs.[46]

The calculation ignores potential changes in provider wage costs, which may be large, as the average NP wage is only half the average physician wage (Bureau of Labor Statistics 2021a,b). If two NPs substitute for one physician, there may be no wage savings when substituting physicians with NPs. For a conservative estimate,

[42] Since a share of patients has health insurance coverage in addition to the VHA's (mainly Medicare), we report robustness checks that include hospital stays outside of the VHA for patients who enroll in both the VHA and traditional Medicare. Note that this is likely an upper-bound estimate of the possible bias for our sample since patients without non-VHA health insurance are much less likely to have hospital stays outside of the VHA. Regardless, the estimates are stable.

[43] As described in Section IC, to focus on a single margin between NPs and physicians, our main analysis focuses on EDs that use only NPs and physicians. In Supplemental Appendix Table A.21, we include all VHA EDs that use NPs regardless of using physician assistants (but exclude cases in ED-day cells with physician assistants).

[44] To calculate the cost of hospital admissions and 30-day preventable hospitalizations, we apply the cost estimate of \$19,220 per VHA hospital stay, based on the average length of stay per hospitalization at the VHA and costs per VHA inpatient day reported by the Department of Veterans Affairs (2021).

[45] As this estimation focuses on preventable hospitalization effects within 30 days of the ED visit, the extra cost estimated can be viewed as that within 30 days of the ED visit. The NP impact on total spending per case may be larger when we consider costs further downstream. The 95 percent confidence interval of the extra cost in this counterfactual is [−\$16 million, \$411 million]. The $p$-value of the estimate is 0.067. The relatively wide confidence interval mainly stems from hospital admissions (as shown in panel C of Table 2, the NP effects on ED costs and 30-day preventable hospitalizations are much more precise). In this counterfactual and Section V, we take the societal perspective and view extra medical spending as costs. While this may reflect the perspective of a vertically integrated delivery system such as the VHA, other hospitals in the United States may not fully internalize the societal cost of hiring low-productivity workers since they may not bear the cost of downstream outcomes. They may even receive profits from extra care utilization, encouraging a wage-productivity relationship different from the societal perspective.

[46] For this estimation, we divide the total number of cases in the 25 percent set by the average caseload of NPs in our sample, finding that 654 NPs would be needed for treating 25 percent of VHA's ED cases annually. We then multiply the number of NPs needed by the NP wage reported by Bureau of Labor Statistics (2021a), yielding a total wage estimate of \$74.9 million per year.

we consider the scenario where one NP may substitute for one physician. Under this scenario, we also arrive at a net cost, $129 million per year, for assigning 25 percent of VHA ED cases to NPs.[47]

Using LATEs, this analysis assesses counterfactual outcomes for compliers, likely to encompass the 25 percent of cases assigned to NPs. Yet we also consider a counterfactual scenario where EDs assign the least complex 25 percent of cases (by the number of Elixhauser comorbidities) to NPs. Applying the estimates for the lowest-complexity quartile cases in Table 5, we find a lower extra medical cost of $97 million per year to the VHA. This cost can be reduced to $29 million per year by lower NP wages when an NP can substitute for one physician, only about one-fifth of the net cost ($129 million per year) in the counterfactual above.

Despite possible extra spending, our findings do not imply that NPs are inefficient to use. When physician capacity is limited, hiring NPs to augment provider supply and reduce wait times may, nevertheless, be necessary and efficiency-improving. To examine this concept, in Supplemental Appendix A.4 we consider the trade-off between reducing wait times and increasing resource use with additional NPs. As overcrowding is a significant issue in the ED, wait time has been an important object of attention for policymakers and ED management alike (e.g., Institute of Medicine 2006; American College of Emergency Physicians 2016). We find that roughly nine-tenths of the additional spending to reduce wait time by hiring NPs comes from the increased medical resource use associated with NPs; only one-tenth comes from additional NP wage costs. However, the extra medical cost decreases with case complexity: It declines by about one-third in the lowest-complexity quartile of cases compared to the average case.

Finally, beyond the scope of this paper, there may be important long-run efficiency gains from increased NP use. For example, broader NP adoption may impact physician wages, enable greater physician specialization in complex cases or better coordination between NPs and physicians, or influence the composition of individuals entering the NP and physician professions. These may be fruitful areas for future research.

## V. Productivity within Professions

Up to this point, we have focused on estimating the average productivity difference between the professional classes of NPs and physicians. In this section, motivated by growing evidence of productivity variation across providers (within professions), we measure the distribution of productivity within each profession and ask how this intraprofessional variation in productivity compares with the difference in productivity between professions, specifically NPs versus physicians.[48]

[47] We use the physician and NP wages reported in Bureau of Labor Statistics (2021a,b) for nationally representative wage estimates. Since the Bureau of Labor Statistics does not provide separate wage estimates for NPs in the ED, we apply the wages of the average physician and the average NP. Using instead the wage estimates in the VHA data for NPs and physicians in the ED, we arrive at similar net costs of $109 million per year.

