NBER WORKING PAPER SERIES

THE PRODUCTIVITY OF PROFESSIONS: EVIDENCE FROM THE EMERGENCY DEPARTMENT

David C. Chan Jr
Yiqun Chen

Working Paper 30608
http://www.nber.org/papers/w30608

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2022

We are grateful to Ricardo Alonso, Sandy Black, Marika Cabral, David Card, Shooshan Danagoulian, Qing Gong, Joshua Gottlieb, Mitch Hoffman, Bapu Jena, Amanda Kowalski, Brad Larsen, Darren Lubotsky, Neale Mahoney, 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, Damien Dong, Akriti Dureja, Jesse Kozler, Matthew Merrigan, Jonatas Prates, Aadit Shah, Kemin Wang, Justine Weng, and Saam Zahedian provided excellent research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Productivity of Professions: Evidence from the Emergency Department
David C. Chan Jr and Yiqun Chen
NBER Working Paper No. 30608
October 2022
JEL No. I11,I18,J24,J44,M53

ABSTRACT

Professions play a key role in determining the division of labor and the returns to skilled work. This paper studies the productivity difference between physicians and nurse practitioners (NPs), two health care professions performing overlapping tasks but with stark differences in background, training, and pay. Using data from the Veterans Health Administration and quasi-experimental variation in the patient probability of being treated by physicians versus NPs in the emergency department, we find that, compared to physicians, NPs significantly increase resource utilization but achieve worse patient outcomes. We find evidence suggesting mechanisms relating to lower human capital among NPs relative to physicians and worker-task assignment responding to the lower skill of NPs. Counterfactual analysis suggests a net increase in medical costs with NPs, even when accounting for NPs' wages that are half as much as physicians'. Despite large productivity differences between professions, we find even larger productivity differences within professions and substantial productivity overlap between professions. Yet there is little overlap in wages between NPs and physicians and, within professions, no significant correlation between productivity and wages.

David C. Chan Jr
Department of Health Policy
117 Encina Commons
Stanford, CA 94305
and NBER
david.c.chan@stanford.edu

Yiqun Chen
Department of Economics
University of Illinois at Chicago
601 South Morgan Street, 708 UH
Chicago, IL 60607
and NBER
yqchen@uic.edu

1 Introduction

Professions play a key role in determining the division of labor and returns to skilled work. As described by sociologists, professional groups engage in activities that impact both the "jurisdictions" of tasks that each profession controls and the economic returns to this work (Abbott 2014). Designs imposed by professional groups to select and train future professionals may at once restrict the supply of professionals and impact their quality. Professional groups may also lobby policymakers and negotiate with other interest groups. Therefore, it is theoretically ambiguous whether potentially large differences in pay between professional groups 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 increase in the demand for health care, outstripping the supply of physicians, and the rise of a class of professionals from the nursing tradition—nurse practitioners (NPs)—seeking to 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 US (Bureau of Labor Statistics 2021*a*,*b*), and various states in the US have responded to the shortfall in physicians by granting NPs "scope of practice" to perform tasks traditionally performed by physicians (e.g., McMichael and Markowitz forthcoming). Yet on the basis of training, income, and social class, the comparison between the two professions of NPs and physicians is stark.

In this paper, we exploit a quasi-experiment in the Veterans Health Administration (VHA) to study the productivity differences between NPs and 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). 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.

Along a variety of measures, we find that NPs use more resources and achieve worse health outcomes than physicians. Our IV results show that NPs increase length of stay by 11 percent and raise the cost of ED care by 7 percent. While we can rule out large effects on inpatient admission and 30-day mortality in the overall sample, we find that NPs raise 30-day preventable hospitalizations by 20 percent. In contrast to our IV estimates, ordinary least squares (OLS) estimates for the benchmark outcomes of length of stay and ED costs are negative in sign, consistent with the descriptive evidence that NPs treat healthier patients than physicians do.

We undertake a range of analyses to assess the validity of our IV quasi-experiment. We show that, conditional on our baseline controls, a broad range of patient characteristics that predict outcomes are well balanced across values of our instrument, the number of NPs on duty. 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 (still opposite-signed relative to IV estimates) of those when only baseline covariates are included. We also assess the 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 a range of characteristics of on-duty physicians and NPs. We also examine potential spillovers between NPs and physicians and find no evidence suggesting such spillovers.[1] Finally, we show that our results are robust to controlling for a series of other factors that may vary with the instrument, including the total level of available staff, patient volume, and patient wait time.

Next, using several analytical lenses, we uncover mechanisms and responses related to the lower productivity of NPs versus physicians. First, we find that experience matters: The NP-physician gap in resource utilization narrows among providers who have seen more prior patients, both in general and for the diagnosis in question. This suggests that differences in training may play some role in the productivity differences between NPs and physicians. Second, we document clinical decision-making responses to lower skill among NPs. Compared to physicians, NPs are likelier to gather (costly) information from other sources: They are likelier to obtain radiology tests and formal consults for their patients. 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. Third, we examine

[1] Specifically, we examine whether NP presence may affect physician performance. E.g., if NPs ask physicians for assistance, it could slow down physicians. We thus examine whether outcomes of patients treated by physicians are affected by the presence of NPs; we find no such evidence. We also examine whether 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.

heterogeneity in the NP-physician gap by patient condition complexity and severity. The NP-physician gap in patient log length of stay, log ED cost, and the probability of inpatient admission grows for patients with more comorbidities and higher severity. Patient-provider assignment seems to respond to NPs' relative and absolute disadvantage: On average, NPs are assigned healthier patients, of those available for assignment; they are also assigned a (modestly) smaller share of patients when the ED is less busy.

To understand the policy implications of the productivity differences between NPs and physicians, we perform two counterfactual analyses. First, we compare the current allocation of approximately a quarter of the patients being treated by NPs in all VHA EDs and the counterfactual of staffing the EDs only with physicians. Relative to the counterfactual of no NPs, the current allocation increases spending by $160 million per year due to higher resource utilization and worse outcomes in the next 30 days. Accounting for lower NP wages, even under the conservative lower bound that one NP can perform the work of one physician, we arrive at a net cost of $74 million per year under the current allocation. That is, despite having wages half that of physicians' wages, NPs are still more costly to employ. As a second counterfactual analysis, we consider the scenario in which hospitals cannot employ additional physicians and can increase provider labor only through additional NPs. The primary benefit of increasing the pool of providers in the ED is to increase throughput, decreasing wait times. In this counterfactual analysis, given the lower productivity of NPs relative to physicians, increasing the number of NPs on duty decreases wait time but increases resource utilization and adverse outcomes. Strikingly, the wage costs of hiring additional NPs account for only one fifth of the total cost to reduce wait times; the lower productivity of NPs accounts for four fifths of this cost.

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 flexible underlying prior 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. Exploiting detailed wage data in the VHA, we find little correlation between productivity and wages within professional class and little overlap in the wage distributions between physicians and NPs. In other words, despite our previous results, wages are highly predictive of professional class, while productivity is much less predictive.

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, heated debates have arisen around the quality of NP-provided

care and on whether NPs should be permitted to substitute for physicians. A recent body of research has investigated the state-level impacts of liberalizing "scope of practice laws" for NPs.[2] By design, these papers study 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. In contrast, our study sheds light on the effect of assigning patients to NPs versus physicians.[3] Our results on the heterogeneity of this effect by case complexity and riskiness imply an optimal use of NPs by organizations—i.e., assign NPs healthier patients and a smaller share of patients when the ED is less busy—and our descriptive evidence points in the direction of this hypothesis of skill-task matching (Acemoglu and Autor 2011).

Second, our research relates to the widespread practice of occupational licensing.[4] The existing literature suggests that occupational licensing increases the earnings of licensed workers (Kleiner and Krueger 2013; Kleiner et al. 2016; Farronato et al. 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 et al. 2020 for a review). Two studies of an earlier, unregulated environment of midwifery, near the beginning of the 20th century, demonstrate meaningful reductions in maternal and infant mortality with the initial implementation of occupational licensing (Lazuka 2018; Anderson et al. 2020). Studies in this literature compare differences in quality within professions (along the margin of occupational licensing), while ours compares two competing professions differing in their historical origins, social status, income, and selection and training processes, not to mention licensing.

A third related literature is concerned with worker human capital and productivity. These issues have received growing attention in the health care setting (e.g., Doyle, Ewer, and Wagner 2010; Currie and MacLeod 2017, 2020; Chen 2021; Chan, Gentzkow, and Yu 2022) and more broadly (e.g., Chetty, Friedman, and Rockoff 2014; Moretti 2004; Gennaioli et al. 2013). Relative to this literature, our study is unique in that it compares the productivity of two distinct classes of professionals, with human capital differences that

[2]The findings of this literature are varied and somewhat mixed. Perry (2009) and Kleiner et al. (2016) find that these laws impact physician and 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. However, 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. McMichael and Markowitz (forthcoming) note that these papers may adopt different definitions of scope of practice laws.

[3]An older medical literature has raised this question, but the small numbers of providers and other features of these earlier studies limit inference on systematic differences between the 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 largest study contains 1,465 patients (Kinnersley et al. 2000); all studies involve at most a handful of clinic locations. The studies tend to focus on primary care settings but 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."

[4]More than a thousand occupations and about a third of all jobs in the US require some form of licensing or certification (Kleiner and Krueger 2013).

may stem from 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. Building on the prior literature on practice variation (e.g., Epstein and Nicholson 2009; Gowrisankaran, Joiner, and Léger 2017), we find that, strikingly, the variation in productivity across providers within professions may be even larger than the differences between professions.

Fourth, a broad set of questions is concerned with 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 et al. 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 that is in fact larger than wage differences, at least in our resource-intensive and information-dependent setting within health care. Yet, we find potentially even larger productivity variation within professions that may not be captured in worker-specific wages, consistent with frictions in observing productivity across similar workers from outside of the firm (e.g., Acemoglu and Pischke 1998). If worker-specific productivity is difficult to observe, then professional class and the actions of organized professional institutions, as richly documented by sociologists (e.g., Starr 1982; Abbott 2014), may matter more than a worker's individual productivity in setting her wage. Entering a profession may represent a costly and imperfect way for workers to distinguish themselves (Spence 1973).

The remainder of this paper is organized as follows. Section 2 describes the institutional setting and the data. Section 3 describes our empirical approach and provides evidence for its validity. Section 4 provides our main results. Section 5 presents evidence on mechanisms and responses. Section 6 presents analyses on policy-relevant counterfactual scenarios. Section 7 reports productivity and wage distributions within professions. Section 8 concludes.

2 Background and Data

2.1 Physicians and NPs in the US

To understand the physician and NP professions in the US, 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 (Larson 1979; Brown 1979).[5]

The "professionalization" of American nursing began in the 1870s, with the establishment of the first three nursing schools in the US. 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). Unlike American physicians, who maintained their autonomy from (and exercised influence over) hospitals and payers, professional nurses were generally employed by hospitals and subject to downward wage pressure from hospitals (Staiger, Spetz, and Phibbs 2010; Maggs 2016).

NPs emerged from the nursing tradition in the 1960s and 1970s, in the setting of increasing specialization in medicine, worsening access to care in urban and rural areas, and new federal funding from Medicare and Medicaid to increase the training of providers (Fairman 2009; Hallett 2016). In the 1980s and 1990s, rising health care costs and pressures to improve the throughput of health care further expanded the boundaries of NP practice in a large variety of settings (e.g., primary care, emergency care) (Fairman 2009; Kleinpell, Cook, and Padden 2018). The growth of the NP workforce has been a distinctly American phenomenon.[6]

Present-day NPs and physicians remain starkly different in terms of training, income, and social class, despite performing overlapping tasks in many settings. Physicians undergo a highly selective process and long periods of training to enter the profession. They comprise the single most common profession in the top percentile of the income distribution (Gottlieb et al. 2020). About half of medical students come from families in the top quintile of the income distribution, while only 5 percent of medical students come from families in the bottom quintile (Kahn and Sneed 2015). In contrast, the income of NPs is roughly half of that of physicians, and the number of years of training is also roughly half.[7] Admission rates to nursing programs are around 10 times higher than admission rates to medical school, and nursing has been highlighted as a realistic path to the middle class for women of working-class backgrounds (Friedman, Laurison, and

[5]Much of this transformation centered around the Flexner (1910) Report, which strongly advocated for a scientific orientation of medical education and the exclusion of alternative practices (Beck 2004). Following the report, more than half of medical schools closed or consolidated (Patel and Rushefsky 2004). All but two of the historically Black medical schools closed as a result (Sullivan and Suez Mittman 2010); medical schools, which had begun admitting women, reverted to male-only admittance.

[6]In a survey of 39 countries, only 11 countries (the Netherlands and 10 English-speaking countries) granted NPs clinical autonomy on seven clinical dimensions and defined educational requirements to become an NP (Maier and Aiken 2016). Of the eight of these countries studied in greater detail, only the US, Canada, and the Netherlands included NPs in their health care workforce planning (Maier et al. 2018). As of 2016, the US had 175,000 NPs, representing 5.6 percent of nurses, while the next two highest countries, the Netherlands and Canada, had 2,700 and 3,600 NPs, representing 1.5 and 1.3 percent of nurses in these countries, respectively (Maier et al. 2016).

[7]According to the American Academy of Medical Colleges (2020), physicians must complete a four-year undergraduate degree, a four-year Doctor of Medicine (MD) degree, and three to seven years of residency training. According to the American Association of Nurse Practitioners (2020), NPs are required to complete a four-year Bachelor of Science in Nursing degree and may choose between one to two year(s) in a Master of Science in Nursing degree or three to four years in a Doctor of Nursing Practice degree.

Macmillan 2017; Searcey, Porter, and Gebeloff 2015).[8]

2.2 VHA Setting

In December 2016, the VHA granted full practice authority to NPs. The policy enables NPs to practice without any requirement for physician supervision at VHA facilities. NPs can treat patients as independently as physicians, regardless of state restrictions that would otherwise limit NPs' practice authority.[9]

Several features make the ED a setting well suited to study the productivity of providers. First, each patient visit is generally assigned to a single ED provider. This setup allows us to attribute patient outcomes to individual providers and compare productivity across 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 types of patients arriving. Third, patients present at the ED with a wide spectrum of conditions, ranging in complexity and severity. This provides an opportunity to investigate productivity across a range of tasks. Finally, the ED has become a major setting that uses the NP workforce. Across the nation, 8 percent of ED visits were seen by NPs between 2010 and 2017 (Wu and Darracq 2021). In the VHA, the share of ED visits seen by an NP has steadily increased to 11 percent in 2019—close to the share of visits seen by NPs in primary care at around 20 percent (Morgan et al. 2012).[10]

In Appendix Table A.1, we report characteristics of NPs and physicians working at the 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). Appendix Table A.1 also compares NPs practicing at the ED and the overall NP workforce, showing that ED NPs are less likely to be female (79 versus 90 percent) and are slightly younger (43 versus 45 years old).

2.3 Data

We use administrative health records from the VHA, the largest health care delivery system in the US, serving more than nine million veterans. For each ED visit, the data allow us to identify the type of provider

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

[9] As of 2021, about half of the states in the US have not granted NPs full practice authority (American Association of Nurse Practitioners 2021).

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

treating the patient (i.e., NP or physician) as well as resource use and patient outcomes (e.g., length of stay, 30-day preventable hospitalization). The data also contain detailed information on patient characteristics that include demographics, comorbidities, vital signs, and prior health care use, as well as information on provider characteristics such as birth date and gender.

Sample Construction. We restrict our analysis sample in the following ways, summarized in Appendix Table A.2. First, we restrict the sample to ED visits between January 2017 and January 2020, that is, after full practice authority was granted to NPs at the VHA and before the onset of the COVID pandemic in the US. Second, we include only cases arriving during the daytime (8 a.m. to 6 p.m.), because the data show that few NPs take evening or night shifts.[11] Third, we focus on visits to VHA EDs using NPs to treat patients and in months after the ED adopted the full practice authority policy. Though the VHA granted full practice authority to NPs at all VHA facilities, local facilities varied in when they adopted the policy and whether they used NPs in the ED.[12] Fourth, to examine a single margin between NPs and physicians, we exclude EDs that used non-physician providers other than NPs (mainly physician assistants).[13] Finally, we drop a small number of cases with missing age or gender or whose age was below 20 or above 99 years. The final sample contains 1.1 million cases over 44 EDs, seen by a total of 156 NPs and 1,348 physicians.

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 independently next).[14] We also include hospital admission, a resource-intensive option that indicates the provider's decision to admit the patient for inpatient care. To measure 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 Agency for Healthcare Research and Quality (2021). We exclude from 30-day preventable hospitalization the inpatient admissions immediately following the ED visit, as they reflect the hospital admission decision described above.

In examining mechanisms behind the effect of NPs, we include the following sets of outcomes: (i)

[11] One possible explanation is that, since patient volumes are on average 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 a.m. and 6 p.m. versus only 0.9 outside of 8 a.m.-6 p.m.), EDs in our sample often staff only one provider for a shift in the evening/night, resulting in a low probability of having NPs in those times since NPs may not be able to handle very severe patients arriving.

[12] We define a VHA ED as having granted full practice authority to NPs and 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.

[13] In addition, unlike NPs, physician assistants had not been granted full practice authority at the VHA as of 2021.

[14] For length of stay, we use detailed time-stamped data on patient assignment and discharge to estimate length of stay. For the cost of ED care, we rely on detailed accounting by the VHA that considers utilization during the ED visit.

whether the provider orders consults from other providers; (ii) whether the provider orders CT scans and X-rays, two primary diagnostics in the ED; and (iii) prescriptions of opioids and antibiotics—two major classes of drugs whose clinical indications for appropriate use are often unclear and may require skill to discern (e.g., Fleming-Dutra et al. 2016; Huang et al. 2018; Neuman, Bateman, and Wunsch 2019).

Descriptive Statistics. Table 1 summarizes characteristics of the cases included in our analysis. In addition to demographics, we measure patient comorbidities as Elixhauser et al. (1998) indices, which are 31 indicators for comorbidities (e.g., cancer, diabetes) that are predictive of clinical outcomes, based on patient medical histories in the prior 365 days. We also report average length of stay, average ED spending (inflation adjusted to 2020 dollars), and the 30-day preventable hospitalization rate. Column 1 shows characteristics for 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 overall 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 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 (0.7 versus 1.4 percentage points). This pattern suggests that simple comparisons of patient outcomes between NPs and physicians may be confounded by differences in underlying patient health.

3 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 a case is treated by an NP or a physician. In this section, we begin by describing our instrumental variables (IV) approach. We then discuss evidence that supports the validity of our identification strategy.

3.1 Specification

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

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

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

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 (i.e., the number of NPs on duty between 8 a.m. and 6 p.m.) at the ED on the day that case i visits.[15] The parameter of interest is δ, which represents a local average treatment effect (LATE), i.e., the average causal effect among cases that would have been assigned to a different type of providers under a different number of NPs on duty.

The vector $\mathbf{T}_i$ encodes interactions between indicators for the ED and indicators for time categories of the patient's arrival, specifically, the year, the month, the day of the week, and the hour of the day of the patient's arrival. We condition on $\mathbf{T}_i$ to allow for 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.[16]

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 three-digit ICD-10 code of patient primary diagnosis of the visit.[17] For each patient covariate with missing values, we add an indicator for missing values and replace missing values with zero. Finally, ε_i and v_i are error terms. We cluster standard errors by provider. In robustness checks, we also show results under alternative clustering approaches, including clustering by ED-day (the level across which the instrument varies) and, more conservatively, two-way clustering by provider and ED-day, neither of which meaningfully affects our results.

3.2 Identification

To interpret δ as the LATE of being treated by NPs, our IV approach requires four identifying assumptions: first stage, conditional independence, exclusion, and monotonicity. In this section, we summarize empirical

[15] Since the data do not include direct information on provider scheduling, we measure Z_i as the number of NPs treating cases 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 a.m. and 6 p.m. in the ED-day cell. In the main analysis, we include the index case in calculating Z_i. In Section 4.5, we show the robustness of our estimates to an alternative Z_i that includes NPs with only one case in a shift as well as one that leaves out the index case in defining whether an NP is on duty.

[16] While fixed effects for the interactions between indicators for the ED and indicators for the hour of the day 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.