[48] Doyle, Ewer, and Wagner (2010) show differences in resource utilization decisions among physician trainees, potentially driven by human capital. Abaluck et al. (2016) documents variation in physician testing thresholds in the context of pulmonary embolism, and Silver (2021) examines returns to time spent on patients by ED physicians and variation in the physicians' productivity. Currie and MacLeod (2017) and Currie and MacLeod (2020) show physician skill variation in the contexts of C-section and depression treatments, respectively. Chan, Gentzkow, and Yu (2022) demonstrate important variation in diagnostic skill among radiologists.

We find productivity variation within each profession severalfold larger than the productivity difference between the two professions. This leads to substantial overlap between the productivity distributions of NPs and physicians; a large share of NPs perform even better than the average physician. Using detailed administrative data from the VHA, we also evaluate the extent to which productivity within each profession relates to the average health risk of cases assigned and the wages paid to individual providers, finding limited relationships.

### A. *Distribution of Productivity*

We first examine the distribution of productivity within each profession. We operationalize this examination by focusing on a measure of the total cost per case. Specifically, for each case, we aggregate the three main cost components investigated in Section III, the same components we considered in Section IV: the cost of ED care, hospital admission, and 30-day preventable hospitalization.[49] We then estimate provider effects on the log of this measure of the total cost of each case, with higher effects indicating *lower* productivity. In practice, hospitals may not fully internalize the societal cost of using higher-spending workers, and extra care utilization may be an opportunity to reap additional revenues, particularly for hospitals outside of the VHA. In this paper, we take the societal perspective: We view extra spending as costs and define higher spending effects as lower productivity.

To account for the provider-specific selection of patients in estimating provider spending effects, we use a just-identified IV model that instruments for indicators for treating providers with indicators for providers on duty in the ED-day cell of the patient's visit, leveraging the quasi-random variation in on-duty providers (conditional on $\mathbf{T}_i$). The model also controls for the baseline control vector $\mathbf{T}_i$ and patient characteristics $\mathbf{X}_i$, as in our baseline IV model in equations (1) and (2). Supplemental Appendix A.5.1 provides details of the empirical specification and shows that the instruments (i.e., the indicators for on-duty providers) are strongly predictive of the treating providers but are independent of arriving patients' characteristics conditional on $\mathbf{T}_i$, supporting the validity of these instruments.

Supplemental Appendix Table A.22 reports estimates of the variance of provider effects on the log total cost of the ED visit defined above. Using a split-sample approach to account for finite-sample estimation error, we find a variance of 0.045 for physicians and 0.048 for NPs (see Supplemental Appendix A.5.2 for details of the split-sample approach). These estimates suggest large variation in provider effects: A one standard deviation costlier physician and NP increase the total cost of the ED visit by 21 and 22 percent per case, respectively, which are about three times the average NP effect of 6.7 percent on the total cost per case defined above from the 2SLS model in equations (1) and (2). That is, the productivity variation within each professional class is much larger than the difference in productivity between the two classes. Related to the wide variation in provider effects within each professional class, Supplemental Appendix Figure A.11 shows that the NP effect varies considerably across EDs (details in Supplemental Appendix A.6). Accounting for sampling

[49] We multiply the latter two components by the average cost of a hospital stay, \$19,220.

error, the standard deviation of the ED-specific NP effect for each outcome is similar in magnitude to the average NP effect reported in Section III.[50]

We next investigate the full distributions of provider effects on the log total cost per case, applying a nonparametric empirical Bayes deconvolution approach adapted by Kline, Rose, and Walters (2022) from Efron (2016). This approach extracts a flexible empirical Bayes prior distribution of population provider effects, using the estimated provider effects and their standard errors from the just-identified IV model described above. Specifically, the procedure first yields a distribution of provider $z$-scores (the ratio of each provider effect to its standard error); it then estimates the distribution of provider effects from the $z$-score density function. Supplemental Appendix A.5.3 describes details of the estimation. We apply this procedure separately for NPs and physicians and ensure that the difference between the means of the deconvolved distributions for NPs and physicians equals the NP effect from the 2SLS model in equations (1) and (2).[51]

Panel A of Figure 8 displays the deconvolved density of provider effects. For interpretation, we convert provider effects on log total spending per case on the $x$-axis to annual spending (not including provider wages) by incorporating the empirical distribution of log total spending per case and the average number of cases a provider treats per year (see Supplemental Appendix A.5.3 for details of the conversion). Within each professional class, the productivity distribution implies greater nonwage spending of about \$910,000 and \$870,000 per year under a provider at the twenty-fifth percentile of productivity than under a provider at the seventy-fifth percentile for NPs and physicians, respectively—both of which are as high as about three times the mean annual nonwage spending difference between NPs and physicians.