[17] Since our study period is from January 2017 to January 2020, disease diagnoses are all 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 below in Section 4, our estimates are remarkably stable regardless of controlling for three-digit diagnosis indicators or not. In Appendix A.1, we also show that NPs and physicians appear to be similar in their coding of three-digit diagnoses.

evidence that supports the validity of the identifying assumptions.

First Stage. Figure 1 shows the first stage of our IV model, controlling for the baseline controls, i.e., ED-by-time-category indicators, $\mathbf{T}_i$. Panel A shows that the number of physicians on duty declines linearly with the number of NPs on duty. Consequently, Panel B shows that patient probability of being treated by an NP increases with the number of NPs on duty: One more NP on duty increases patient probability of being treated by NPs by 18.6 percent. The increase is highly significant (with an F-statistic of 149.2, conditioning on ED-by-time-category indicators) and is close to linear.

To provide context, 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-days have zero, one, and two NP(s) on duty, respectively; 3.4 percent of ED-days have more than two NPs on duty. A possible question is what drives the variation in the number of NPs on duty. NPs are less likely to work on weekends (77 percent of weekday ED-day observations versus 24 percent of weekend ED-day observations have NPs on duty). However, conditional on day-of-the-week and other time-category (year and month) indicators, we still observe substantial variation in the number of NPs on duty within EDs across days (standard deviation of 0.49).[18]

Conditional Independence. For our instrument to be valid, the number of NPs on duty must be uncorrelated with patient potential outcomes, conditional on our vector of baseline controls, $\mathbf{T}_i$. 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 $\mathbf{T}_i$. For completeness, Appendix Figure A.2 reports similar coefficients for the instrument using each of the various patient characteristics included in $\mathbf{X}_i$ as the dependent variable. Despite the fact that these characteristics are strong predictors of patient outcomes (F-statistics around 100 for joint significance, even controlling for ED-time-category indicators, see Appendix Figure A.3), there is little significant relationship between our instrument and the broad range of patient characteristics, conditioning on $\mathbf{T}_i$.

As a second set of evidence, we examine the stability of our IV estimates under different sets of controls for patient covariates. Specifically, we divide patient observable characteristics into eight groups and estimate separate regressions that control for each of the $2^8 = 256$ different combinations of patient covariates.[19] We

[18] Scheduling in the ED is usually done several months in advance. Providers state shifts on which they will be available to work, and scheduling is often done manually by administrators with the aid of rudimentary software (e.g., spreadsheets). Quasi-random variation in provider scheduling has been exploited in other hospital settings (e.g., Chan 2021; Chen 2021).

[19] We divide patient characteristics into the following eight groups: (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 three-digit patient primary diagnosis of the visit.

show in empirical results below that controlling for any combination of patient covariates results in virtually no change in our IV estimates of the NP effect. Following the logic of Altonji, Elder, and Taber (2005), this evidence implies limited selection bias due to either observed or unobserved patient characteristics that are predictive of 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 patient probability of being treated by NPs, not through any other channels. After discussing our main results, we present in Section 4.4 empirical evidence to support this assumption in our setting. We show that both physicians' and NPs' characteristics are well balanced across the number of NPs on duty conditional on $\mathbf{T}_i$, and there appears to be little evidence of spillovers from NPs to physicians (e.g., NPs may ask physicians for help, which may slow down physicians). We also investigate a series of alternative explanations, finding little evidence indicating 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, i.e., 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.

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 Appendix Figure A.4, 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, total number of Elixhauser comorbidities, and predicted 30-day mortality.[20] Appendix Figure A.4 shows that for all subsamples, the first-stage estimates are positive and statistically different from zero, consistent with the validity of the monotonicity assumption.[21]

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

[21] An interesting pattern is that the first-stage coefficients appear to be larger for healthier patients. Appendix Figure A.4 shows that the first-stage coefficient is larger for patients who are younger, have a lower number of Elixhauser comorbidities, and have a lower predicted 30-day mortality. This pattern reflects that compliers are more heavily concentrated among healthy patients. In Section 4.3, we compute characteristics of compliers and never-takers. We find consistent evidence that compliers are healthier than the overall sample, while never-takers are riskier.

4 Main Results

In this section, we present our main findings, showing the effect of NPs on resource use and patient outcomes. We find that, compared to physicians, NPs use more medical resources: They require longer lengths of stay and incur higher costs. However, they achieve less favorable patient outcomes, as measured by 30-day preventable hospitalizations. We also characterize compliers relative to the overall sample, present evidence supporting the exclusion restriction, and consider a series of additional robustness checks.

4.1 Length of Stay and Cost

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. Figure 3 shows the reduced-form effect of the instrument (i.e., the number of NPs on duty) against patient log length of stay and log cost of the ED visit, controlling for our 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 figure patient predicted log length of stay and predicted log cost, both of which are well balanced across the instrument.[22]

Table 2 reports 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 reduced-form coefficients on the instrument for these outcomes. All regressions control for the full set of controls described in Section 3.1. The OLS estimates (Columns 1 and 4) show that cases treated by NPs have significantly shorter lengths of stay and lower costs, which could reflect that NPs treat healthier and easier cases than do physicians, at least in terms of observable characteristics shown in Table 1. Exploiting plausibly quasi-random variation in patient probability of being treated by NPs, the IV estimates (Columns 3 and 6) suggest that NPs raise patient medical resource use during 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 the mean of the sample, the NP effect equals an 18-minute increase in length of stay and a $66 increase in cost per ED visit. Appendix Figure A.5 shows a visual IV plot of the NP effect on length of stay and cost.

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 three-digit primary diagnosis of the visit. We

[22] We form these predictions using linear regressions of actual outcomes on the full set of patient controls that include five-year age-bin indicators, marital status, gender, race, indicators for 31 Elixhauser comorbidities, prior health care use, vital signs, and indicators for three-digit ICD-10 code of patient primary diagnosis of the visit.

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 (i.e., $\mathbf{T}_i$), 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 attenuates to 10 percent. The lower health risks of cases treated by NPs (Table 1) and the sensitivity of the OLS estimates to patient controls suggest selection bias due to patient unobservable characteristics. In contrast, the IV estimates are remarkably robust to controlling for any combination of patient covariates: Regardless of the additional controls, the IV estimates for the effect of NPs 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.[23]

4.2 Hospital Admission and Patient Outcomes

Having examined resource use in the ED, we next assess NP effects on hospital admission and downstream patient outcomes, in Table 3. In our overall sample, the hospital admission rate does not differ significantly between NPs and physicians, but we show in Section 5 that NPs increase admissions for severe cases. We find no significant effect of NPs on 30-day mortality for most cases, but we find evidence in Section 5 for increases in mortality for a subset of highly severe cases.[24]

Finally, we find a significant NP effect on increasing patient 30-day preventable hospitalizations: Compared to physicians, NPs raise patient 30-day preventable hospitalization rate by 0.25 percentage points, which is equivalent to a 20 percent increase compared to the mean of the sample. The higher 30-day pre-

[23] Another potential explanation for the divergence between the OLS and IV estimates is heterogeneity in treatment effects, since OLS reports the average effect of NPs 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 developed by Bhuller et al. (2020) and reweight the analysis sample to match the sample of compliers using predicted 30-day mortality, i.e., a composite index of all patient observables. The OLS estimates with the reweighting still differ in sign compared with 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.

[24] We find a marginally significant and clinically meaningful NP-driven increase in mortality for a highly severe type of patients: those with a sepsis diagnosis (point estimate: 24.5 percentage points, p-value: 0.106). 30-day mortality among patients with a sepsis diagnosis is 11.5 percentage points, as against 1.25 percentage points for the average patient.

ventable hospitalization rate of NPs may reflect two possibilities: (i) NPs have poorer decision-making over whom to admit to the hospital, resulting in under-admission of patients who should have been admitted and a net increase in return hospitalizations, despite NPs using longer lengths of stay to evaluate patients' need for hospital admission; (ii) NPs produce lower quality of care conditional on admitting decisions, despite spending more resources on treating the patient (as measured by costs of the ED care). Both possibilities imply lower skill of NPs relative to physicians.

Taken together, empirical evidence suggests that NPs and physicians operate on different production functions: NPs use more inputs (longer lengths of stay and higher costs), but achieve less favorable patient outcomes (higher 30-day preventable hospitalization rates). In other words, comparing NPs and physicians as two professional classes, NPs exhibit lower productivity than do physicians. Yet note that this evidence may not indicate that we can cut back on care for NPs without compromising patient outcomes. A lower production function may still exhibit positive returns to inputs; that is, higher intensity of care may be allocatively efficient for NPs (Chandra and Staiger 2007; Silver 2021; Chan, Gentzkow, and Yu 2022).

4.3 Complier Characteristics

Our IV estimates represent the LATE, i.e., the average causal effect among complier cases that are quasi-randomly assigned to NPs versus physicians due to the instrument. To better understand this LATE, we examine complier characteristics relative to the overall sample following the approach developed by Abadie (2003), as described in Appendix A.2. Appendix Table A.3 reports the results. Consistent with NPs treating less severe cases, we find that compliers are healthier than the average case. Compared to the average case, compliers are younger, have a lower number of Elixhauser comorbidities, have a lower number of inpatient stays in the prior year, and exhibit lower predicted mortality. Appendix Table A.3 also examines characteristics of never-takers (of NPs), following an approach from Dahl, Kostøl, and Mogstad (2014) that we detail in Appendix A.2.[25] In line with the notion that NPs treat healthier cases than physicians do, we find that never-takers are riskier than the average case, both of which are riskier than compliers.

4.4 Exclusion Restriction

As discussed in Section 3.2, 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 the patient probability of being treated by NPs, not through any other channels. In this section, we present evidence supporting the validity of the

[25] As described in Appendix A.2, we define never-takers as cases treated by physicians in ED-day cells with the residual share of cases treated by NPs at least as high as the 90th percentile of ED-days with at least one case treated by NPs. There are no (conventional) always-takers in our setting since no patients can be assigned to NPs on days without NPs.

exclusion restriction. We first show that the characteristics of physicians on duty are well balanced across the number of NPs on duty. Second, we show that NP characteristics remain stable regardless of the number of NPs on duty. Third, we assess and find little support for the possibility of productivity spillovers between NPs and physicians in our setting (e.g., NPs may ask physicians in the ED for help, which may slow down physicians, or NP peer identity affects physician productivity). Finally, we examine a range of factors that may vary across days and find no evidence to suggest these factors are driving our IV estimates.

Balance in Provider Characteristics. To start, we investigate whether 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 of physicians. Figure 5 reports the balance for various physician characteristics. Specifically, we consider an ED-day level analysis 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.[26]

We examine three sets of physician characteristics: (i) demographics of age and gender; (ii) measures 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 Appendix A.3 for details). Figure 5 shows that each of these physician characteristics is well balanced across the instrument, conditional on the baseline controls. Panels A and B show that on-duty physicians' average age and gender composition are remarkably stable across the instrument. Panels C-E show that, while there is large variation in physician value-added and practice style, the average value-added and practice style of physicians on duty is constant despite the instrument.

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.[27] Following Figure 5, Figure 6 shows that an analogous set of average NP characteristics are well balanced across days with differing numbers of NPs, conditional on baseline controls.

[26]Specifically, 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. A potential question is, since we have only 44 EDs, 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 by using Wild cluster bootstrap as suggested by Cameron and Miller (2015); we find no meaningful change in the standard error estimates.

[27]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 4.5, 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.

Assessing Productivity Spillovers. We then consider the possibility of spillovers between NPs and physicians. If NPs ask physicians in the ED for assistance, this could slow down physicians. Alternatively, a change in peers from days without any NP to days with NPs may influence physicians' treatment decisions as, for example, 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 might expect the outcomes of patients treated by NPs to depend on the quality of physicians on duty. Using the value-added measure described in Appendix A.3 as a proxy for physician quality, we find in Appendix Table A.4 no such relationship. Second, with spillovers (either through the assistance or peer pressure channel), 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 look at cases arriving between 5 and 8 a.m., i.e., 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.[28] As shown in Appendix Table A.5, Panel A, we find no evidence of spillovers from (subsequent) NP presence on physicians' patients overall; in Panel B, we focus on days with high workloads, when NP spillovers may be more detectable, and again find no evidence of spillovers in this subsample.

Robustness to Additional Factors. Finally, we investigate factors that may vary across days, including the total number of cases arriving, the total number of physician equivalents on duty, and patient wait times (defined as the time between arrival at the ED and assignment to a treating provider). We control for the total number of physician equivalents on duty to mitigate the concern that the effective level of providers may vary across days with different numbers of NPs on duty.[29] Turning to wait time, since patient wait time is potentially endogenous (healthier cases could be assigned a lower priority and thus wait longer), we instrument for wait time using the average wait time of cases visiting on the same day at the same ED as the index case. While potentially important, the factors listed above do not affect our estimates: Appendix Table

[28] To further restrict the possibility of patient selection between NPs and physicians, we exclude patients arriving between 5 and 8 a.m. in ED-day cells with any patient assigned to NPs.

[29] We calculate the number of physician equivalents on duty as the sum of the number of on-duty physicians and the number of on-duty NPs multiplied by 0.341, where 0.341 is the coefficient reported in Panel A of Figure 1 and assumed to be the extent of substitution between NPs and physicians. As a robustness check, we also apply a more conservative substitution rate of 0.5; the results are stable. We do not directly control for the total number of providers on duty because, conditional on the total number of providers, a higher number of NPs indicates lower staffing (since one NP appears unable to substitute for one physician given that NPs take longer to discharge patients but do not seem to handle more patients simultaneously); therefore, the 2SLS estimates for the NP effect conditional on the total number of on-duty providers would be confounded by lower staffing.

A.6 shows that our IV estimates are remarkably robust to controlling for these factors.

We also ask if the estimated NP effect is driven by patient-provider gender mismatch, since the majority of patients are male (91 percent), while physicians are primarily male (74 percent) and NPs are mostly female (79 percent). Appendix Table A.7 explores this possibility by asking whether NPs treat patients of the opposite gender differently compared to the same gender. The results show little heterogeneity.

4.5 Additional Robustness Checks

Appendix Tables A.8-A.11 report additional robustness checks. Appendix Table A.8 shows that our findings are stable with alternative standard error clustering approaches: clustering by ED-day or two-way clustering by ED-day and provider. The standard errors become smaller when clustering by ED-day compared to the baseline model that clusters by provider, but no conclusion on statistical significance is changed. The standard errors change only minimally with two-way clustering by ED-day and provider relative to the baseline model clustering by provider only. Panels A and B of Appendix Table A.9 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 two cases) in the analysis time window of an ED-day cell and another count that includes any NP with any case besides the index case.[30] In Appendix Table A.9, Panels C-D, we construct two alternative instruments—the share of cases in the ED-day cell treated by NPs (leaving out the index case) and an indicator for any NP on duty; we show the results are stable. Appendix Table A.10 shows that our results remain similar when looking at the margin between when there is no NP versus only one NP on duty. Appendix Table A.11 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.[31]

5 Mechanisms and Responses

The evidence in the previous section suggests that NPs have lower productivity than physicians: They use more medical resources and produce worse patient outcomes. In this section, we examine mechanisms and responses related to this productivity gap. First, we show that experience may play a role behind the productivity gap. Second, we show NP responses to lower skill in their clinical decision-making, in calling on external resources and setting prescription thresholds; these responses in turn manifest as lower productivity.

[30] See Section 3.1 and footnote 15 for explanations for these two robustness checks.

[31] Since a proportion of patients have 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.

Third, we show that the NP effect is larger for more complex and more severe patients, suggesting a comparative disadvantage for NPs in treating these patients. Finally, we show patient assignment responds toward optimality given the NP comparative and absolute disadvantage: NPs receive healthier patients and take on a lower share of the caseload when the ED is less constrained to meet demand.

5.1 Provider Experience

First, we ask whether experience impacts the magnitude of the performance difference between NPs and physicians. The professions of NPs and physicians entail stark differences both in the training that new members undergo and in the selectivity of choosing new members. Whether physicians are more productive than NPs because of their training or their innate ability has clear policy relevance. 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, then we may see that the gap narrows with experience. On the other hand, if the gap derives from lower innate ability, then we may see that the gap persists or even widens with experience.

We form measures of both general and specific experience. We measure general experience as the number of cases the provider has treated since 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 three-digit primary diagnosis that is the same as the current case the provider has treated since 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:

$$y_i = \delta_1 \text{NP}_i \times \text{Experience}_i + \delta_2 \text{NP}_i + \delta_3 \text{Experience}_i + \mathbf{T}_i \eta + \mathbf{X}_i \beta + \varepsilon_i. \tag{3}$$

Experience_i denotes standardized experience within each provider type.[32] We instrument for NP_i and $\text{NP}_i \times \text{Experience}_i$ using Z_i (i.e., the number of NPs on duty) and $Z_i \times \text{Experience}_i$.

We find that, for several outcomes, experience predicts a smaller NP-physician performance gap. Panel A of Table 4 examines the impact of specific experience. For length of stay, Column 1 indicates that a one-standard-deviation increase in specific experience among both NPs and physicians lowers the performance gap by 5.8 percent, reducing the gap at the mean levels of experience (10.1 percent) by more than half. The NP-physician gaps in costs, CT scan ordering, and consult ordering similarly declines with spe-

[32] 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.

cific experience. However, specific experience does not appear to reduce the NP-physician gap in 30-day preventable hospitalizations. Panel B shows a similar pattern for general experience in the outcomes of length of stay and consult ordering. Increasing general experience by one standard deviation in both NPs and physicians is associated with a 10 percent decline in the NP-physician performance gap in length of stay and a 2 percentage-point reduction in the gap in consult ordering. Overall, general experience is correlated with a reduced NP-physician performance gap in fewer dimensions than is specific experience, possibly reflecting task-specific human capital that accrues to a greater degree within tasks than across tasks (Gibbons and Waldman 2004).[33]

We examine several alternative measures of experience to address measurement concerns. One concern is that, because we do not observe cases treated by providers since the start of their careers, our measures of experience are imperfect representations of providers' true levels of experience. To mitigate this concern, we restrict the sample to cases visiting after January 2018, so that our measures of experience have at least a one-year look-back window. We also measure experience based on cases seen in the prior year (i.e., in the 365 days before the day of the current case's visit), so that the estimates precisely represent heterogeneity by prior-year experience.[34] 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 have an effect on 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. Appendix Tables A.12 to A.14 show that our estimates remain qualitatively similar under these alternative measurements.

5.2 Clinical Decision-Making

Next, we examine clinical decision-making that may respond to the lower human capital, particularly lower diagnostic skill, of NPs. With lower diagnostic skill, providers may draw on more formal resources, such as

[33] Although the magnitude of the decline in the NP-physician difference in length of stay and consult ordering is larger with a one-standard-deviation increase in general experience than with a one-standard-deviation increase in specific experience, it is not necessarily that general experience is more effective in reducing the NP-physician performance gap in these two outcomes. As described above, a one-standard-deviation increase in general experience equals 1,855 and 1,258 cases for NPs and physicians, respectively, while a one-standard-deviation increase in specific experience equals a smaller number, 108 and 107 cases for NPs and physicians, respectively.

[34] Specifically, we measure general experience as the number of cases the provider treated in the proceeding 365 days; we measure specific experience as the number of cases with the same three-digit ICD-10 primary diagnosis as the current case the provider treated in the proceeding 365 days. We exclude from this regression cases visiting in the first year of our analysis period, since we cannot fully observe their providers' experience in the proceeding 365 days.

consults and radiology tests. They may also adjust their treatment thresholds for decisions with asymmetric costs between type I and type II errors (Chan, Gentzkow, and Yu 2022). While these represent appropriate responses for providers with lower human capital, they increase the cost of providing care, manifesting in lower productivity.

Informational Resources. Columns 1-3 of Table 5 report the effect of NPs on the use of informational resources, using the 2SLS estimation specified in Equations (1) and (2). Column 1 shows that, 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, which are the two primary diagnostics in the ED. 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 are. This could directly increase lengths of stay as well as medical costs, since consults and diagnostics take time and resources to complete. On the other hand, consults and diagnostics allow lower-skilled providers to improve decision-making by incorporating more information and interpretation from other experts.

Prescription Thresholds. 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 in the face of less information. Specifically, providers with less information may more frequently opt for a 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 a treatment when false positives are costlier than false negatives.