Finally, using the deconvolved density of NP and physician effects on log total spending per case, we estimate the probability that a randomly drawn NP is costlier than a randomly drawn physician (Supplemental Appendix A.5.3 provides details of the estimation). We find that the probability that a randomly drawn NP is costlier than a randomly drawn physician, as measured by the effect on the total spending for the ED visit defined above, is only 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 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.

[50] Consistent with the results in Section III, we find that provider productivity is correlated with professional class. Within professions, we find no significant correlation between provider productivity and provider observable characteristics (age, gender, whether they were born in the United States, whether they received medical training in the United States (observable for physicians only), and whether they have a doctoral degree (for NPs only)). This pattern is consistent with the education literature, which suggests that teacher value added is not strongly predicted by observables (e.g., Rivkin, Hanushek, and Kain 2005; Jackson, Rockoff, and Staiger 2014). Yet we do not exclude the possibility that provider productivity may be more predictable with a richer set of characteristics.

[51] The average difference in total spending per case based on the 2SLS model and the difference between the means of the NP and physician effects from the just-identified IV model weighted by the number of cases seen by each provider are close: 6.7 versus 7.2 percent.

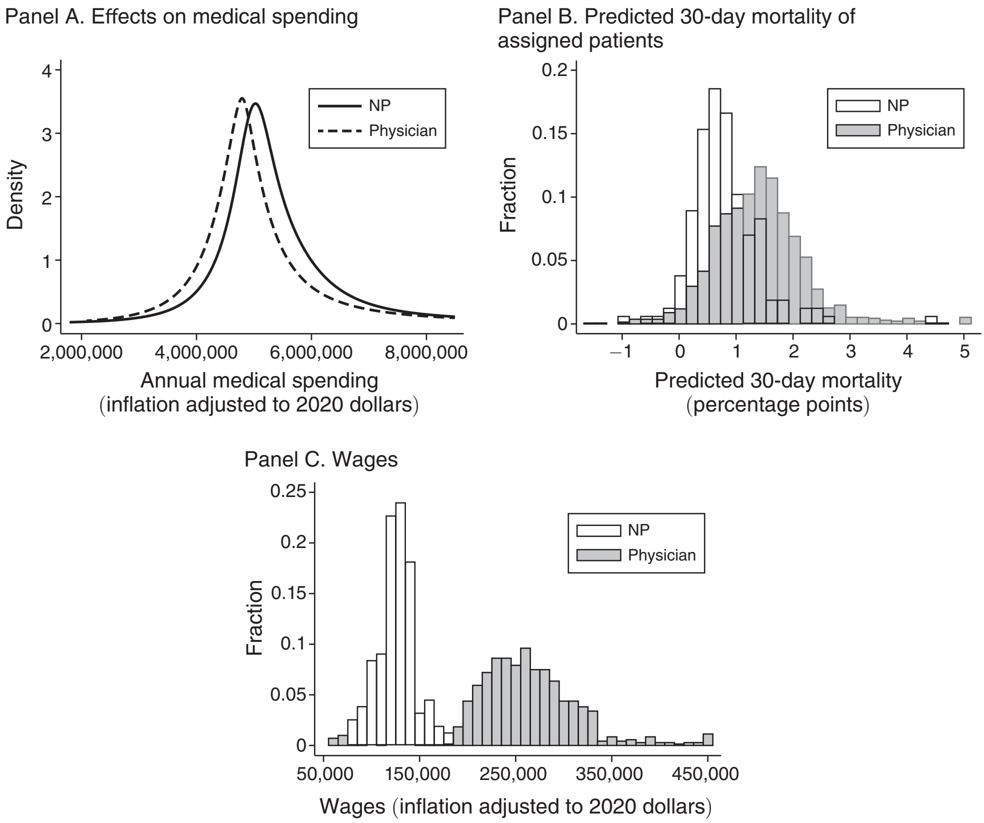


FIGURE 8. DISTRIBUTION OF PROVIDER PRODUCTIVITY, PATIENT ASSIGNMENTS, AND WAGES

*Notes*: Panel A reports the deconvolved distributions of provider effects on log total spending per case, with the provider effects on log total spending per case converted to total annual spending on the *x*-axis. See Supplemental Appendix A.5.3 for details of the estimation. The solid and dashed lines show the deconvolved distributions for NPs and physicians, respectively. Panel B plots histograms of the average predicted 30-day mortality of patients assigned to each provider. Predicted 30-day mortality is winsorized at the value of 5 percentage points. Panel C plots histograms of provider wages observed in the VHA data (inflation-adjusted to the year 2020). Wages are winsorized at the value of $450,000. In panels B and C, the white and gray bins show results for NPs and physicians, respectively.

Taken together, the findings in this section suggest a nuanced role for professions in determining workers' productivity. Despite stark differences in training and the relative complexity of cases in the ED setting, there is still substantial overlap in the productivity distributions of NPs and physicians: A large share of NPs may perform even better than the average physician. Therefore, despite the average productivity difference between the two professions, our findings do not necessarily imply inefficiency in using NPs. Not only could NPs provide a valuable labor supply for health care, but the large productivity overlap between NPs and physicians also suggests scope for efficiency gains beyond organizing work solely around professional lines. In addition, the large productivity overlap between NPs and physicians, even in the ED setting, may shed a positive light on the use of NPs in less complex settings (e.g., primary care).