We choose two important prescriptions with different asymmetries in the costs of type I and type II errors: opioids and antibiotics. For both of these prescriptions, the clinical indications for appropriate use are often unclear and require clinical judgment.[35] We estimate the NP-physician difference in prescriptions using the 2SLS estimation specified in Equations (1) and (2). Column 4 of Table 5 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 have a lower prescription likelihood relative to physicians: NPs lower opioid prescriptions by 1.8 percentage

[35] See, e.g., Fleming-Dutra et al. (2016), Huang et al. (2018), Butler et al. (2019), Neuman, Bateman, and Wunsch (2019).

points, or 20 percent of the sample mean. In contrast, for antibiotics, which have higher false-negative costs—e.g., non-treatment of a potentially life-threatening infection compared to antibiotic resistance—NPs show a higher prescription likelihood relative to physicians: NPs increase antibiotic prescriptions by 4.0 percentage points, or 6.3 percent of the sample mean.[36] The joint evidence from prescription thresholds is consistent with NPs responding to lower skill.[37]

5.3 Case Complexity and Severity

In this section, we exploit the wide variety of cases that arrive at the ED to examine heterogeneous 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:

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

where $y_i \cdot \mathrm{NP}_i$ is the interaction between patient outcome and the indicator for being treated by an NP, $\mathbf{1}(\mathrm{Group}_i = g)$ 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 $\left\{\mathbf{1}(\mathrm{Group}_i = g)\right\}_{g=1}^{G}$ and NP_i by interacting $\left\{\mathbf{1}(\mathrm{Group}_i = g)\right\}_{g=1}^{G}$ 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 $y_i \cdot (\mathrm{NP}_i - 1)$.

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 by whether condition severity measured by 30-day mortality of the three-digit diagnosis is equal to or above the 95th percentile of the sample. The 30-day mortality rate for this top-severity group is 8.6 percentage points, compared to 1.2 percentage points for the whole analysis sample. Included in this group are cases with relatively severe conditions such as heart failure and acute kidney failure, potentially requiring higher human capital to manage.[38]

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

[37] The different treatment thresholds between NPs and physicians (consults, diagnostic tests, and prescriptions) could reflect NPs responding to lower skill, perceiving that they have lower skill (i.e., being less confident in their skill), or exhibiting different risk preferences without any difference in skill. While we cannot rule out the latter two explanations, the pattern that NPs achieve less favorable outcomes (as measured by 30-day preventable hospitalizations) despite using more resources supports lower skill of NPs.

[38] Appendix Table A.15 summarizes the 10 most common three-digit diagnosis codes in this group. The largest diagnosis category is heart failure, followed by acute kidney failure. The top 10 diagnoses also include acute myocardial infarction, a form of sepsis,

As shown in Figure 7, the effect of NPs on raising lengths of stay and medical costs grows with case complexity and severity. For example, for cases in the lowest complexity quartile, NPs increase their length of stay by about 5 percent, while for cases in the highest complexity quartile, NPs increase their length of stay by around 25 percent. For cases with a condition severity at least as high as the 95th percentile, we find an NP effect that doubles their length of stay. For these cases, NPs also increase their hospital admission rate by about 30 percentage points, nearly a 100 percent increase from the potential admission rate under physicians. Appendix Table A.16 summarizes the heterogeneity in treatment effects.[39] The table further investigates the NP effect among four severe conditions with high mortality: stroke, acute myocardial infarction (AMI), sepsis, and heart failure. We consistently find larger NP effects among cases with these severe conditions than among other cases.

We do not find increasing NP effects on 30-day preventable hospitalizations with case complexity or severity (Column 5 of Appendix Table A.16). If NPs are less skilled at treating more complex or more severe cases, they may obtain worse patient outcomes for these cases. On the other hand, as NPs increase their intensity of care for these cases, the incremental care could mitigate the NP effect on raising preventable hospitalizations. As shown in Panel A of Appendix Table A.16, for cases in the highest complexity quartile, NPs sizably increase their lengths of stay and medical costs, without leading to a significant change in 30-day preventable hospitalizations. In contrast, for cases in the lower complexity quartiles, NPs reveal a smaller-magnitude and sometimes insignificant effect on lengths of stay and medical costs, but significantly raise 30-day preventable hospitalizations.

5.4 Patient Assignment

Finally, we examine patient assignment between NPs and physicians. On the whole, the evidence on case heterogeneity in Section 5.3 suggests a comparative disadvantage for NPs in treating complex and severe cases. We do not find a set of cases in which NPs outperform physicians, which suggests that NPs are also at an absolute disadvantage on average 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 have bandwidth to see them.

Appendix Figure A.6 provides descriptive insight into patient assignment by exploiting variation in NP staffing (i.e., the instrument) and variation in patient arrivals, conditional on our baseline controls. Panels

and a form of respiratory failure. The 30-day mortality rate ranges between 5 and 17 percentage points.

[39] 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: $y_i = y_i \cdot \mathrm{NP}_i - y_i \cdot (\mathrm{NP}_i - 1))$.

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, despite the average complexity and severity of all cases remaining stable regardless of NP staffing. 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.[40] Panel D of the figure assesses the probability that a patient is assigned to an NP as a function of the number of other patients arriving in the analysis time window (i.e., 8 a.m. to 6 p.m.) of the ED-day cell. We find a modest but clear trend in which NPs are more likely to be assigned patients when the ED is busier, consistent with NPs providing some degree of back-up labor when physicians are unavailable due to external shocks.[41]

6 Counterfactual Scenarios

In this section, we consider two counterfactual policy scenarios. First, we use estimates from previous sections to consider overall cost implications of substituting physicians with NPs by assigning 25 percent of cases across VHA EDs to NPs, instead of physicians. Second, we perform auxiliary analyses to consider the policy of augmenting the existing supply of physicians with NPs. Overall, these analyses highlight that productivity differences between classes of workers may have even larger cost implications than the sizable differences in wages.

6.1 Substituting Physicians with NPs

We first perform a simple calculation of the cost of assigning 25 percent of all cases in VHA EDs to NPs—approximately the share of cases treated by NPs in our analysis sample, which consists of EDs that are early adopters of NPs under full practice authority. We assume that treatment effects of NPs across all EDs would be similar to that in our sample and consider three components of extra costs due to NPs: the costs of ED care (Section 4.1), the costs of hospital admission for severe cases (Section 5.3), and the costs

[40] As a result, cases left for physicians also become riskier with more NPs. A potential question is whether the increased average risk of patients treated by physicians on days with more NPs may affect physicians' overall productivity, violating an exclusion restriction. To assess this concern, Panel F of Appendix Table A.6 controls for the average health risk (predicted 30-day mortality) of patients treated by physicians in the ED-day cell. The results show that our IV estimates are highly robust.

[41] 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 4.4, we also show that our IV estimates are highly robust to controlling for the number of patients arriving.

of 30-day preventable hospitalizations (Section 4.2). We find an extra cost of $160 million per year for the VHA.[42] This figure is approximately twice the yearly NP wage costs that the VHA would encounter to assign 25 percent of its ED cases to NPs.[43]

The calculation ignores potential changes in provider wage costs, although the average NP wage is only half of the average physician wage (Bureau of Labor Statistics 2021*a*,*b*). If two NPs substitute for one physician, which could be within the possible range given the coefficient reported in Panel A of Figure 1, there would be no wage saving when substituting physicians with NPs. For a conservative estimate, we consider the scenario in which one NP may substitute for one physician. Under this scenario, we nonetheless find net costs of $74 million per year for the VHA for the policy of assigning 25 percent of cases to NPs.

In using previously estimated LATEs from our quasi-experiment, this analysis is well suited to counterfactual outcomes for compliers, likely to encompass the 25 percent of cases that would be assigned to NPs. Alternatively, we consider a more conservative counterfactual scenario where EDs assign the least complex 25 percent of cases (by the number of Elixhauser comorbidities) to NPs. We note that such an allocation may not be always feasible: For example, in hours when all arriving patients are relatively complex, EDs may have to assign some complex patients to NPs. Applying the estimates for the lowest complexity quartile patients in Appendix Table A.16, we calculate extra costs of $81 million per year to the VHA. To overcome these extra costs through lower wages, an NP would need to be able to substitute for about 0.85 physicians.

6.2 Augmenting Provider Supply with NPs

While our analysis up to this point has centered on substituting physicians with NPs, much of the policy motivation for hiring NPs has been to augment provider supply. When physician capacity is limited, it may nevertheless be efficient to hire additional NPs to improve throughput and reduce wait times. To examine this concept, we consider the trade-off between reducing wait times and worsening resource use—measured by ED length of stay and total cost per case—induced by additional NPs. As overcrowding is a significant issue in EDs, 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).

[42] To calculate the cost of increased 30-day preventable hospitalizations and hospital admissions in the ED visit, we apply the cost estimate of $19,220 per VHA hospital stay, on the basis of the average length of stay per hospitalization at the VHA and costs per VHA inpatient day reported by the VHA's Health Economics Resource Center (2021).

As this cost estimation focuses on preventable hospitalization effects within 30 days of the ED visit, the extra cost estimated can be viewed as that in the 30 days of the ED visit. To the extent that the cost-increasing effect of NPs may accrue over time and extend into other dimensions of post-ED care, the extra spending with using NPs may be larger than the estimate reported above.

[43] For this back-of-the-envelope estimation, we divide the total number of cases in the 25 percent set by the average caseload of NPs in our sample, finding that 827 NPs would be needed for treating 25 percent of VHA's ED cases annually. We then multiply the number of NPs needed with the average wage of NPs reported by Bureau of Labor Statistics (2021*a*), yielding a total wage estimate of $94.7 million per year.

In this auxiliary analysis, we hold fixed the number of cases arriving and the number of physicians on duty, and ask how additional NPs may affect patient wait time and downstream outcomes. We use the following empirical design:

$$y_i = \sum_{n=0}^{N} \delta_n \times \mathbf{1}(Z_i = n) + N_i^c \gamma_1 + N_i^p \gamma_2 + \mathbf{T}_i \eta + \mathbf{X}_i \beta + \varepsilon_i. \tag{5}$$

$\mathbf{1}(Z_i = n)$ is an indicator for $n \in \{0, \ldots, 5\}$ NPs being on duty at the ED on the day case i visits.[44] N_i^c and N_i^p are, respectively, the number of cases arriving and the number of physicians on duty at the ED on the day case i visits. We apply Equation (5) to several outcomes of interest: (i) wait time (i.e., the time between patient arrival at the ED and assignment to a provider); (ii) ED length of stay (i.e., the time between assignment to a provider and discharge from the ED); (iii) cost of ED care; (iv) hospital admission; and (v) 30-day preventable hospitalization.

Figure 8 presents the trade-off based on estimates of $\{\delta_n\}_{n=0}^{5}$ in Equation (5) across outcomes. Panel A shows the trade-off between wait time and ED length of stay. The first dot plots δ_0 (i.e., no NP on duty) for wait time on the y-axis and δ_0 for length of stay on the x-axis. Each of the subsequent dots plots a pair of estimates corresponding to $n \in \{1, \ldots, 5\}$ with the estimate for wait time on the y-axis and the estimate for ED length of stay on the x-axis. Panel B similarly plots the trade-off between wait time and total cost per case. To arrive at the latter, we apply a cost estimate of $19,220 per VHA hospital stay, as in Section 6.1, and calculate the sum of the coefficients for the three cost-related outcomes: the cost of ED care, the cost of hospital admission (for high-severity patients),[45] and the cost of 30-day preventable hospitalization.

As implied in Figure 8, decreasing wait time by 30 minutes per case by hiring additional NPs would increase length of stay by 16 minutes per case (Panel A), as well as increase total medical spending by 15 percent, or about $238 per case, not including the cost of additional NP wages (Panel B).[46] Including the wage costs of additional NPs would bring this figure to about $300 per case.[47] This suggests that roughly four fifths of the additional spending to reduce wait time by hiring NPs come from the lower productivity of NPs, while only one fifth comes from additional NP wage costs.

[44] The maximum of n is 6, but only a small share of ED-days has 5 or 6 NPs on duty (Appendix Figure A.1). We therefore group $n = 5$ and $n = 6$ together.

[45] As shown in Section 5.3, we only observe significant effects of NPs on increasing hospital admission in the ED visit among severe cases, i.e., cases with a three-digit ICD-10 diagnosis whose 30-day mortality is equal to or above the 95th percentile of the sample; thus, in calculating the extra cost due to increased hospital admission in the ED visit, we include only that incurred by the severe cases.

[46] The results in Figure 8 show that a 34-minute decrease in wait time increases lenghth of stay by 18 minutes and total spending by $268 per case (comparing the first and last dot of each panel). We rescale the tradeoff to a 30-minute decrease in wait time for ease of interpretation.

[47] For this wage cost estimation, we divide the yearly NP wage by the average number of cases an NP treats per year, and then multiply this figure by the number of additional NPs required.

7 Productivity and Wage Distributions

Up to this point, we have focused on estimating the average productivity difference between the professional classes of NPs and physicians. We show that this difference is large, likely even larger than the difference in average wages between the two professions, despite the average physician wage being double the average NP wage. In this section, we turn to variation in productivity and wages within professions. Motivated by growing evidence of productivity variation across providers (within profession), we ask how this intra-professional variation in productivity compares to the difference in productivity between professions, in the case of NPs versus physicians.[48] Using the detailed data of the VHA, we then ask how this variation in productivity relates to variation in wages within profession.

We operationalize this examination by focusing on a measure of total cost per case. Specifically, for each case, we aggregate the three components of resource utilization in which we find significant NP effects, the same components we previously considered in Section 6: ED costs, hospital admission, and 30-day preventable hospitalizations (we multiply the latter two components by the average cost of a hospital stay, $19,220). We then estimate provider effects on the log of this measure of total cost associated with each ED visit. To account for provider selection of patients (Chang and Obermeyer 2020), we use a just-identified IV model that instruments for indicators for treating providers with indicators for on-duty providers in the ED-day cell of the patient's visit. Appendix A.4.1 describes details of the estimation, and shows that these instruments are strongly predictive of the treating providers but are independent of arriving patients' characteristics conditional on our baseline controls, supporting the validity of these instruments.

Appendix Table A.17 reports estimates of the variance of provider effects on log total medical spending associated with the ED visit defined above. Using a split-sample approach to account for measurement error resulting from the fact that provider effects are estimated on a finite sample, we find a variance of 0.045 for physicians and 0.048 for NPs (see Appendix A.4.2 for details of the split-sample approach). These estimates suggest large variation in provider effects: A one-standard-deviation costlier physician and NP increases total cost associated with the ED visit by 21 and 22 percent per case, respectively, which are about three times of the average NP effect of 6.7 percent from the 2SLS model in Equations (1) and (2). Related to the wide variation in provider effects within physician and NP professional classes, Appendix Figure A.7 shows that the NP effect varies considerably across EDs (details in Appendix A.5). Accounting for sampling error, the standard deviation of the ED-specific effects for each outcome is similar in magnitude to the average NP

[48] Doyle, Ewer, and Wagner (2010) show differences in resource utilization decisions among physician trainees, potentially driven by human capital. Gowrisankaran, Joiner, and Léger (2017) provide evidence of variation in diagnostic and treatment skill, and Silver (2021) examines returns to time spent on patients by ED physicians and variation in the physicians' productivity. Chan, Gentzkow, and Yu (2022) demonstrate important variation in diagnostic skill in the setting of radiology.

effect reported in Section 4.

We then investigate the full distributions of provider effects on total cost per case, applying a non-parametric 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 associated standard errors from the previously described just-identified IV model. 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) (i.e., 6.7 percent). Panel A of Figure 9 displays the deconvolved density of provider effects for NPs and physicians. The figure shows large variation in provider effects within professions. The deconvolved distributions imply that the probability that a randomly drawn NP is costlier than a randomly drawn physician is 62 percent (Appendix Figure A.8).[49] Appendices A.4.3 and A.4.4 describe details of these estimations.

Separately, we characterize the full distributions of annualized provider wages, for NPs and physicians. Since we observe actual wage payments to providers, we do not estimate them in a regression framework, nor do we apply deconvolution to obtain a population prior distribution. Rather, we simply describe the empirical distributions of actual annualized wages.[50] In contrast to the distributions of provider effects on total spending per case, Panel B of Figure 9 shows virtually no overlap between the wage distributions for NPs and physicians. That is, to a much larger degree than for provider effects, wages are extremely predictive of professional class.

Finally, we explore the relationship between wages and our measure of provider productivity (i.e., provider effects on total spending per case). In previous sections, we have demonstrated a large difference in productivity between NPs and physicians, with economic implications possibly larger than the difference in wages between the two professions. Here, we compute the empirical Bayes posterior mean for each provider's effect on total spending per case, using the previously described just-identified IV coefficients for each provider and the deconvolved empirical Bayes prior distributions for NPs and physicians. We also

[49]This statistic is equivalent to the c-statistic of a receiver operating characteristic (ROC) curve in which underlying provider effects from the deconvolved distributions, if observed, were to be used to predict professional class. The probability that a randomly chosen NP is less costly than a randomly chosen physician, i.e., $100-62 = 38$ percent, is robust to accounting for possible differences in treatment effects between the overall population and compliers. When assuming the average treatment effect is as large as that among the highest complexity quartile patients which is twice of the LATE estimate, the probability that NPs are less costly remains large at 28 percent.

[50]For each provider, we access detailed payment records of the full-time equivalents and wages for each pay period between the years 2011 to 2020, inclusive. 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 considers a number of 80 hours to be one full-time-equivalent, and (iii) multiplying this figure by 26 pay periods and 80 hours per pay period.

shrink provider effects linearly to the grand mean using precision weights given by the signal and noise.[51] Appendix Figure A.9 shows the results with binned scatter plots—separately for NPs and physicians—in the space of wages on the y-axis and empirical Bayes posterior means on the x-axis, residualizing both by ED indicators. We find no significant correlation between wages and productivity within each of the professions. Appendix A.4.5 provides further details of the estimation.

8 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 process of selecting and training members of the profession from society at large, and the economic returns to working as a professional. However, very little is known empirically about the impact of qualitatively distinct professions on productivity. This is because, by their very nature, 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 classes of professionals—nurse practitioners (NPs) and physicians. While physicians have occupied a dominant position in society since the turn of the 20th century (Starr 1982), the rising demand for health care in an aging population and the limited supply of physicians have set the stage for the rise of NPs to challenge the monopoly of physicians over the independent provision of medical care. The professions of NPs and physicians differ widely in their incomes, training, selectivity, and social standing.

Yet the coexistence of these two professional classes for providing medical care raises questions about whether professional class matters for productivity. Equal productivity between NPs and physicians would imply that the process to become a physician is unnecessarily selective, that the additional years of training is wasteful from society's perspective, and that the two-fold higher salary of physicians reflects monopoly rents and perhaps institutions based on the mistaken concept that physicians are worth the higher price of their labor. On the other hand, lower NP productivity could suggest additional productivity costs when using NPs to meet rising demand for health care. Differences in productivity may also imply an optimal matching of tasks to workers that accounts for professional class.

Our empirical setting allows us to study the quasi-experimental assignment of ED patients to NPs versus physicians in the Veterans Health Administration (VHA). Beginning in December 2016, the VHA directed its stations to allow full practice authority to NPs. We use the quasi-random arrival of patients at the ED

[51] Such shrinkage estimators are often used (e.g., Chetty, Friedman, and Rockoff 2014; Chandra et al. 2016; Abaluck et al. 2021) and equivalent to empirical Bayes posterior means when assuming the prior distribution is normal.

between times that may differ in the availability of NPs on shift, which drives the probability of being treated by an NP versus a physician. Compared to physicians, NPs incur greater resource costs to treat patients but achieve worse patient outcomes.

We also shed light on behavioral mechanisms and responses that connect productivity to human capital. We show that the performance gap between NPs and physicians narrows as NPs gain more experience, suggesting that differences in training could explain some of the gap. We demonstrate clinical decision-making in response to lower human capital: NPs are more likely to gather external information, suggesting compensation for less information that they can perceive on their own. NPs are also less likely to prescribe drugs with potentially high errors of commission (i.e., opioids), while they are more likely to prescribe drugs with potentially high errors of omission (i.e., antibiotics). Finally, the productivity gap between NPs and physicians is higher for more complex and more severe patients. We show descriptive evidence of patient allocation that responds to NP comparative and absolute disadvantage: On average, NPs receive healthier patients and take on a smaller share of the caseload when the ED is less constrained to meet demand.