### B. *Relationship with Case Assignments and Wages*

Next, we evaluate the distribution of assigned patient risks and paid wages across providers within each professional class. In panel B of Figure 8, we show the average patient risks, measured by predicted 30-day mortality, across providers in each professional class. While there is large overlap in the distributions for NPs and physicians, the difference in means between the two distributions is clear. In panel C of Figure 8, we plot the distributions of annual wages across providers.[52] The panel shows substantial wage variation within each profession: The gaps between the tenth and ninetieth percentiles of annual wages are about \$48,000 and \$117,000 for NPs and physicians, respectively. Despite the large wage variation within each profession, there is virtually no overlap in the wage distributions between the two professions.

In Supplemental Appendix Figure A.12, we present the receiver operating characteristic (ROC) curves of provider productivity, case assignments, and wage payments as characteristics that may distinguish NPs and physicians (details in Supplemental Appendix A.5.4). The area under the curve (AUC) values are 0.62, 0.75, and 0.99 for productivity, case assignments, and wage payments, respectively. Put differently, assigned patient risks are more predictive of professional class than productivity is; yearly wages are extremely predictive of professional class. Productivity, which exhibits the largest overlap between the distributions of NPs and physicians (Figure 8), has the poorest performance in classifying providers.

Finally, given the potential for large efficiency gains under policies using provider-specific productivity, we explore how our measure of provider productivity (i.e., the reverse of provider effects on the log total spending per case) relates to assigned patient risks and wages paid to each provider. To account for sampling error, we compute the empirical Bayes posterior mean for each provider's effect on log total spending per case (details in Supplemental Appendix A.5.5). In contrast to the clear gradient between the professional classes of NPs and physicians, we find that provider productivity within each professional class has little bearing on the average predicted 30-day mortality of cases assigned to each provider (panel A of Supplemental Appendix Figure A.13).[53] Similarly, we find that, within each professional class, higher productivity is not correlated with higher wages (panel B of Supplemental Appendix Figure A.13).[54] Taken together, the evidence indicates that organizations make little use of provider productivity within each professional class when assigning cases or setting wages for providers.

[52] For each provider, we access detailed payment records of the full-time equivalents and wages for each pay period from 2011 to 2020. We convert these data to annualized provider wages by (i) inflation-adjusting payments in any year to corresponding payments in 2020 dollars; (ii) computing a per hour wage by dividing the sum of (inflation-adjusted) payments by the sum of work hours across all pay periods, where each pay period covers two weeks; and (iii) multiplying this figure by 26 pay periods and 80 hours per pay period (the number of hours in a full-time-equivalent pay period).

[53] In an analysis that divides cases into low- versus high-risk by whether predicted 30-day mortality is below versus above the sample median and separately identifies provider effects on low- and high-risk patients, we find that the gap in spending effects between low- and high-productivity providers is greater for riskier patients. This implies that assigning riskier patients to higher-productivity providers could be efficient.

[54] For NPs, though panel B of Supplemental Appendix Figure A.13 shows a relationship between wages and productivity, the relationship is only marginally significant and is opposite-signed: NPs with higher effects on spending, that is, lower-productivity NPs, are paid higher wages.

## VI. Conclusion

Professionals perform some of the most important tasks across a variety of economic sectors. In turn, professional groups play a central role in determining the division of professional labor, the selection and training of future members, and the economic returns to professional work. However, very little is known empirically about productivity differences between distinct professions performing overlapping tasks and how they compare to within-profession productivity variation, largely because professions exclude other groups from providing tasks within their "jurisdictions" (Abbott 2014).

In this paper, we exploit a unique opportunity to study two starkly different professional groups—nurse practitioners and physicians. Our empirical setting allows us to study the quasi-experimental assignment of ED patients to NPs versus physicians within the Veterans Health Administration, in the context of a directive granting NPs full practice authority. We use the quasi-random arrival of patients at the ED across times that may differ in the availability of NPs on shift, which drives the probability of being treated by an NP versus a physician. In the ED setting, compared to physicians, on average, NPs use more resources during the visit (longer lengths of stay and higher spending) and exhibit a higher 30-day preventable hospitalization rate.

We further examine behavioral mechanisms and responses to productivity in our setting. We show that experience may play a role in the NP-physician performance gap. We find clinical decision-making in response to skill differences: NPs are likelier to call on external information, and they shift their prescribing decisions to avoid costly errors. We also find that the performance gap between NPs and physicians lessens when cases are less complex and less severe. In response, NPs receive healthier cases and take on a smaller share of the caseload when the ED is less constrained to meet demand. While we find in the ED that, on average, the resource cost implied by the lower productivity of NPs versus physicians outweighs salary savings from hiring NPs instead of physicians, this cost declines when NPs are assigned healthier patients. These patterns suggest returns to training, decision support, and optimal case assignment for NPs.