Even under the most conservative assumptions, the resource costs implied by the lower productivity we find outweigh any salary savings from hiring NPs, despite NP wages that are half as much as physician wages. This reflects the outsize importance of productivity in modern health care, in which the utilization of considerable resources rests on the judgment of workers. However, we also find productivity variation within each of the professions that are even larger than the difference between professions. The distributions of NP and physician productivity—at least according to our measures of provider effects on total spending per case—substantially overlap, in contrast to negligible overlap in the wages of NPs and physicians. Furthermore, there is no significant correlation between wages and productivity within professions, suggesting that employers at best use professional class as a coarse signal of productivity for setting wages.

Considered together, our findings paint a nuanced picture of the role of professions in determining the productivity and wages of workers. While the intensive processes of professional selection and training may imply important productivity differences between professional classes, perhaps justifying sizable differences in wages, it appears likely that professional institutions are more effective at compressing wages rather than at standardizing productivity within professions. These relationships may derive from frictions in observing productivity in the labor marketplace (Acemoglu and Pischke 1998), where professions may provide an important function of certifying membership. Professional membership may sometimes be exclusive, as in the case of physicians, but it may nonetheless be a highly imperfect proxy for productivity.

References

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

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

Abbott, Andrew. 2014. *The System of Professions: An Essay on the Division of Expert Labor.* Chicago: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, ed. David Card and Orley Ashenfelter, Chapter 12, 1043–1171. Amsterdam:Elsevier.

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

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

Alexander, Diane, and Molly Schnell. 2019. "Just What the Nurse Practitioner Ordered: Independent Prescriptive Authority and Population Mental Health." *Journal of Health Economics*, 66: 145–162.

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): 151–184.

American Academy of Medical Colleges. 2020. "The Road to Becoming a Doctor." Accessed February 25, 2022. `https://www.aamc.org/system/files/2020-11/aamc-road-to-becoming-doctor-2020.pdf`.

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

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

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

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): 4337–4383.

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

Beck, Andrew H. 2004. "The Flexner Report and the Standardization of American Medical Education." *JAMA*, 291(17): 2139–2140.

Benkard, Lanier. 2000. "Learning and Forgetting: The Dynamics of Aircraft Production." *American Economic Review*, 90(4): 1034–1054.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, 119(1): 249–275.

Bhuller, Manudeep, Gordon B. Dahl, Katrine V. Løken, and Magne Mogstad. 2020. "Incarceration, Recidivism, and Employment." *Journal of Political Economy*, 128(4): 1269–1324.

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

Bureau of Labor Statistics. 2021*a*. "Occupational Employment and Wages, May 2020 (Nurse Practitioners)." Accessed November 10, 2021. https://www.bls.gov/oes/current/oes291171.htm.

Bureau of Labor Statistics. 2021*b*. "Occupational Employment and Wages, May 2020 (Physicians, All Other; and Ophthalmologists, Except Pediatric)." Accessed November 10, 2021. https://www.bls.gov/oes/current/oes291228.htm.

Butler, Christopher C., David Gillespie, Patrick White, Janine Bates, Rachel Lowe, Emma Thomas-Jones, Mandy Wootton, Kerenza Hood, Rhiannon Phillips, Hasse Melbye, et al. 2019. "C-Reactive Protein Testing to Guide Antibiotic Prescribing for COPD Exacerbations." *New England Journal of Medicine*, 381(2): 111–120.

Cameron, A. Colin, and Douglas L. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources*, 50(2): 317–372.

Chan, David C. 2016. "Teamwork and Moral Hazard: Evidence from the Emergency Department." *Journal of Political Economy*, 124(3): 734–770.

Chan, David C. 2021. "Influence and Information in Team Decisions: Evidence from Medical Residency." *American Economic Journal: Economic Policy*, 13(1): 106–137.

Chan, David C., Matthew Gentzkow, and Chuan Yu. 2022. "Selection with Variation in Diagnostic Skill: Evidence from Radiologists." *Quarterly Journal of Economics*, 137(2): 729–783.

Chandra, A., and D. O. Staiger. 2007. "Productivity Spillovers in Healthcare: Evidence from the Treatment of Heart Attacks." *Journal of Political Economy*, 115(1): 103–140.

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

Chang, Cindy Y., and Ziad Obermeyer. 2020. "Association of Clinical Characteristics With Variation in Emergency Physician Preferences for Patients." *JAMA Network Open*, 3(1): e1919607.

Chen, Yiqun. 2021. "Team-Specific Human Capital and Team Performance: Evidence from Doctors." *American Economic Review*, 111(12): 3923–3962.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review*, 104(9): 2593–2632.

Currie, Janet M., 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 M., and W. Bentley MacLeod. 2020. "Understanding Doctor Decision Making: The Case of Depression Treatment." *Econometrica*, 88(3): 847–878.

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

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): 866–882.

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, A. J., and S. Nicholson. 2009. "The Formation and Evolution of Physician Treatment Styles: an Application to Cesarean Sections." *Journal of Health Economics*, 28(6): 1126–1140.

Fairman, Julie. 2009. *Making Room in the Clinic: Nurse Practitioners and the Evolution of Modern Health Care.* New Brunswick: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, Jeffrey S. Gerber, David Y. Hyun, Jeffrey A. Linder, et al. 2016. "Prevalence of Inappropriate Antibiotic Prescriptions Among US Ambulatory Care Visits, 2010-2011." *JAMA*, 315(17): 1864–1873.

Flexner, Abraham. 1910. *Medical Education in the United States and Canada: A Report to the Carnegie Foundation for the Advancement of Teaching.* New York:Carnegie Foundation for the Advancement of Teaching.

Freidson, Eliot. 1974. *Professional Dominance: The Social Structure of Medical Care.* New Brunswick:Transaction Publishers.

Friedman, Sam, Daniel Laurison, and Lindsey Macmillan. 2017. "Social Mobility, the Class Pay Gap and Intergenerational Worklessness: New Insights from The Labour Force Survey." UK Social Mobility Commission.

Gennaioli, Nicola, Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Shleifer. 2013. "Human Capital and Regional Development." *Quarterly Journal of Economics*, 128(1): 105–164.

Gibbons, Robert, and Michael Waldman. 2004. "Task-Specific Human Capital." *American Economic Review*, 94(2): 203–207.

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. 2017. "Physician Practice Style and Healthcare Costs: Evidence from Emergency Departments." NBER Working Paper 24155.

Hallett, Christine E., ed. 2016. *The History of Nursing. Major Themes in Health and Social Welfare*, New York:Routledge.

Health Economics Resource Center. 2021. "HERC Inpatient Average Cost Data." Accessed February 23, 2022. `https://www.herc.research.va.gov/include/page.asp?id=inpatient`.

Huang, David T., Donald M. Yealy, Michael R. Filbin, Aaron M. Brown, Chung-Chou H. Chang, Yohei Doi, Michael W. Donnino, Jonathan Fine, Michael J. Fine, Michelle A. Fischer, et al. 2018. "Procalcitonin-Guided Use of Antibiotics for Lower Respiratory Tract Infection." *New England Journal of Medicine*, 379(3): 236–249.

Imbens, Guido W., and Donald B. Rubin. 1997. "Estimating Outcome Distributions for Compliers in Instrumental Variables Models." *Review of Economic Studies*, 64(4): 555–574.

Institute of Medicine. 2006. "Hospital-Based Emergency Care: At the Breaking Point."

Kahn, Marc J., and Ernest J. Sneed. 2015. "Promoting the Affordability of Medical Education to Groups Underrepresented in the Profession: The Other Side of the Equation." *AMA Journal of Ethics*, 17(2): 172–175.

Kinnersley, Paul, Elizabeth Anderson, Kate Parry, John Clement, Luke Archard, Pat Turton, Andrew Stainthorpe, Aileen Fraser, Chris C. Butler, and Chris Rogers. 2000. "Randomised Controlled Trial of Nurse Practitioner versus General Practitioner Care for Patients Requesting 'Same Day' Consultations in Primary Care." *BMJ*, 320(7241): 1043–1048.

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): 261–291.

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–S202.

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): 140–149.

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.* Berkeley: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.

Lazuka, Volha. 2018. "The Long-Term Health Benefits of Receiving Treatment from Qualified Midwives at Birth." *Journal of Development Economics*, 133: 415–433.

Maggs, Christopher. 2016. "Profit and loss and the hospital nurse." In *The History of Nursing*. Vol. 3 of *Major Themes in Health and Social Welfare*, ed. Christine Hallett, 83–93. New York:Routledge.

Maier, Claudia B., and Linda H. Aiken. 2016. "Task shifting from physicians to nurses in primary care in 39 countries: a cross-country comparative study." *European Journal of Public Health*, 26(6): 927–934.

Maier, Claudia B., Hilary Barnes, Linda H. Aiken, and Reinhard Busse. 2016. "Descriptive, cross-country analysis of the nurse practitioner workforce in six countries: size, growth, physician substitution potential." *BMJ Open*, 6(9): e011901.

Maier, Claudia B., Ronald Batenburg, Stephen Birch, Britta Zander, Robert Elliott, and Reinhard Busse. 2018. "Health workforce planning: which countries include nurse practitioners and physician assistants and to what effect?" *Health Policy*, 122(10): 1085–1092.

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

McMichael, Benjamin J., and Sara Markowitz. forthcoming. "Toward a Uniform Classification of Nurse Practitioner Scope of Practice Laws." *Medical Care Research and Review*.

Moretti, Enrico. 2004. "Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions." *American Economic Review*, 94(3): 656–690.

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): 1–8.

Neuman, Mark D., Brian T. Bateman, and Hannah Wunsch. 2019. "Inappropriate Opioid Prescription After Surgery." *The Lancet*, 393(10180): 1547–1557.

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

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.

Searcey, Dionne, Eduardo Porter, and Robert Gebeloff. 2015. "Health Care Opens Stable Career Path, Taken Mainly by Women." *The New York Times*.

Shapiro, Carl. 1986. "Investment, Moral Hazard, and Occupational Licensing." *Review of Economic Studies*, 53(5): 843–862.

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

Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick. 2019. "Capitalists in the Twenty-First Century." *Quarterly Journal of Economics*, 134(4): 1675–1745.

Spence, Michael. 1973. "Job Market Signaling." *Quarterly Journal of Economics*, 87(3): 355–374.

Staiger, Douglas O., Joanne Spetz, and Ciaran S. Phibbs. 2010. "Is There Monopsony in the Labor Market? Evidence from a Natural Experiment." *Journal of Labor Economics*, 28(2): 211–236. Publisher: The University of Chicago Press.

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.* New York:Basic Books.

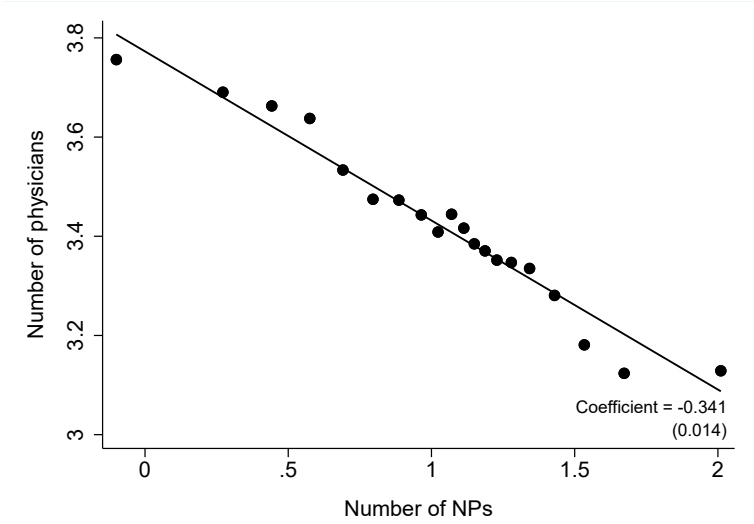
Sullivan, Louis W., and Ilana Suez Mittman. 2010. "The state of diversity in the health professions a century after Flexner." *Academic Medicine: Journal of the Association of American Medical Colleges*, 85(2): 246–253.

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

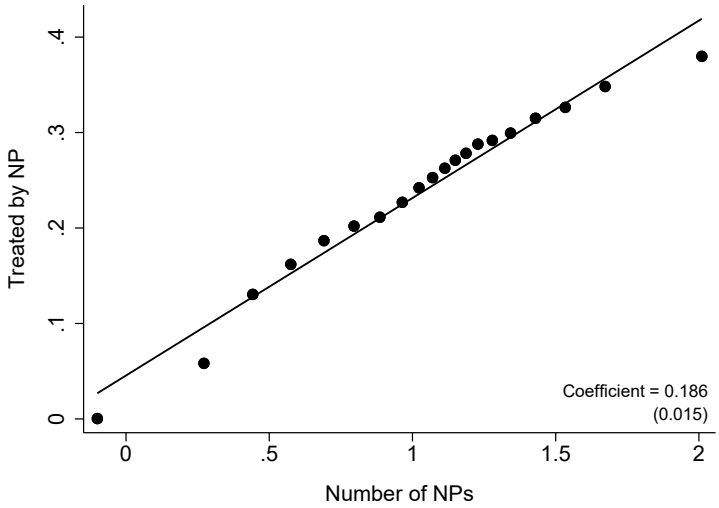
Wu, Fred, and Michael A. Darracq. 2021. "Comparing Physician Assistant and Nurse Practitioner Practice in US Emergency Departments, 2010–2017." *Western Journal of Emergency Medicine*, 22(5): 1150–1155.

Figure 1: First Stage

A. Number of Physicians

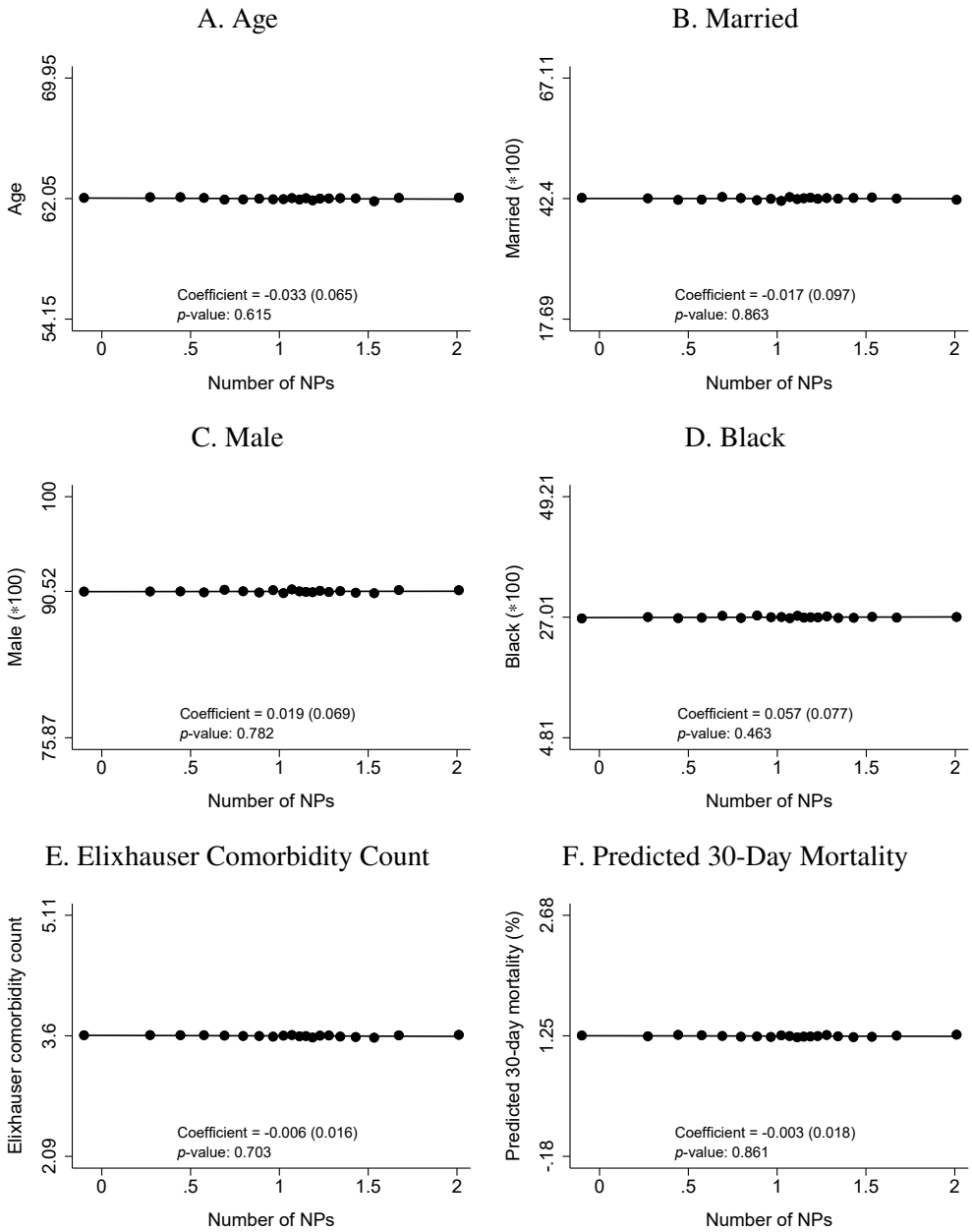


B. Treated by NP



Notes: This figure represents a graphical illustration of the first-stage estimation. Panel A shows a binned scatter plot of the number of physicians on duty versus the number of NPs on duty. Panel B shows a binned scatter plot of whether the case is treated by an NP versus the number of NPs on duty. To construct these binned scatter plots, we first residualize both the y-axis and x-axis variable with respect to the baseline control vector (i.e., indicators for ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day) 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 variable (conditional on the baseline control vector), with standard errors clustered by provider reported in parentheses.

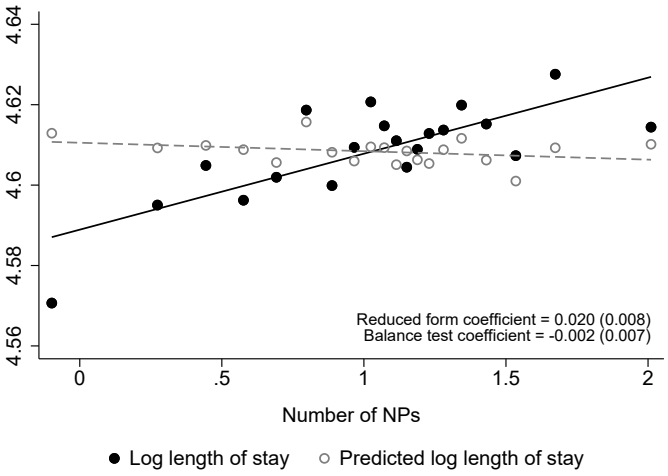
Figure 2: Balance in Patient Characteristics



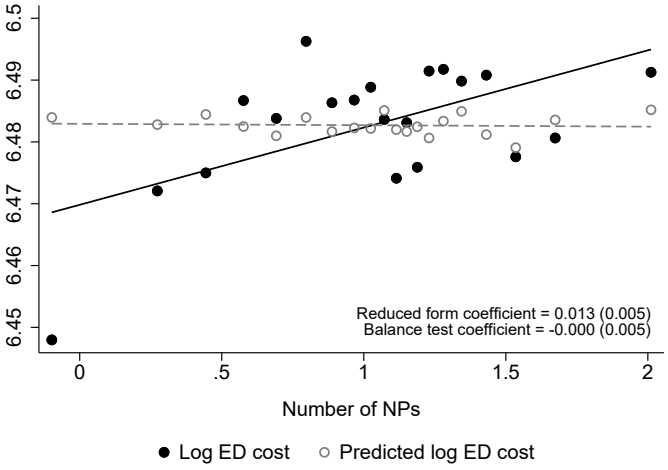
Notes: This figure shows balance in patient characteristics across the number of NPs on duty, conditional on the baseline controls (i.e., indicators for ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day). To construct these binned scatter plots, we first residualize both the *y*-axis and *x*-axis variable 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 number 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 variable (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 $\mathbf{X}_i$ included in Equations (1) and (2), including demographics, comorbidities, prior health care use, vital signs, and three-digit diagnosis indicators.

Figure 3: Reduced-Form and Balance

A. Log Length of Stay



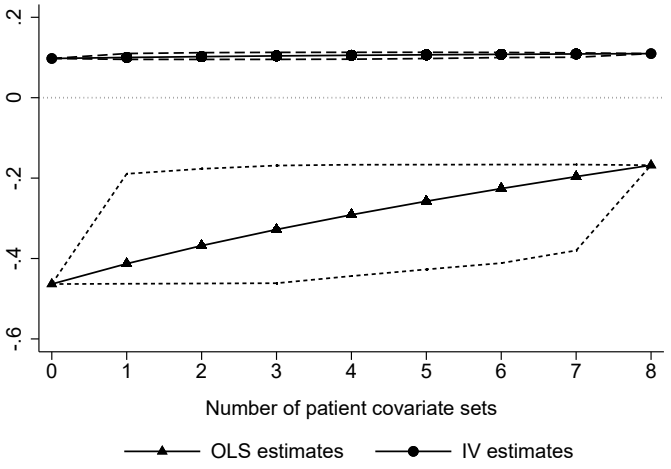
B. Log ED Cost



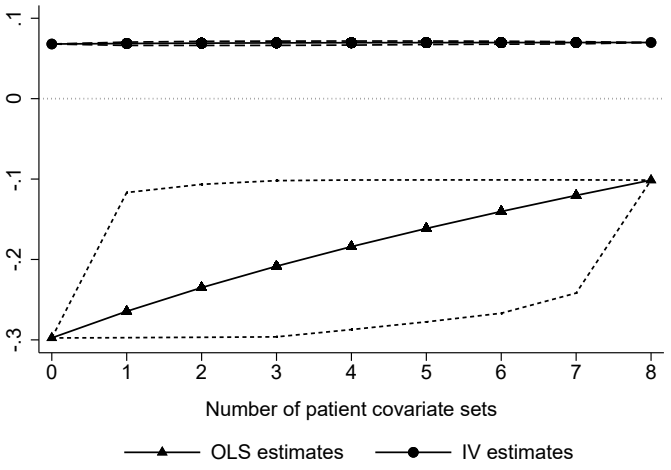
Notes: This figure shows binned scatter plots of patient actual and predicted outcomes on the *y*-axis versus the number of NPs on duty on the *x*-axis, controlling for the baseline control vector (i.e., indicators for ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day). Panel A reports results for log length of stay; Panel B reports results for log cost of the ED visit. The solid circles and lines represent patient actual outcomes. The hollow circles and dashed lines represent patient 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 diagnosis indicators. The reduced-form coefficients are estimated using Equation (2), with patient actual outcomes as the dependent variable; the balance-test coefficients are estimated by regressing patient predicted outcomes on the number of NPs on duty, conditional on the baseline control vector. Standard errors clustered by provider are reported in parentheses.

Figure 4: Stability of OLS and IV Estimates

A. Log Length of Stay

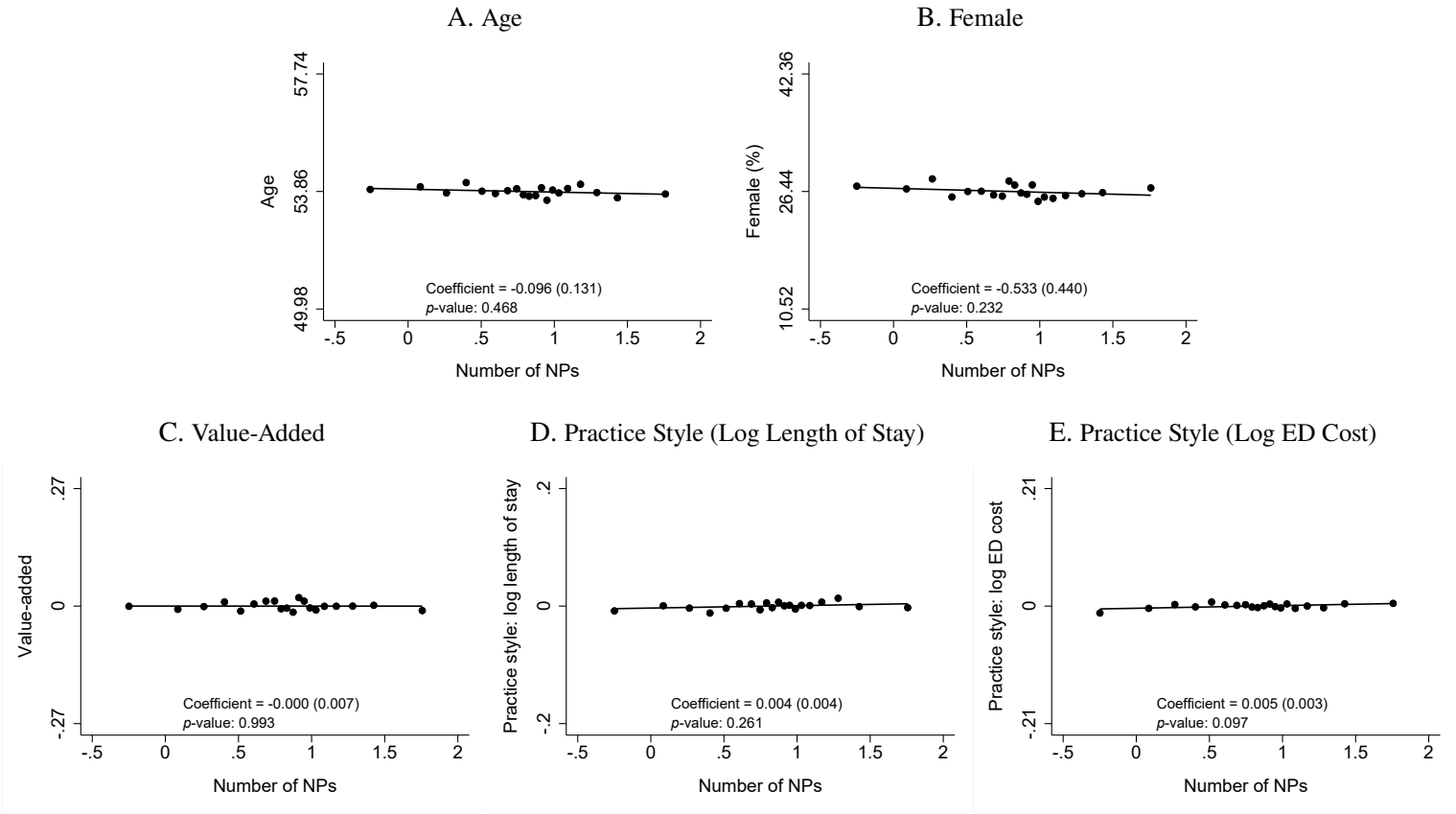


B. Log ED Cost



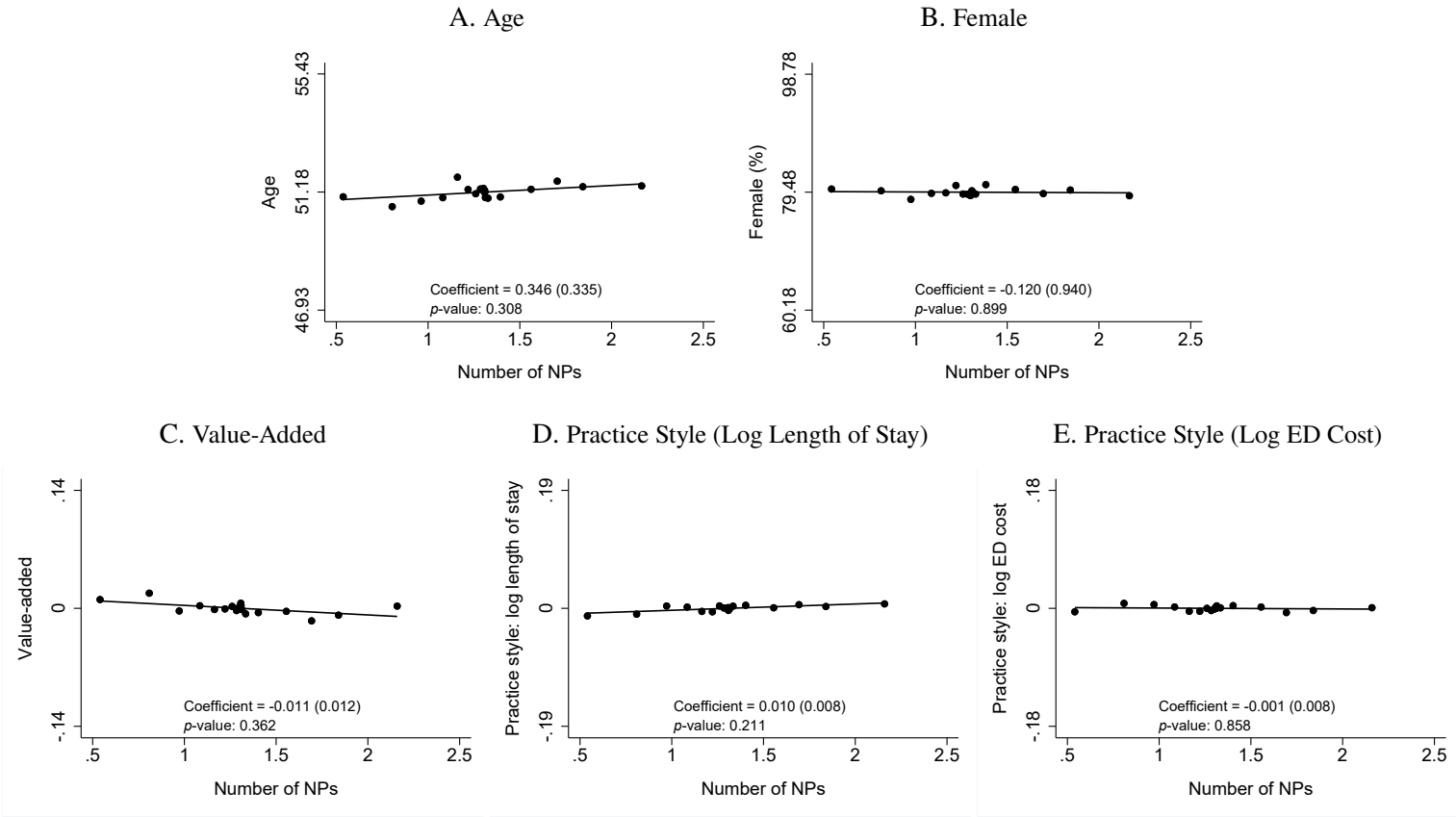
Notes: This figure shows the robustness of our OLS and IV estimates to the inclusion of different sets of patient controls. We divide patient observable characteristics into eight subsets: (i) five-year age-bin indicators; (ii) marital status; (iii) gender; (iv) race indicators; (v) indicators for 31 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 mean 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. Panel A reports results for log length of stay. Panel B reports results for log cost of the ED visit.

Figure 5: Balance in Physician Characteristics



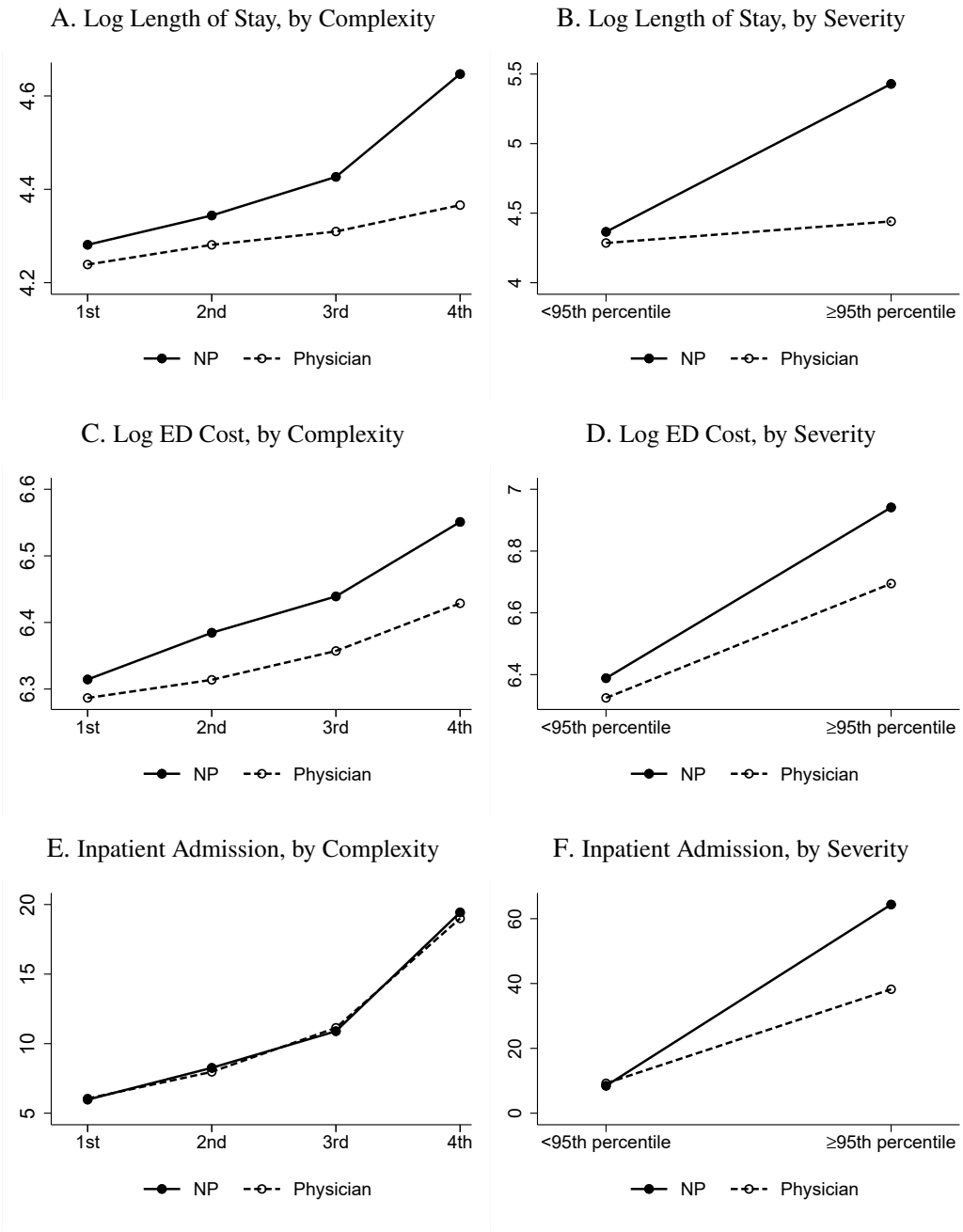
Notes: These panels are graphical representations of the balance-test regression at the ED-day level of physician average characteristics (weighted by the number of cases treated by each physician) on the number of NPs on duty, conditional on ED-by-year, ED-by-month, and ED-by-day-of-the-week indicators. Coefficients from the regressions are reported in each panel, along with standard errors (shown in parentheses) and p-values. To construct the binned scatter plots, 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 indicators for ED-by-year, ED-by-month, and ED-by-day-of-the-week, 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 number 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 Appendix A.3.

Figure 6: 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 scatter plots, 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 indicators for ED-by-year, ED-by-month, and ED-by-day-of-the-week, 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 number 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 Appendix A.3.

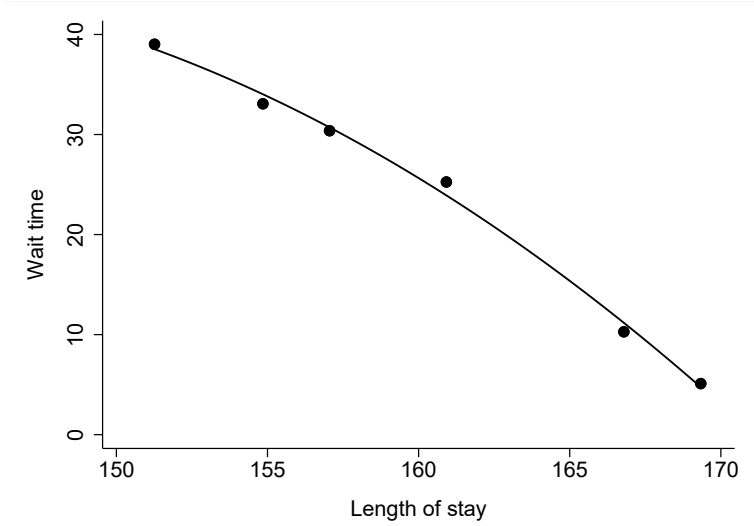
Figure 7: Heterogeneous Effects by Case Complexity and Severity



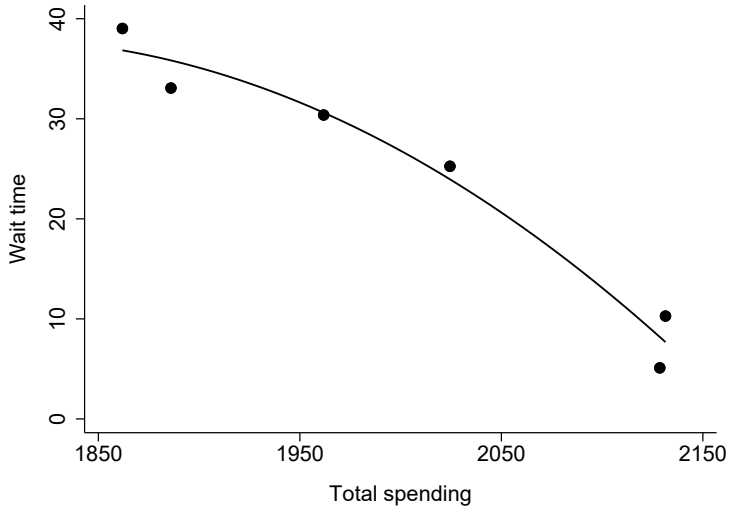
Notes: This figure shows heterogeneous effects of NPs by case complexity and severity. Panels A, C, and E divide cases into quartiles by their total number of Elixhauser comorbidities, with higher quartiles indicating more complex cases. Panels B, D, and F divide cases by whether condition severity measured by 30-day mortality of cases with the same three-digit ICD-10 primary diagnosis is equal to or above the 95th percentile of the sample. 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 $y_i \times (\text{NP}_i - 1)$, i.e., 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.

Figure 8: Trade-Off: Wait Time versus Length of Stay and Total Spending

A. Wait Time versus Length of Stay



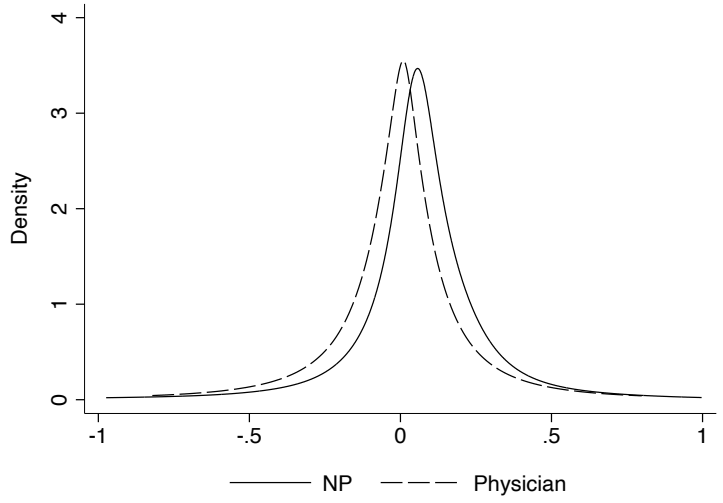
B. Wait Time versus Total Spending



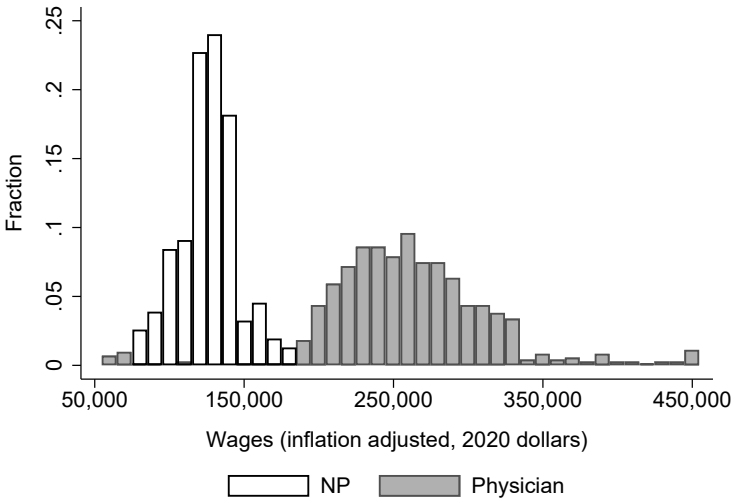
Notes: This figure shows changes in patient average wait time and other outcomes with incremental numbers of NPs on duty, conditional on the number of physicians on duty, the number of cases arriving, and the controls included in the main specification (see details in Equation (5)). Panel A presents the trade-off between wait time and length of stay. Panel B presents the trade-off between wait time and total spending associated with the ED visit, which is the sum of the cost of care at the ED and the cost due to hospital admission in the ED visit and preventable hospitalizations in the 30 days after the ED visit. The solid lines show the quadratic fit estimated on the plotted points.