We find striking productivity variation within each profession, severalfold larger than the difference between the two professions. This leads to substantial overlap between the productivity distributions of NPs and physicians, such that a large share of NPs outperform the average physician. These findings imply that the professional class is, at best, a coarse indicator of productivity. While NPs and physicians receive qualitatively different patient risks and starkly different wages, within each profession, tasks and wages bear limited relationships with worker productivity. The large productivity overlap between the two professions suggests that organizing work solely around professional lines may forgo large gains in efficiency.

Considered together, our findings paint a nuanced picture of the role of professions in determining and revealing workers' productivity, tasks, and wages. Intensive professional selection and training may imply average productivity differences between professional classes that justify sizable wage differences in industries such as health care, where worker judgments may influence the utilization of considerable resources. Nevertheless, professional institutions are likely more effective at separating wages and organizing tasks between professions than at standardizing productivity within professions. These relationships may derive from frictions in observing or acting on productivity in the labor marketplace (Acemoglu and Pischke 1998), where professions may provide a function of certifying membership and organizing labor. Professional membership may sometimes be exclusive, as in the case of physicians, but it may nonetheless be a highly imperfect proxy for productivity. When organizations cannot observe or act on individual productivity, they may need to rely on second-best policies that adhere to professional lines to improve efficiency. However, if individual productivity could be better ascertained and utilized, there may be scope for even larger efficiency gains for workplaces moving beyond the instrument of the professional class.

## REFERENCES

**Abadie, Alberto.** 2003. "Semiparametric Instrumental Variable Estimation of Treatment Response Models." *Journal of Econometrics* 113 (2): 231–63.

**Abaluck, Jason, Leila Agha, Chris Kabrhel, Ali Raja, and Arjun Venkatesh.** 2016. "The Determinants of Productivity in Medical Testing: Intensity and Allocation of Care." *American Economic Review* 106 (12): 3730–64.

**Abbott, Andrew.** 2014. *The System of Professions: An Essay on the Division of Expert Labor*. University of Chicago Press.

**Acemoglu, Daron, and David Autor.** 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." In *Handbook of Labor Economics*, Vol. 4B, edited by David Card and Orley Ashenfelter, 1043–171. Elsevier.

**Acemoglu, Daron, and Jörn-Steffen Pischke.** 1998. "Why Do Firms Train? Theory and Evidence." *Quarterly Journal of Economics* 113 (1): 79–119.

**Adhvaryu, Achyuta, Namrata Kala, and Anant Nyshadham.** 2022. "Management and Shocks to Worker Productivity." *Journal of Political Economy* 130 (1): 1–47.

**Agency for Healthcare Research and Quality.** 2021. "Prevention Quality Indicators Technical Specifications Updates." Agency for Healthcare Research and Quality. https://qualityindicators.ahrq.gov/archive/pqi_techspec/icd10_v2021 (accessed October 12, 2022).

**Alexander, Diane, and Molly Schnell.** 2019. "Just What the Nurse Practitioner Ordered: Independent Prescriptive Authority and Population Mental Health." *Journal of Health Economics* 66: 145–62.

**Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber.** 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 113 (1): 151–84.

**American Association of Nurse Practitioners.** 2020. "The Path to Becoming a Nurse Practitioner (NP)." American Association of Nurse Practitioners. https://www.aanp.org/news-feed/explore-the-variety-of-career-paths-for-nurse-practitioners (accessed February 23, 2022).

**American Association of Nurse Practitioners.** 2021. "State Practice Environment." American Association of Nurse Practitioners. https://www.aanp.org/advocacy/state/state-practice-environment (accessed February 23, 2022).

**American College of Emergency Physicians.** 2016. "Emergency Department Crowding: High Impact Solutions." American College of Emergency Physicians. https://www.acep.org/globalassets/sites/acep/media/crowding/empc_crowding-ip_092016.pdf (accessed February 20, 2022).

**Anderson, D. Mark, Ryan Brown, Kerwin Kofi Charles, and Daniel I. Rees.** 2020. "Occupational Licensing and Maternal Health: Evidence from Early Midwifery Laws." *Journal of Political Economy* 128 (11): 4337–83.

**Andrews, Michelle, and McKenzie Beard.** 2024. "Like Doctors, More Nurse Practitioners Are Heading into Specialty Care." *Washington Post*, June 17. https://www.washingtonpost.com/politics/2024/06/17/like-doctors-more-nurse-practitioners-are-leaving-primary-care-behind/.

**Ashley, Jo Ann.** 1976. *Hospitals, Paternalism, and the Role of the Nurse*. Teachers College Press.

**Association of American Medical Colleges.** 2020. "The Road to Becoming a Doctor." Association of American Medical Colleges. https://www.aamc.org/system/files/2020-11/aamc-road-to-becoming-doctor-2020.pdf (accessed February 25, 2022).