Figure 9: Distribution of Provider Effects on Medical Spending and Provider Wages

A. Provider Effects on Medical Spending



B. Wages



Notes: Panel A reports the deconvolved distributions of provider effects on log total spending associated with the ED visit, i.e., the sum of the cost of care at the ED, the cost due to hospital admission in the ED visit, and the cost due to preventable hospitalizations in the 30 days after the ED visit. See Appendix A.4.3 for details of the deconvolution estimator. The solid and dashed lines show the deconvolved distributions of NPs and physicians, respectively. Panel B plots histograms of provider wages observed in the VHA data (inflation adjusted to year 2020). The white and gray bins show NPs' and physicians' wages, respectively. Wages are winsorized at the value of $450,000.

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.

Table 2: Effect of NPs on Length of Stay and ED Cost

	Log length of stay			Log ED cost		
	OLS	Reduced form	IV	OLS	Reduced form	IV
	(1)	(2)	(3)	(4)	(5)	(6)
NP assignment	-0.168		0.110	-0.101		0.070
	(0.034)		(0.045)	(0.023)		(0.030)
Number of NPs		0.020			0.013	
		(0.008)			(0.005)	
Full controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	4.608	4.608	4.608	6.483	6.483	6.483
S.D. dep. var.	1.161	1.161	1.161	0.878	0.878	0.878
Observations	1,110,798	1,110,798	1,110,798	1,108,961	1,108,961	1,108,961

Notes: This table shows OLS, reduced-form, and IV estimates of the effect of NPs on patient log length of stay and log cost of the ED visit. Columns 1 and 4 report the OLS estimates; Columns 2 and 5 report the reduced-form estimates; Columns 3 and 6 report the IV estimates. Sample sizes 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 indicators for ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day, 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.

Table 3: Effect of NPs on Additional Outcomes

	Dependent variable		
	Admission	30-day mortality	30-day prevent. hosp.
	(1)	(2)	(3)
Reduced form	0.019	-0.021	0.047
	(0.108)	(0.021)	(0.021)
IV estimate	0.103	-0.116	0.252
	(0.585)	(0.115)	(0.120)
Full controls	Yes	Yes	Yes
Mean dep. var.	16.625	1.247	1.234
S.D. dep. var.	37.230	11.099	11.041
Observations	1,118,836	1,118,836	1,118,836

Notes: This table shows reduced-form and IV estimates of the effect of NPs on various outcomes. Inpatient admission is an indicator for whether the patient is admitted to the hospital in the ED visit; 30-day mortality indicates whether the patient dies within 30 days of the ED visit; 30-day preventable hospitalization is defined as having any preventable hospitalization in the 30 days after the ED visit. All estimations include the full set of controls described in the notes to Table 2. Standard errors clustered by provider are shown in parentheses.

Table 4: Heterogeneous Effects by Provider Experience

	Dependent variable							
	Log length of stay	Log ED cost	Admission	30-day mortality	30-day prevent. hosp.	Consult	CT	X-ray
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Provider specific experience								
NP assignment	0.101	0.072	0.092	-0.115	0.255	0.025	0.011	0.019
	(0.044)	(0.030)	(0.579)	(0.116)	(0.121)	(0.009)	(0.007)	(0.009)
NP assignment × experience	-0.058	-0.042	-0.504	-0.001	0.016	-0.014	-0.010	-0.003
	(0.025)	(0.019)	(0.331)	(0.041)	(0.030)	(0.008)	(0.003)	(0.008)
Experience	-0.001	0.006	0.238	-0.001	-0.016	-0.009	-0.003	0.005
	(0.006)	(0.010)	(0.314)	(0.018)	(0.014)	(0.006)	(0.002)	(0.002)
Panel B: Provider general experience								
NP assignment	0.086	0.062	0.089	-0.100	0.255	0.022	0.010	0.018
	(0.043)	(0.029)	(0.608)	(0.116)	(0.121)	(0.009)	(0.007)	(0.009)
NP assignment × experience	-0.103	-0.035	0.340	0.088	0.043	-0.020	-0.008	-0.012
	(0.056)	(0.033)	(1.048)	(0.093)	(0.068)	(0.011)	(0.010)	(0.010)
Experience	-0.036	-0.013	-0.719	-0.012	-0.048	-0.006	-0.006	-0.001
	(0.015)	(0.011)	(0.221)	(0.023)	(0.025)	(0.007)	(0.003)	(0.004)
Full controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	4.608	6.483	16.625	1.247	1.234	0.226	0.145	0.291
S.D. dep. var.	1.161	0.878	37.230	11.099	11.041	0.418	0.352	0.454
Observations	1,110,798	1,108,961	1,118,836	1,118,836	1,118,836	1,118,836	1,118,836	1,118,836

Notes: 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 since 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 (despite diagnoses) the provider has treated since the start of the study period to the day before the current case's visit. For ease of interpretation, both specific and general experience are standardized to have a mean of zero and a standard deviation of one for NPs and physicians separately. The outcome variables in Columns 1-8 are, respectively, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, 30-day preventable hospitalization, whether the patient receives formal consults in the ED visit, whether the patient receives CT scans in the ED visit, and whether the patient receives X-rays in the ED visit. All estimations include the full set of controls described in the notes to Table 2. Standard errors clustered by provider are shown in parentheses.

Table 5: Clinical Decision-Making

	Dependent variable				
	Consult	CT	X-ray	Opioid	Antibiotic
	(1)	(2)	(3)	(4)	(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
S.D. 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 in the ED visit, respectively, formal consults, CT scans, X-rays, opioid prescriptions, and antibiotic prescriptions. Since antibiotics generally only apply to patients with infections, Column 5 restricts the sample to patients with respiratory or genitourinary system infections, which are 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.

Online Appendix

A.1 Diagnosis Coding: NPs versus Physicians

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

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

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

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

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

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

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

A.2 Characterizing Compliers and Never-Takers

This appendix describes estimation of characteristics of compliers and never-takers. Following the approach developed by Abadie (2003), we characterize compliers by δ estimated through the 2SLS model specified in Equations (A.1) and (A.2), replacing the outcome variable y_i with $x_i \times \text{NP}_i$, i.e., the interaction between each patient characteristic x_i and the indicator for being treated by an NP. Results are discussed in Section 4.3 and shown in Columns 2-3 of Appendix Table A.3.

To estimate characteristics of never-takers, we follow a method by Dahl, Kostøl, and Mogstad (2014). We first collapse the data to the ED-day level. We then residualize the share of cases treated by NPs by indicators for ED-by-year, ED-by-month, and ED-by-day-of-the-week. We define never-takers as cases treated by physicians in ED-day cells with the residual share of cases treated by NPs at least as high as the 90th percentile of ED-days with at least one case treated by NPs. There are no always-takers in our setting since patients cannot be assigned to NPs on days without any NPs on duty.

Columns 4 of Appendix Table A.3 report the average characteristics of never-takers. For each characteristic, we compute the mean of never-takers as well as the ratio between the mean and the overall sample mean. We estimate standard errors for the means by bootstrap, blocking observations by provider with 500 replications. In line with the notion that NPs treat healthier cases than physicians do, Appendix Table A.3 shows that never-takers are the riskiest, followed by the overall sample, and finally, compliers. For example, the total number of Elixhauser comorbidities among never-takers, the overall sample, and compliers are, respectively, 4.0, 3.6, and 3.3; the average predicted 30-day mortality among these three types of cases are, respectively, 1.7, 1.2, and 0.9 percentage points.

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

This appendix describes our construction of measures of provider value-added and practice styles, used to examine the exclusion restriction in Section 4.4. We consider physician value-added as a measure of risk-adjusted mortality outcomes and form these measures using leave-out data. Specifically, for physician p on day d, we measure

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

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

Still, since cases are not experimentally assigned among physicians, $A_{p,d}$ may reflect both a physician's effect on patient outcomes and systematic patient-physician sorting under imperfect risk adjustment. As one way to assess the degree of such potential biases, we investigate the robustness of physician value-added

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

estimates to patient predicted mortality constructed on the basis of different risk adjusters, analogous to the test of student sorting biases in the teacher value-added literature (e.g., Chetty, Friedman, and Rockoff 2014). If patient sorting is important, the estimated physician value-added will change meaningfully with the addition of risk adjusters. Otherwise, our estimates should remain stable. Appendix Figure A.11 shows that physician value-added estimates are stable regardless of patient risk adjusters. We compare physician value-added measures constructed using (i) the most parsimonious set of risk adjusters that includes only age-bin and three-digit primary diagnosis indicators, (ii) the less parsimonious set that adds non-age demographics (gender, race, and marital status), and (iii) the set that further adds dummies for 31 Elixhauser comorbidities, with the baseline physician value-added constructed using the full set of patient covariates (i.e., demographics, Elixhauser comorbidities, prior health care use, vital signs, and three-digit diagnosis indicators). The correlations between measures (i)–(iii) and the baseline measure are all above 0.99. Note that these risk adjusters are important predictors of patient 30-day mortality: They alone explain 7 percent of the variation in 30-day mortality, with an F-statistic of 88 for joint significance.

We consider physician practice styles as measures of physician-chosen inputs to care. Specifically, we define practice style measures by Equation (A.3), but instead set $\tilde{Y}_i$ as the difference between patient actual and predicted log length of stay or log cost of the ED visit. As with value-added, we show the robustness of practice style estimates to different patient risk adjusters in Appendix Figure A.11.

We construct similar measures of value-added and practice style for NPs. As with physicians, we show the robustness of these estimates to different patient risk adjusters. Appendix Figure A.11 shows that NP valued-added and practice-style estimates are highly stable among those constructed using (i) the most parsimonious set of risk adjusters that includes only age-bin and three-digit primary diagnosis indicators, (ii) the less parsimonious set that adds non-age demographics (gender, race, and marital status), (iii) the set that additionally includes dummies for 31 Elixhauser comorbidities, and (iv) the full set that further adds detailed controls for prior health care use and vital signs upon arrival at the ED.

A.4 Distribution of Provider Effects on Total Spending

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

A.4.1 Estimating Provider Effects

We generate a measure of total spending associated with the ED visit, as the sum of the three main components of costs that we find significant NP effects: ED costs, hospital admission, and 30-day preventable hospitalizations (we multiply the latter two components by the average cost of a hospital stay, $19,220).

We then estimate provider effects on total spending associated with the ED visit. To mitigate the effect

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

$$y_i = \sum_j \theta_j \mathbf{1}_{\{j(i)=j\}} + \mathbf{T}_i\eta + \mathbf{X}_i\beta + \varepsilon_i, \tag{A.4}$$

$$\mathbf{1}_{\{j(i)=j\}} = \sum_j \lambda_j \mathbf{1}_{\{j\in\mathcal{I}_i\}} + \mathbf{T}_i\zeta + \mathbf{X}_i\gamma + v_i. \tag{A.5}$$

$\mathbf{1}_{\{j(i)=j\}}$ is an indicator for whether case i is treated by provider j, and $\mathcal{I}_i$ is the set of providers on duty in the ED-day cell of case i's visit. The coefficients of interest are θ_j, representing provider effects. Since θ_j is only identified relative to one another for providers within the same ED, we make the natural normalization that the case-weighted mean of θ_j is 0 within each ED, using linear constraints in the 2SLS estimation to yield valid standard errors.[3]

The F-statistics for the joint significance of on-duty provider indicators in the first-stage regressions, i.e., Equation (A.5), have shares of 0.99 and 0.68 above 10 and 100, respectively, suggesting that provider availability is strongly predictive of the treating provider.[4] Appendix Figure A.12 shows that patient characteristics are well balanced across the average characteristics (age, gender, and practice style) of on-duty providers, conditional on the baseline controls, i.e., ED-by-time-category indicators.[5] In addition, the F-statistics for the joint significance of on-duty provider indicators from regressions of patient predicted log length of stay and predicted log cost of the ED visit on on-duty provider indicators conditioning on ED-by-time-category indicators, are 2.4 and 2.1, respectively. These are much smaller than the corresponding F-statistics using the actual log length of stay and log cost of the ED visit as the outcome—which are 10.3 and 9.9, respectively. These results make plausible the assumption that the set of on-duty providers is conditionally independent of the set of patients arriving, supporting the validity of our instruments.

A.4.2 Estimating Variance of Provider Effects

We estimate the variance of provider effects, within each professional class, on log total spending associated with the ED visit. The estimated provider effects $\hat{\theta}_j$ from Appendix A.4.1 yields a case-weighted variance of 0.054 for NPs, and 0.064 for physicians (see Appendix Table A.17).[6] However, these estimates are upward

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

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

[5] We compute case-weighted average characteristics of on-duty providers, with the index case left out. For practice style examined in this balance test, to deal with the concern on patient sorting between NPs and physicians, we use provider effects on patient log length of stay and log cost of the visit estimated by the 2SLS model in Equations (A.4) and (A.5).

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

biased, due to sampling error resulting from the fact that provider effects are estimated on a finite sample. To account for such biases, we leverage a split-sample approach, resembling that employed in earlier studies (e.g., Chetty, Friedman, and Rockoff 2014; Silver 2021). Specifically, we randomly split a provider's patients within each day to two approximately equal-sized partitions. We then estimate the 2SLS model in Equations (A.4) and (A.5) using each partition separately, yielding two fixed effect estimates for each provider $\hat{\theta}_{j,a}$ and $\hat{\theta}_{j,b}$. Suppressing the j subscript for simplicity, we have

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

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

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

We perform this calculation for NPs and physicians separately.

Appendix Table A.17 reports the case-weighted variance of provider effects from the split-sample approach. The variance for physicians is estimated to be 0.045, which is about 70 percent of the calculated variance without accounting for the bias due to sampling error. The variance from the split-sample approach suggests that on average, a one-standard-deviation costlier physician increases medical spending associated with the ED visit by 21 percent per case. For NPs, the split-sample variance estimate is 0.048, suggesting that a one-standard-deviation costlier NP raises spending by 22 percent per case.

A.4.3 The Population Distribution of Provider Effects

We now estimate the distribution of provider effects by applying a non-parametric empirical Bayes deconvolution approach adapted by Kline, Rose, and Walters (2022) from Efron (2016). This approach extracts a flexible estimate of the distribution of population provider effects using provider effects $\hat{\theta}_j$ and their standard errors s_j estimated in Equations (A.4) and (A.5). Assuming provider z-scores $z_j = \hat{\theta}_j / s_j$ are distributed as

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

where $c_j = \theta_j / s_j$ (i.e., the population analogue of z_j), the procedure first applies the Efron (2016) deconvolution procedure to yield a distribution of provider z-scores $\hat{G}_c$ with density function $\hat{g}_c(\cdot)$. The Efron (2016) procedure estimates $\hat{G}_c$ by maximum likelihood of parameters that represent coefficients on a set of splines, with a regularization parameter to tamp down excursions from a flat prior.

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

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

Following Kline, Rose, and Walters (2022), we assess the independence of z_j and s_j by reporting regressions of z_j on s_j. To account for possible correlated estimation errors in z_j and s_j, we also present split-sample versions of these regressions that randomly split cases for each provider into two approximately equal-sized partitions and regress z-scores from one partition on standard errors from the other partition. The results are reported in Appendix Table A.18, which show no significant relationship between z_j and s_j, suggesting that independence between z-scores and standard errors is plausible.

We apply the empirical Bayes deconvolution estimator to NPs and physicians separately.[7] As in Kline, Rose, and Walters (2022), we calibrate the regularization parameter in the maximum likelihood estimation so that the variance of the deconvolved distribution of provider effects matches the corresponding split-sample variance estimates reported in Appendix Table A.17. We demean both the physician and NP distributions to have a mean of zero, and then shift the distribution of NPs to the right by 0.067, where 0.067 is the 2SLS estimate of the NP effect on the log total spending associated with the ED visit obtained by Equations (1) and (2). Panel A of Figure 9 displays the deconvolved density of provider effects for NPs and physicians.

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

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

where $\hat{F}_{MD}(x)$ and $\hat{F}_{NP}(x)$ are the deconvolved cumulative density functions of physician effects and NP effects, respectively, and $\mathcal{J}_{MD}$ and $\mathcal{J}_{NP}$ are the sets of providers who are physicians and NPs, respectively. We find the probability that a randomly drawn NP is costlier than a randomly drawn physician, in terms of the total spending associated with the ED visit defined above, is 62 percent. Put differently, the probability that a randomly drawn NP is less costly than a randomly drawn physician is as high as 38 percent. This statistic remains large when we adjust the deconvolved productivity distributions to account for possible differences in treatment effects between the overall population and compliers: When assuming the average treatment effect is as large as that among the highest complexity quartile patients which is twice of the LATE estimate (0.136 versus 0.067), the probability that NPs are less costly remains large at 28 percent.

A.4.4 ROC Curve Representation

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

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

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

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

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

A.4.5 Correlation Between Productivity and Wages

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

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

The dependent variable wage_j is the yearly wage of provider j (inflation-adjusted to 2020 dollars). $\mathbf{L}_j$ is a vector of ED indicators since provider effects are only identified relative to one another within EDs. Since provider effects $\hat{\theta}_j$ is estimated with noise, we use empirical Bayes posterior means of each provider effects, $\tilde{\theta}_j$, which we calculate as

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

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

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

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

where φ denotes the standard normal density.

A.5 ED-Specific NP Effects

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

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

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

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

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

The shrinkage factor is given by $w_\ell = \frac{\hat{\pi}^2}{s_\ell^2 + \hat{\pi}^2}$, where $\hat{\pi}^2$ and s_ℓ^2 are, respectively, the variance of the prior distribution of $\hat{\delta}_\ell$ and the variance of the sampling error for each $\hat{\delta}_\ell$. We calculate s_ℓ^2 as the square of the standard error of $\hat{\delta}_\ell$. We calculate $\hat{\pi}^2$ as the difference between the case-weighted variance of $\hat{\delta}_\ell$ and the case-weighted mean of s_ℓ^2. Finally, $\hat{\delta}$ is the overall IV estimate of the NP effect in Equations (1) and (2), which is reported in Section 4.

The gray bins in Appendix Figure A.7 plot the empirical Bayes posterior mean $\hat{\delta}_\ell$ for each ED in our sample.[8] The distribution of posteriors is more compressed than that of the raw estimates of ED-specific effects, reflecting shrinkage due to sampling error in the raw estimates. The results show a fair amount of heterogeneity. Nonetheless, most EDs exhibit positive effects of NPs on raising patient length of stay, cost of the ED visit, and 30-day preventable hospitalization rate.

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

References

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

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

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

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review*, 104(9): 2593–2632.

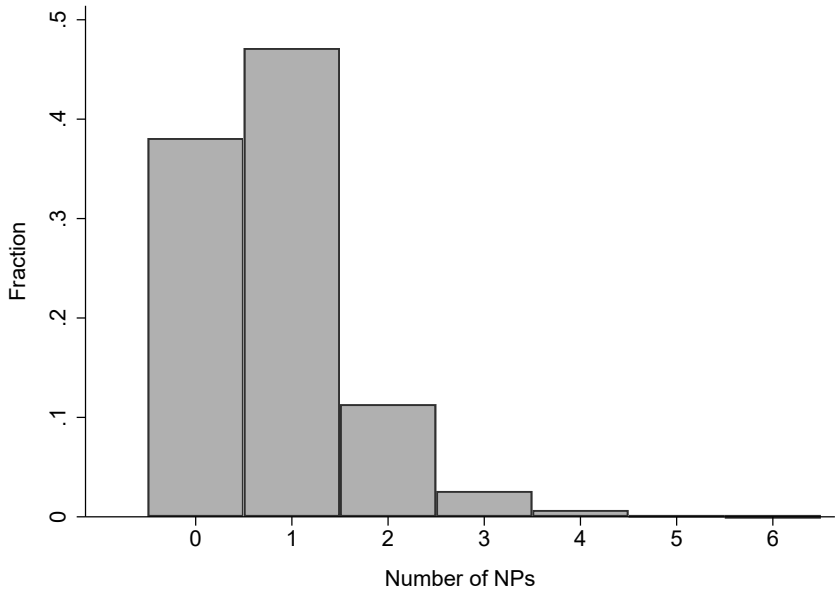
Dahl, Gordon B., Andreas Ravndal Kostøl, and Magne Mogstad. 2014. "Family Welfare Cultures." *Quarterly Journal of Economics*, 129(4): 1711–1752.