**Association of American Medical Colleges.** 2021. *Physician Specialty Data Report*. Association of American Medical Colleges. https://www.aamc.org/data-reports/workforce/data/number-people-active-physician-specialty-2021 (accessed April 25, 2024).

**Beck, Andrew H.** 2004. "The Flexner Report and the Standardization of American Medical Education." *JAMA* 291 (17): 2139–40.

**Benkard, Lanier.** 2000. "Learning and Forgetting: The Dynamics of Aircraft Production." *American Economic Review* 90 (4): 1034–54.

**Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan.** 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics* 119 (1): 249–75.

**Bhuller, Manudeep, Gordon B. Dahl, Katrine V. Løken, and Magne Mogstad.** 2020. "Incarceration, Recidivism, and Employment." *Journal of Political Economy* 128 (4): 1269–324.

**Brown, E. Richard.** 1979. *Rockefeller Medicine Men: Medicine and Capitalism in America*. University of California Press.

**Bureau of Labor Statistics.** 2021a. *Occupational Employment and Wages, May 2020* (*Nurse Practitioners*). Bureau of Labor Statistics. https://www.bls.gov/oes/2020/may/oes291171.htm (accessed February 16, 2024).

**Bureau of Labor Statistics.** 2021b. *Occupational Employment and Wages, May 2020* (*Physicians, All Other; and Ophthalmologists, Except Pediatric*). Bureau of Labor Statistics. https://www.bls.gov/oes/2020/may/oes291228.htm (accessed February 16, 2024).

**Butler, Christopher C., David Gillespie, Patrick White, Janine Bates, Rachel Lowe, Emma Thomas-Jones, Mandy Wootton, et al.** 2019. "C-Reactive Protein Testing to Guide Antibiotic Prescribing for COPD Exacerbations." *New England Journal of Medicine* 381 (2): 111–20.

**Cairns, Christopher, and Kai Kang.** 2022. "National Hospital Ambulatory Medical Care Survey: 2019 Emergency Department Summary Tables." https://dx.doi.org/10.15620/cdc:115748 (accessed February 3, 2023).

**California Health and Human Services Agency.** 2025. *Rates of Preventable Hospitalizations for Selected Medical Conditions by County* (*LGHC Indicator*). California Health and Human Services Agency. https://data.chhs.ca.gov/dataset/rates-of-preventable-hospitalizations-for-selected-medical-conditions-by-county (accessed April 20, 2025).

**Cameron, A. Colin, and Douglas L. Miller.** 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources* 50 (2): 317–72.

**Carrell, Scott, and James West.** 2010. "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors." *Journal of Political Economy* 118 (3): 409–32.

**Chan, David C.** 2016. "Teamwork and Moral Hazard: Evidence from the Emergency Department." *Journal of Political Economy* 124 (3): 734–70.

**Chan, David, and Yiqun Chen.** 2026. *Data and Code for: "The Productivity of Professions: Evidence from the Emergency Department."* American Economic Association; distributed by Inter-university Consortium for Political and Social Research. https://doi.org/10.3886/E241348V1.

**Chan, David C., Matthew Gentzkow, and Chuan Yu.** 2022. "Selection with Variation in Diagnostic Skill: Evidence from Radiologists." *Quarterly Journal of Economics* 137 (2): 729–83.

**Chan, David C., and Jonathan Gruber.** 2020. "Provider Discretion and Variation in Resource Allocation: The Case of Triage Decisions." *AEA Papers and Proceedings* 110: 279–83.

**Chen, Yiqun.** 2021. "Team-Specific Human Capital and Team Performance: Evidence from Doctors." *American Economic Review* 111 (12): 3923–62.

**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): 2593–632.

**Currie, Janet, and W. Bentley MacLeod.** 2017. "Diagnosing Expertise: Human Capital, Decision Making, and Performance Among Physicians." *Journal of Labor Economics* 35 (1): 1–43.

**Currie, Janet, and W. Bentley MacLeod.** 2020. "Understanding Doctor Decision Making: The Case of Depression Treatment." *Econometrica* 88 (3): 847–78.

**Currie, Janet, and Jonathan Zhang.** 2025. "Doing More with Less: Predicting Primary Care Provider Effectiveness." *Review of Economics and Statistics* 107 (2): 289–305.

**Dahl, Gordon B., Andreas Ravndal Kostøl, and Magne Mogstad.** 2014. "Family Welfare Cultures." *Quarterly Journal of Economics* 129 (4): 1711–52.

**Department of Veterans Affairs.** 2021. *HERC Inpatient Average Cost Data*. Department of Veterans Affairs. https://www.herc.research.va.gov/include/page.asp?id=inpatient (accessed February 23, 2022).

**Dillender, Marcus, Anthony T. Lo Sasso, Brian J. Phelan, and Michael R. Richards.** 2022. "Occupational Licensing and the Healthcare Labor Market." NBER Working Paper 29665.

**Doyle, Joseph J., Steven M. Ewer, and Todd H. Wagner.** 2010. "Returns to Physician Human Capital: Evidence from Patients Randomized to Physician Teams." *Journal of Health Economics* 29 (6): 866–82.

**Efron, Bradley.** 2016. "Empirical Bayes Deconvolution Estimates." *Biometrika* 103 (1): 1–20.

**Elixhauser, Anne, Claudia Steiner, D. Robert Harris, and Rosanna M. Coffey.** 1998. "Comorbidity Measures for Use with Administrative Data." *Medical Care* 36 (1): 8–27.
**Epstein, Andrew, and Sean Nicholson.** 2009. "The Formation and Evolution of Physician Treatment Styles: An Application to Cesarean Sections." *Journal of Health Economics* 28 (6): 1126–40.
**Fairman, Julie.** 2009. *Making Room in the Clinic: Nurse Practitioners and the Evolution of Modern Health Care*. Rutgers University Press.
**Farronato, Chiara, Andrey Fradkin, Bradley Larsen, and Erik Brynjolfsson.** 2020. "Consumer Protection in an Online World: An Analysis of Occupational Licensing." NBER Working Paper 26601.
**Fleming-Dutra, Katherine E., Adam L. Hersh, Daniel J. Shapiro, Monina Bartoces, Eva A. Enns, Thomas M. File, Jonathan A. Finkelstein, et al.** 2016. "Prevalence of Inappropriate Antibiotic Prescriptions Among US Ambulatory Care Visits, 2010-2011." *JAMA* 315 (17): 1864–73.
**Flexner, Abraham.** 1910. *Medical Education in the United States and Canada: A Report to the Carnegie Foundation for the Advancement of Teaching.* Carnegie Foundation for the Advancement of Teaching.
**Freidson, Eliot.** 1974. *Professional Dominance: The Social Structure of Medical Care*. Transaction Publishers.
**Gennaioli, Nicola, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer.** 2013. "Human Capital and Regional Development." *Quarterly Journal of Economics* 128 (1): 105–64.
**Gottlieb, Joshua D., Maria Polyakova, Kevin Rinz, Hugh Shiplett, and Victoria Udalova.** 2020. "Who Values Human Capitalists' Human Capital? Healthcare Spending and Physician Earnings." CES Working Paper CES-20-23.
**Gowrisankaran, Gautam, Keith Joiner, and Pierre Thomas Léger.** 2023. "Physician Practice Style and Healthcare Costs: Evidence from Emergency Departments." *Management Science* 69 (6): 3202–19.
**Hallett, Christine E., ed.** 2016. *The History of Nursing. Major Themes in Health and Social Welfare*. Routledge.
**Health Resources and Services Administration.** 2010. *The Registered Nurse Population: Findings from the 2008 National Sample Survey of Registered Nurses*. Health Resources and Services Administration. https://data.hrsa.gov/DataDownload/NSSRN/GeneralPUF08/rnsurveyfinal.pdf (accessed April 5, 2024).
**Huang, David T., Donald M. Yealy, Michael R. Filbin, Aaron M. Brown, Chung-Chou H. Chang, Yohei Doi, Michael W. Donnino, et al.** 2018. "Procalcitonin-Guided Use of Antibiotics for Lower Respiratory Tract Infection." *New England Journal of Medicine* 379 (3): 236–49.
**Imbens, Guido W., and Donald B. Rubin.** 1997. "Estimating Outcome Distributions for Compliers in Instrumental Variables Models." *Review of Economic Studies* 64 (4): 555–74.
**Institute of Medicine.** 2006. *Hospital-Based Emergency Care: At the Breaking Point*. National Academies Press.
**Jackson, C. Kirabo, Jonah E. Rockoff, and Douglas O. Staiger.** 2014. "Teacher Effects and Teacher-Related Policies." *Annual Review of Economics* 6 (1): 801–25.
**Kleiner, Morris M., and Alan B. Krueger.** 2013. "Analyzing the Extent and Influence of Occupational Licensing on the Labor Market." *Journal of Labor Economics* 31 (2): S173–202.
**Kleiner, Morris M., Allison Marier, Kyoung Won Park, and Coady Wing.** 2016. "Relaxing Occupational Licensing Requirements: Analyzing Wages and Prices for a Medical Service." *Journal of Law and Economics* 59 (2): 261–91.
**Kleinpell, Ruth, Michelle L. Cook, and Diane L. Padden.** 2018. "American Association of Nurse Practitioners National Nurse Practitioner Sample Survey: Update on Acute Care Nurse Practitioner Practice." *Journal of the American Association of Nurse Practitioners* 30 (3): 140–49.
**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.
**Larson, Magali S.** 1979. *The Rise of Professionalism: A Sociological Analysis*. University of California Press.
**Laurant, Miranda, David Reeves, Rosella Hermens, Jose Braspenning, Richard Grol, and Bonnie Sibbald.** 2005. "Substitution of Doctors by Nurses in Primary Care." *Cochrane Database of Systematic Reviews* 2: CD001271.
**Lazuka, Volha.** 2018. "The Long-Term Health Benefits of Receiving Treatment from Qualified Midwives at Birth." *Journal of Development Economics* 133: 415–33.
**Liu, Chuan-Fen, Paul L. Hebert, Jamie H. Douglas, Emily L. Neely, Christine A. Sulc, Ashok Reddy, Anne E. Sales, and Edwin S. Wong.** 2020. "Outcomes of Primary Care Delivery by Nurse Practitioners: Utilization, Cost, and Quality of Care." *Health Services Research* 55 (2): 178–89.
**Lutfiyya, May Nawal, Lisa Tomai, Bianca Frogner, Frank Cerra, Daniel Zismer, and Stephen Parente.** 2017. "Does Primary Care Diabetes Management Provided to Medicare Patients Differ between Primary Care Physicians and Nurse Practitioners?" *Journal of Advanced Nursing* 73 (1): 240–52.