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

Kline, Patrick M., Evan K. Rose, and Christopher R. Walters. 2022. "Systemic Discrimination Among Large U.S. Employers." *Quarterly Journal of Economics*, 137(4): 1963–2036.

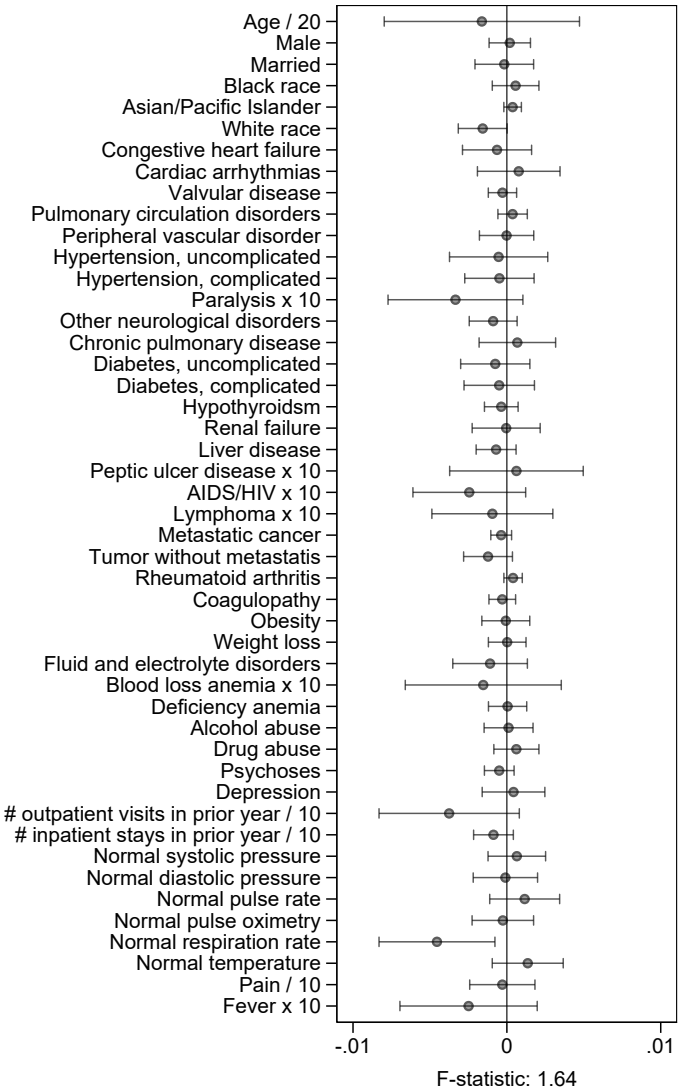
Silver, David. 2021. "Haste or Waste? Peer Pressure and Productivity in the Emergency Department." *Review of Economic Studies*, 88(3): 1385–1417.

Figure A.1: Number of NPs on Duty



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

Figure A.2: Balance in Patient Characteristics

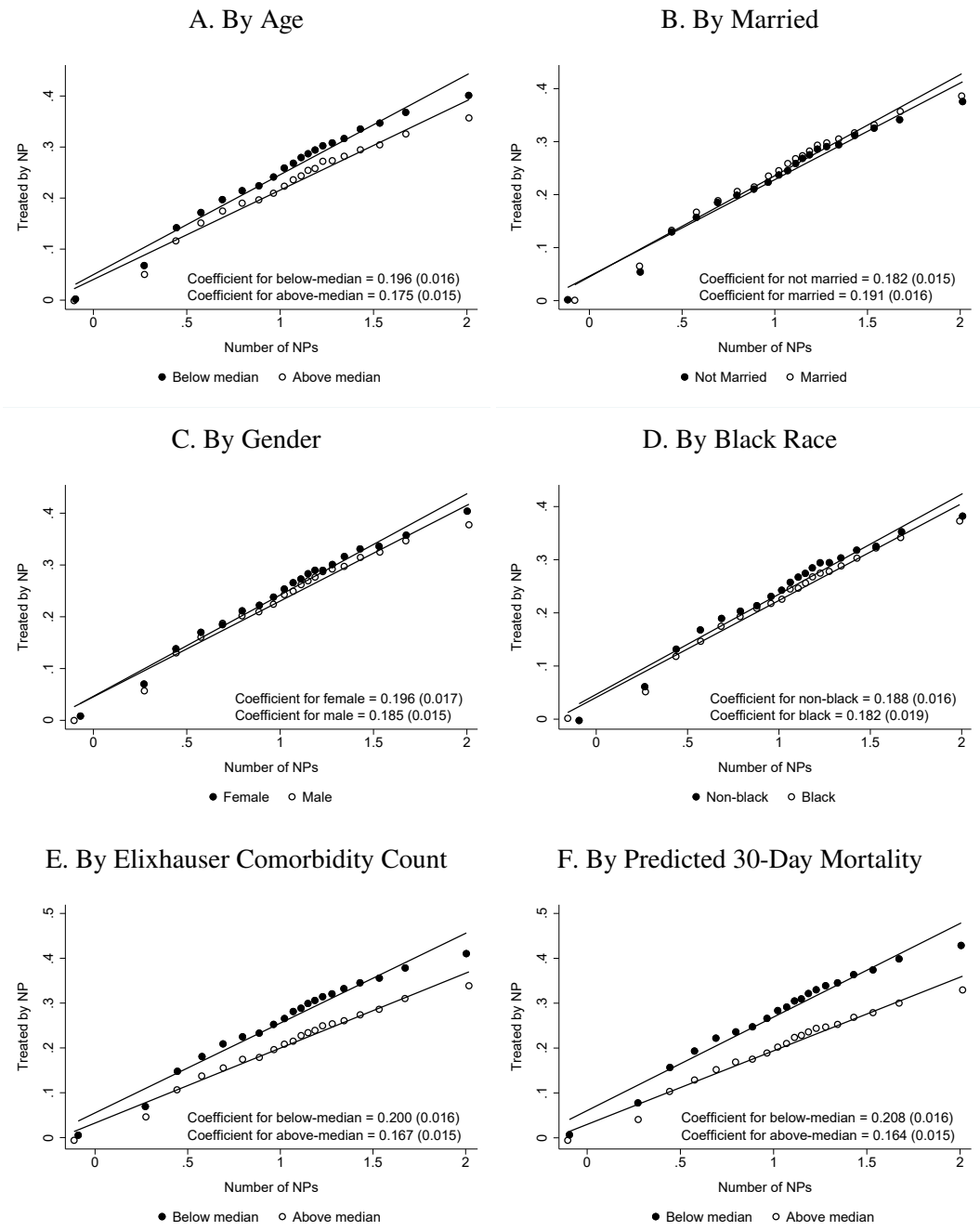


Notes: This figure shows estimated coefficients and 95% confidence intervals from regressions of each patient characteristic listed on the y-axis on the number of NPs on duty, controlling for the baseline control vector (i.e., indicators for ED-by-year, ED-by-month, ED-by-day-of-the-week, and ED-by-hour-of-the-day).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

Figure A.3: Predicting Log Length of Stay and Log ED Cost

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

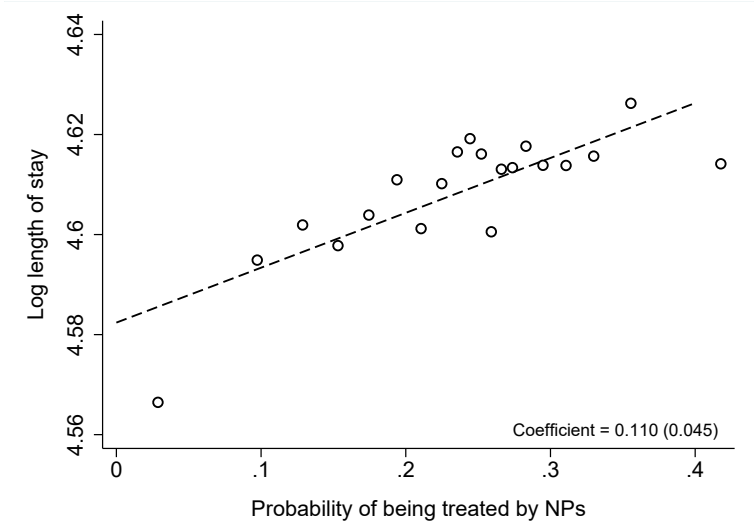
Figure A.4: Monotonicity Test



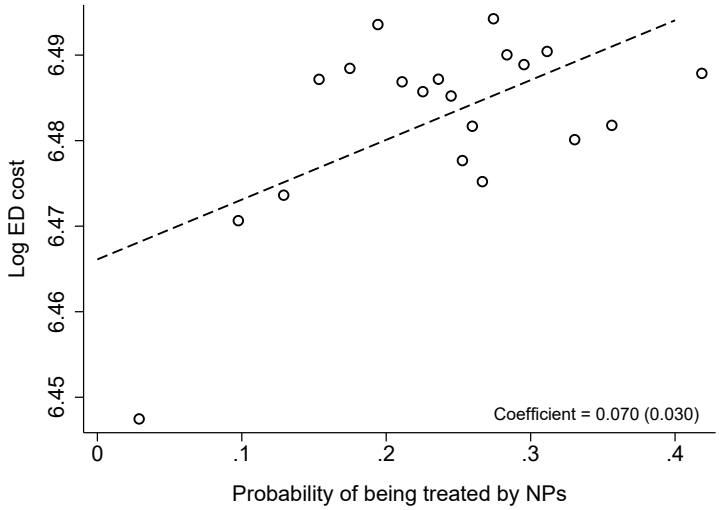
Notes: This figure shows the first-stage regression for cases of different characteristics. Panels A-F split the sample by, respectively, age (above versus below the median of the sample), marital status, gender, race (Black versus non-Black), total number of Elixhauser comorbidities (above versus below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

Figure A.5: Visual IV

A. Log Length of Stay

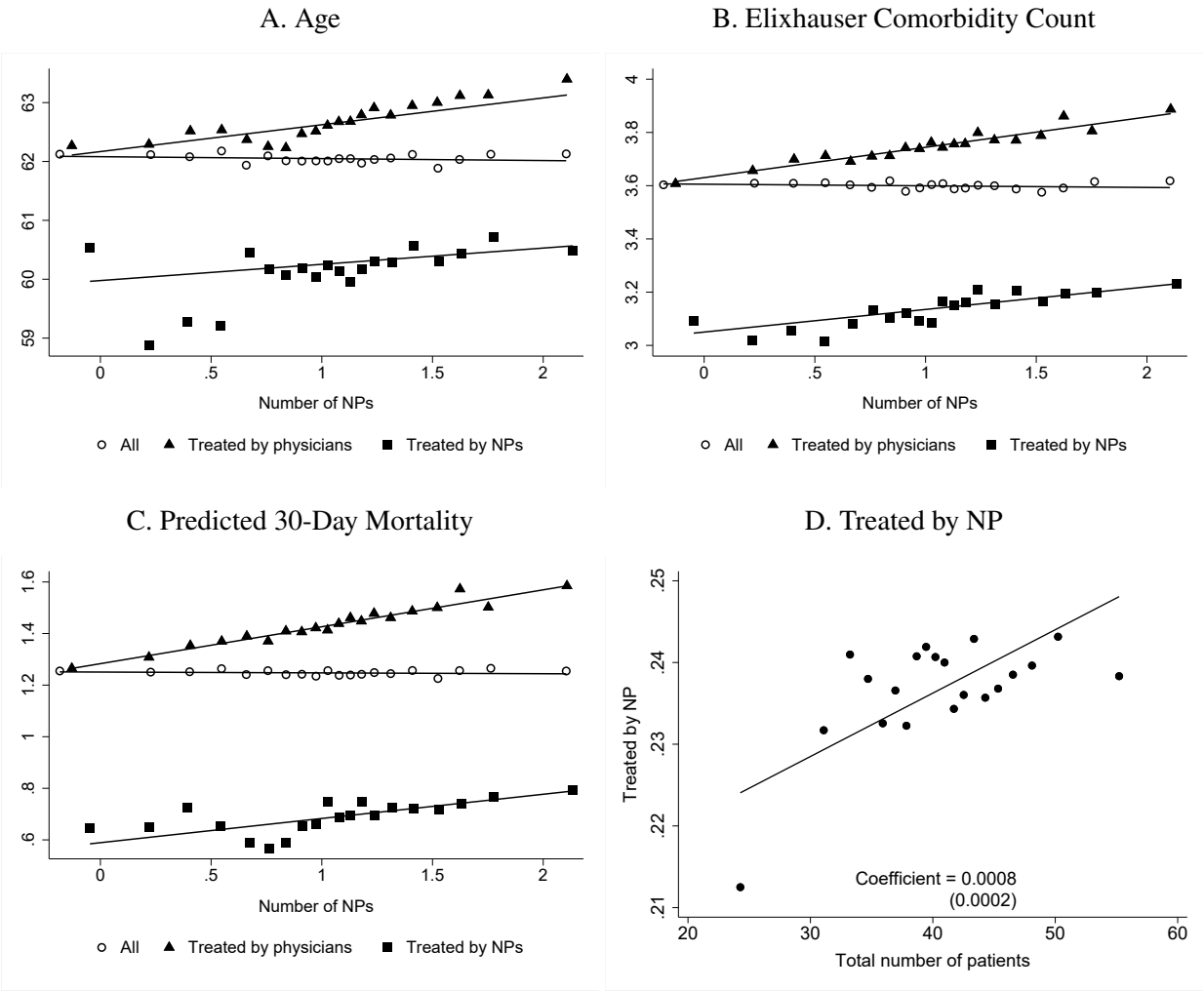


B. Log ED Cost



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

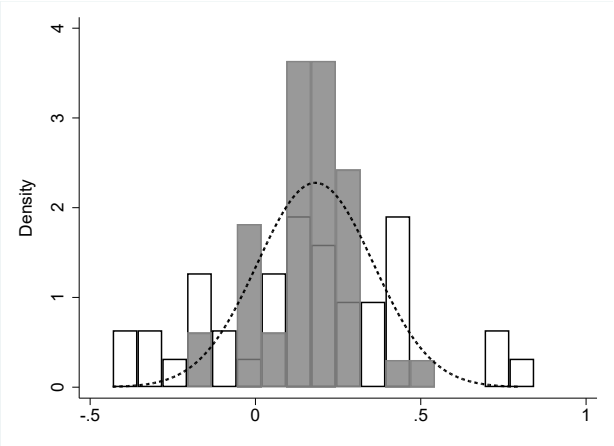
Figure A.6: Patient Assignment



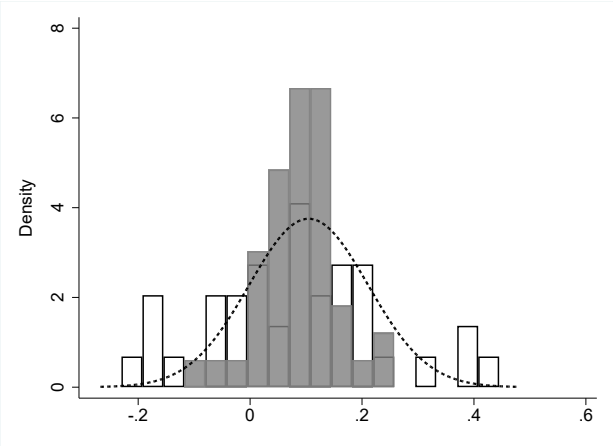
Notes: Panels A-C report patient characteristics (age, total number of Elixhauser comorbidities, and predicted 30-day mortality, respectively) on the *y*-axis against the number of NPs on duty on the *x*-axis. The unit of observation is at the case level. Both the *y*-axis and *x*-axis variables are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

Figure A.7: ED-Specific Estimates of NP Effect

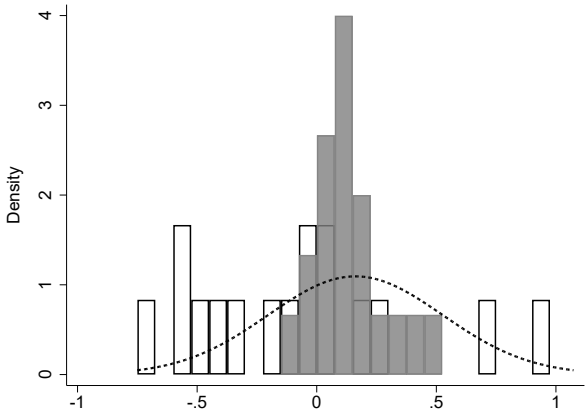
A. Log Length of Stay



B. Log ED Cost

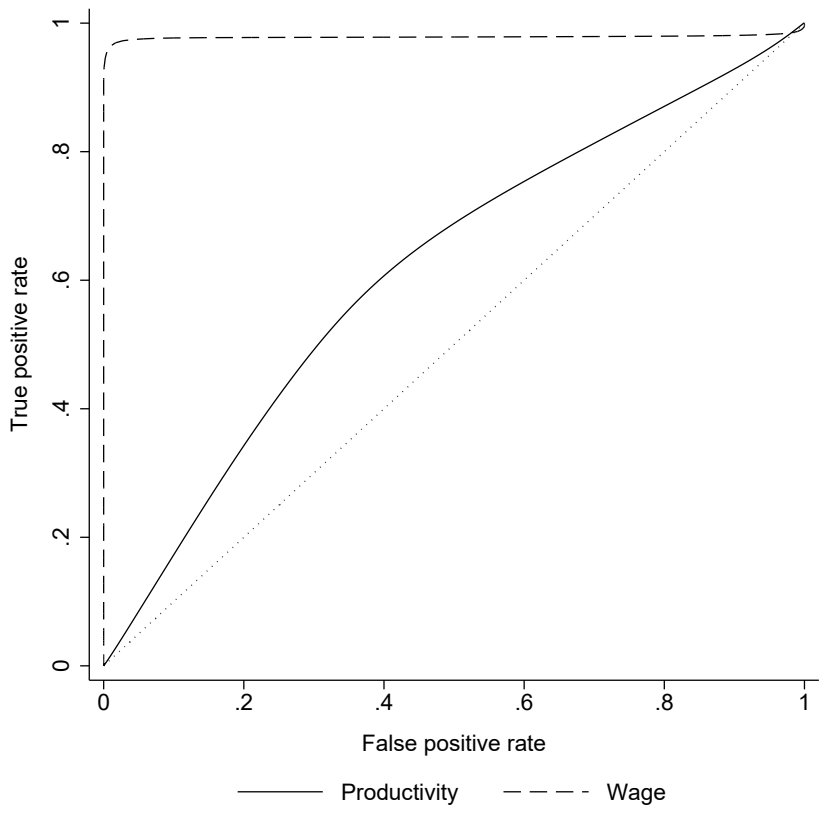


C. 30-day Preventable Hospitalization



Notes: This figure reports the distribution of ED-specific IV estimates of the NP effect. Panels A, B and C report results for the NP effect on log length of stay, log cost of the ED visit, and 30-day preventable hospitalization, respectively. The white bins show the histogram of ED-specific IV estimates without any adjustment to account for estimation noise. The gray bins show the histogram of ED-specific IV estimates with empirical Bayes adjustments (see details in Appendix A.5). The dashed lines show the standard normal density with a variance of the prior distribution of ED-specific IV estimates for each outcome. Panels A and B display estimates for all 44 EDs in our sample. As 30-day preventable hospitalization is not common (occurs in less than 2 percent of the sample), the estimates are relatively imprecise when using observations from a single ED, Panel C thus includes only EDs with at least 25,000 cases in the analysis sample (in total 20 such EDs).