**Markovits, Daniel.** 2020. *The Meritocracy Trap: How America's Foundational Myth Feeds Inequality, Dismantles the Middle Class, and Devours the Elite*. Penguin.

**McMichael, Benjamin J., and Sara Markowitz.** 2023. "Toward a Uniform Classification of Nurse Practitioner Scope of Practice Laws." *Medical Care Research and Review* 80 (4): 444–54.

**Milbank Memorial Fund.** 2024. "No One Can See You Now Five Reasons Why Access to Primary Care Is Getting Worse (and What Needs to Change)." Milbank Memorial Fund. https://www.milbank.org/wp-content/uploads/2024/02/Milbank-Scorecard-2024-APPENDIX_v02.pdf (accessed June 4, 2024).

**Morgan, Perri A., David H. Abbott, Rebecca B. McNeil, and Deborah A. Fisher.** 2012. "Characteristics of Primary Care Office Visits to Nurse Practitioners, Physician Assistants and Physicians in United States Veterans Health Administration Facilities, 2005 to 2010: A Retrospective Cross-Sectional Analysis." *Human Resources for Health* 10 (1): 42–49.

**National Sample Survey of Registered Nurses.** 2022. *2022 National Sample Survey of Registered Nurses*. National Sample Survey of Registered Nurses. https://data.hrsa.gov/topics/health-workforce/nursing-workforce-survey-data\#RegisteredNurses (accessed May 25, 2024).

**Neuman, Mark D., Brian T. Bateman, and Hannah Wunsch.** 2019. "Inappropriate Opioid Prescription After Surgery." *Lancet* 393 (10180): 1547–57.

**New Jersey Department of Health.** 2011. "Health Care Quality Assessment." New Jersey Department of Health. https://www.nj.gov/health/healthcarequality/documents/qi/pqi2011.pdf (accessed May 16, 2025).

**Nyweide, David J., Denise L. Anthony, Julie P.W. Bynum, Robert L. Strawderman, William B. Weeks, Lawrence P. Casalino, and Elliott S. Fisher.** 2013. "Continuity of Care and the Risk of Preventable Hospitalization in Older Adults." *JAMA Internal Medicine* 173 (20): 1879–85.

**Patel, Kant, and Mark E. Rushefsky.** 2004. *The Politics of Public Health in the United States*. M.E. Sharpe.

**Perloff, Jennifer, Catherine M. DesRoches, and Peter Buerhaus.** 2016. "Comparing the Cost of Care Provided to Medicare Beneficiaries Assigned to Primary Care Nurse Practitioners and Physicians." *Health Services Research* 51 (4): 1407–23.

**Perry, John J.** 2009. "The Rise and Impact of Nurse Practitioners and Physician Assistants on their Own and Cross-Occupation Incomes." *Contemporary Economic Policy* 27 (4): 491–511.

**Rivkin, Steven G., Eric A. Hanushek, and John F. Kain.** 2005. "Teachers, Schools, and Academic Achievement." *Econometrica* 73 (2): 417–58.

**Shapiro, Carl.** 1986. "Investment, Moral Hazard, and Occupational Licensing." *Review of Economic Studies* 53 (5): 843–62.

**Silver, David.** 2021. "Haste or Waste? Peer Pressure and Productivity in the Emergency Department." *Review of Economic Studies* 88 (3): 1385–417.

**Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick.** 2019. "Capitalists in the Twenty-First Century." *Quarterly Journal of Economics* 134 (4): 1675–745.

**Stange, Kevin.** 2014. "How Does Provider Supply and Regulation Influence Health Care Markets? Evidence from Nurse Practitioners and Physician Assistants." *Journal of Health Economics* 33: 1–27.

**Starr, Paul.** 1982. *The Social Transformation of American Medicine: The Rise of a Sovereign Profession and the Making of a Vast Industry*. Basic Books.

**Traczynski, Jeffrey, and Victoria Udalova.** 2018. "Nurse Practitioner Independence, Health Care Utilization, and Health Outcomes." *Journal of Health Economics* 58: 90–109.