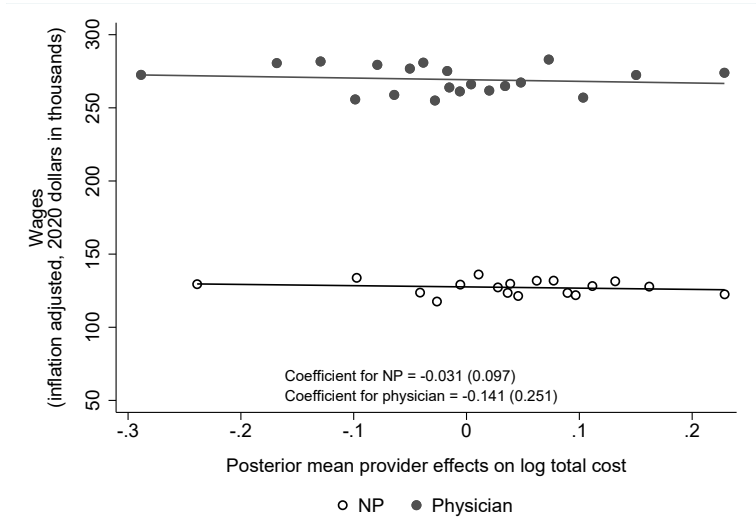
Figure A.8: Receiver Operating Characteristic (ROC) Curve



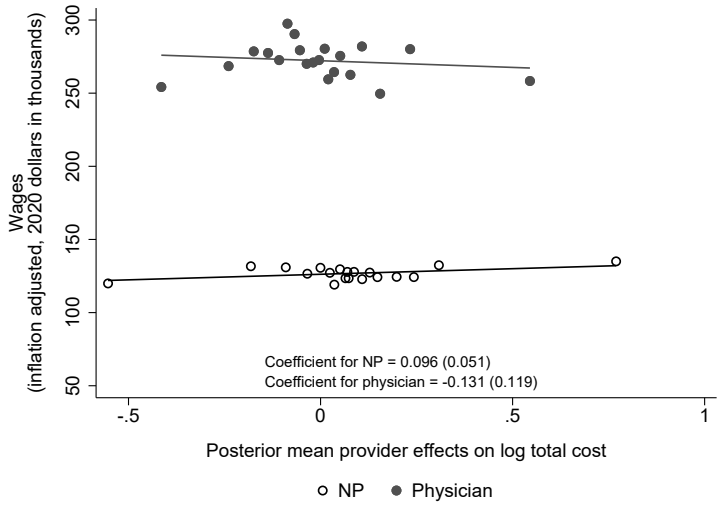
Notes: This figure displays the receiver operating characteristic curves for productivity (in the solid line) and wages (in the dashed line). The dotted line plots the 45-degree line. Productivity is defined as the additive inverse of provider-specific effects on log total spending associated with the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

Figure A.9: Productivity versus Wages

A. Linear Shrinkage



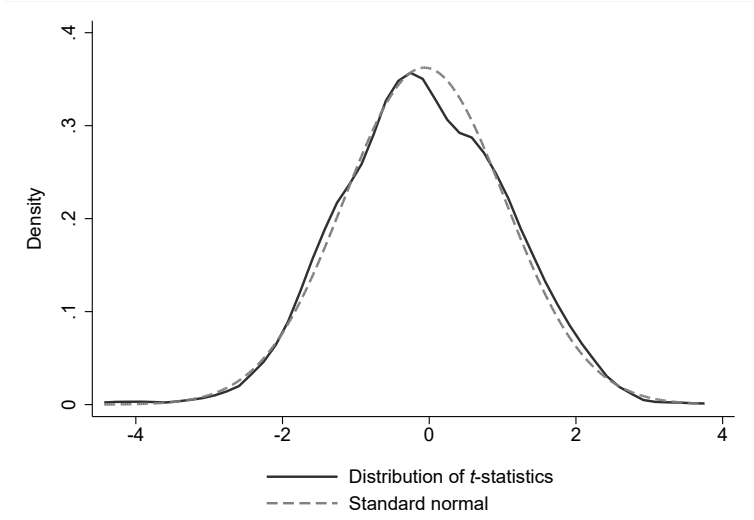
B. Deconvolution Shrinkage



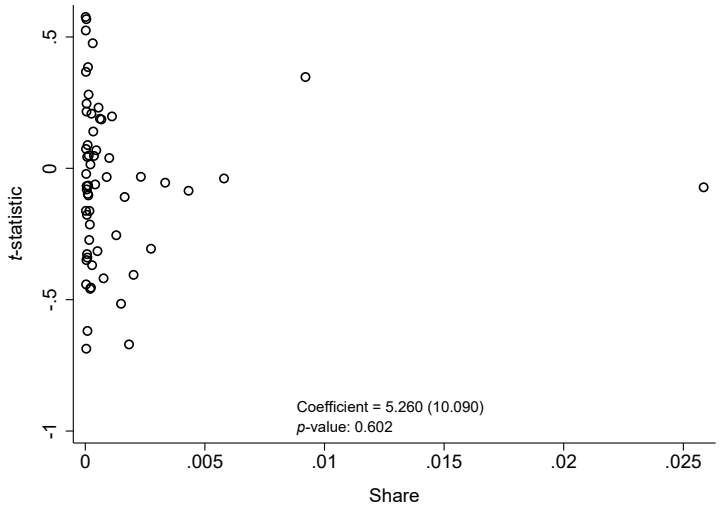
Notes: This figure shows binned scatter plots of provider yearly wage on the y-axis versus posterior mean provider effects on log total cost associated with the ED visit on the x-axis. Both the y-axis and x-axis variables are residualized with respect to ED indicators, with means added back for ease of interpretation. Wages are inflation adjusted to year 2020. Coefficients from regressions of wages on posterior mean provider effects controlling for ED indicators are reported, with standard errors clustered by ED shown in parentheses. The hollow circles report results for NPs; the solid circles report results for physicians. Panel A shrinks provider effects linearly towards the grand mean with weights $w_j = \frac{\hat{\psi}^2}{s_j^2+\hat{\psi}^2}$, where $\hat{\psi}^2$ and s_j^2 are, respectively, the variance of the prior distribution of provider effects and the variance of the sampling error for each provider. Panel B constructs posterior mean provider effects using the deconvolved density. See more details in Appendix A.4.5.

Figure A.10: Diagnosis Coding: NPs versus Physicians

A. Distribution of *t*-Statistics

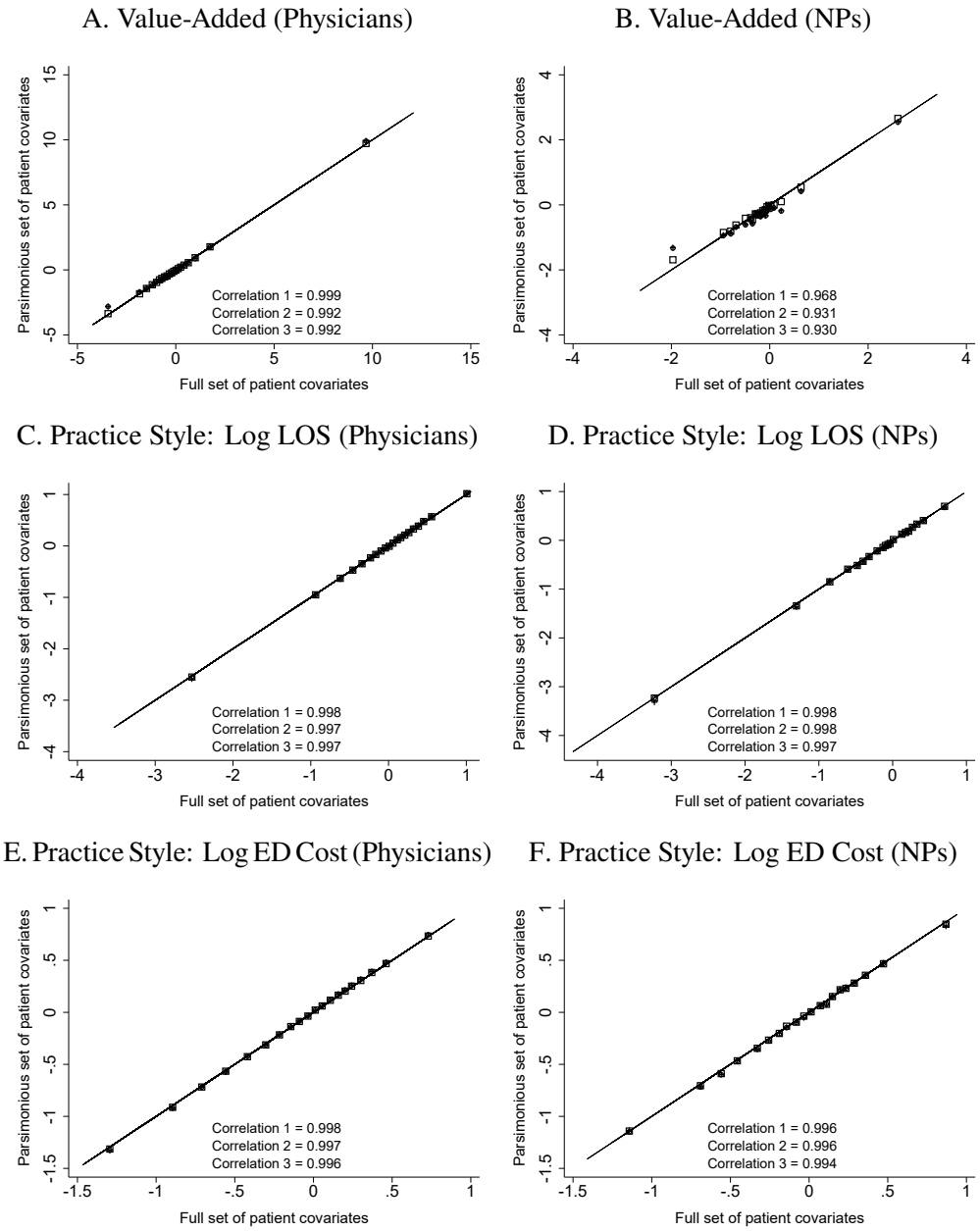


B. *t*-Statistics versus Diagnosis Prevalence



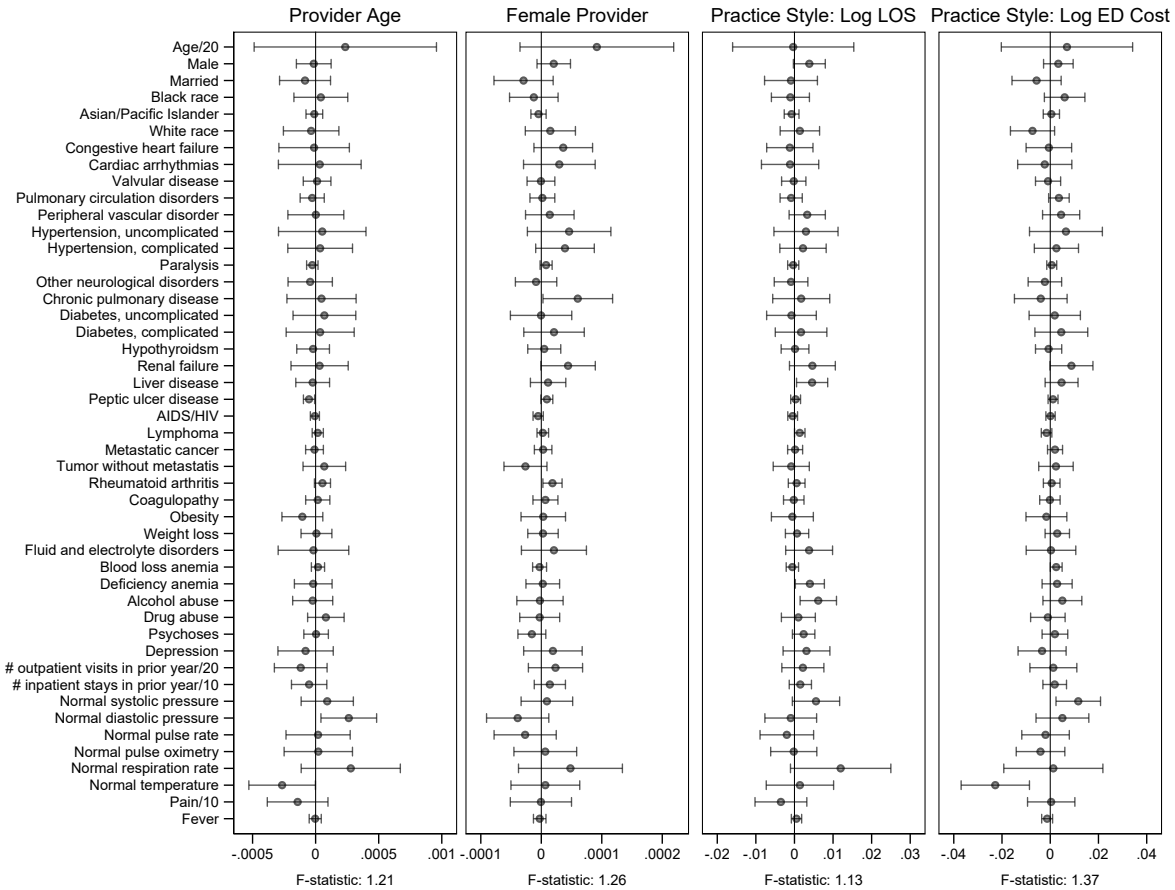
Notes: Panel A plots the distribution of the *t*-statistics on whether NPs and physicians are significantly different in diagnosis coding from 836 separate regressions that use each three-digit diagnosis indicator as the outcome variable. The distribution is estimated using an Epanechnikov kernel with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

Figure A.11: Stability of Provider Value-Added and Practice Style with Varying Patient Covariates



Notes: This figure shows the stability of provider value-added and practice style estimated using alternative patient covariates. See Appendix A.3 for details. The *x*-axis in each panel reports provider value-added/practice style constructed using the full set of patient covariates, including demographics (five-year age-bin indicators, marital status, gender, and race indicators), indicators for 31 Elixhauser comorbidities, prior health care use (the number of outpatient visits and the number of inpatient stays in VHA facilities in the prior 365 days), vital signs, and indicators for three-digit ICD-10 code of the primary diagnosis of the visit. The *y*-axis reports provider value-added/practice style constructed using alternative sets of patient covariates: Parsimonious set 1 that includes demographics, three-digit diagnosis indicators, and 31 Elixhauser comorbidities; parsimonious set 2 that includes demographics and three-digit diagnosis indicators; parsimonious set 3 that includes five-year age-bin and three-digit diagnosis indicators. Valued-added and practice style constructed using parsimonious sets 1-3 are shown in squares, circles, and "+", respectively. Correlations 1-3 report correlations of value-added/practice style estimated using the full set of patient covariates with those using parsimonious sets 1-3, respectively. The solid lines show the 45-degree line. Panels A, C and E report results for physicians. Panel B, D and F report results for NPs.

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



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

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

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

Notes: Panel A reports summary statistics for NPs; Panel B reports summary statistics for physicians. Column 1 reports summary statistics for NPs/physicians working at the ED in our analysis sample. Column 2 reports summary statistics for NPs/physicians working at the ED observed in the 20 percent Medicare data (with age and gender information obtained from the Medicare Data on Provider Practice and Specialty (MD-PPAS)). To provide a description of providers outside of the ED, Column 3 reports characteristics of all NPs/physicians (regardless of working at the ED) observed in the 20 percent Medicare data. VHA ED NPs' and physicians' mean age and female share reported in Column 1 are slightly different from those reported in Figures 5 and 6 because the latter weight means by the number of ED-days a provider works and patient volume.

Table A.2: Selection of Baseline Sample

Sample step	Description	Cases	Providers		EDs
			NPs	Physicians	
1. Build sample of ED cases from January 1, 2017, to January 31, 2020.	We restrict the sample to cases after the VHA directive granting NPs full practice authority in December 2016 and before COVID pandemic in the US.	7,886,164	547	5,749	146
2. Include only cases visiting during daytime.	Empirically NPs do not work outside of the hours of 8 a.m. to 6 p.m. We drop outside of these cases to focus on cases that could be assigned to an NP.	5,766,296	539	5,665	145
3. Restrict EDs to those with NPs, in months with full practice authority.	We restrict the sample to EDs where NPs work. We restrict to months in which these EDs have granted NPs full practice authority.	3,597,347	521	3,781	111
4. Restrict EDs to those in which NPs and physicians are the only providers.	To focus attention on the margin between NPs and physicians and hold the population of cases seen by an NP or physician fixed, we drop EDs that use other provider types, mainly physician assistants.	1,119,396	156	1,348	44
5. Drop cases with missing demographics or extreme ages.	We drop cases with missing age or gender, or age above 99 or below 20.	1,118,836	156	1,348	44

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

Table A.3: Complier and Never-Taker Characteristics

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

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

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

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

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

Table A.5: NP Presence and Outcomes of Patients Treated by Physicians

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

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

Table A.6: Robustness to Additional Controls

	Dependent variable				
	Log length of stay (1)	Log ED cost (2)	Admission (3)	30-day mortality (4)	30-day prevent. hosp. (5)
Panel A: Baseline					
NP assignment	0.110	0.070	0.103	-0.116	0.252
	(0.045)	(0.030)	(0.585)	(0.115)	(0.120)
Panel B: Control for patient volume					
NP assignment	0.095	0.081	0.597	-0.082	0.237
	(0.045)	(0.030)	(0.606)	(0.116)	(0.118)
Panel C: Control for doctor equivalents (1 NP = 0.341 physicians)					
NP assignment	0.110	0.070	0.103	-0.116	0.252
	(0.044)	(0.029)	(0.584)	(0.115)	(0.120)
Panel D: Control for doctor equivalents (1 NP = 0.5 physicians)					
NP assignment	0.085	0.061	-0.019	-0.104	0.245
	(0.043)	(0.029)	(0.574)	(0.113)	(0.117)
Panel E: Control for wait time					
NP assignment	0.109	0.069	0.317	-0.074	0.250
	(0.045)	(0.030)	(0.594)	(0.124)	(0.121)
Panel F: Control for average risks of patients treated by physicians					
NP assignment	0.100	0.068	0.191	-0.114	0.258
	(0.044)	(0.030)	(0.596)	(0.117)	(0.124)
Full controls	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	4.608	6.483	16.625	1.247	1.234
S.D. dep. var.	1.161	0.878	37.230	11.099	11.041
Observations	1,110,798	1,108,961	1,118,836	1,118,836	1,118,836

Notes: Panel A repeats our main estimates reported in Tables 2 and 3. Panel B adds a control for patient volume in the analysis time window (i.e., 8 a.m. to 6 p.m.) of the ED-day cell of the patient's visit. Panels C and D add a control for the total number of doctor equivalents on duty at the ED on the day the patient visits. Panel C assumes a substitution rate of 0.341 between NPs and physicians; Panel D assumes a substitution rate of 0.5. Panel E adds a control for patient wait time. As wait time is potentially endogenous (healthier cases could be assigned a lower priority and hence wait longer), we add an instrument for wait time: the average wait time of cases visiting on the same day at the same ED as the index case. Panel F adds a control for the average predicted 30-day mortality risk of patients treated by physicians in the ED-day cell. The outcomes in Columns 1-5 are, respectively, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, and 30-day preventable hospitalization. All estimations include the full set of controls described in the notes to Table 2. Standard errors clustered by provider are reported in parentheses.

Table A.7: Patient-Provider Gender Match

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

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

Table A.8: Alternative Standard Error Clustering

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

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

Table A.9: Alternative Instruments

	Dependent variable				
	Log length of stay	Log ED cost	Admission	30-day mortality	30-day prevent. hosp.
	(1)	(2)	(3)	(4)	(5)
Panel A: Include NPs with only one case					
NP assignment	0.102	0.076	-0.011	-0.127	0.285
	(0.045)	(0.030)	(0.594)	(0.116)	(0.125)
Panel B: Leave out the index case					
NP assignment	0.123	0.074	0.198	-0.110	0.260
	(0.046)	(0.031)	(0.605)	(0.121)	(0.126)
Panel C: Leave-out share of cases treated by NPs					
NP assignment	0.117	0.069	0.926	-0.033	0.208
	(0.052)	(0.032)	(0.628)	(0.118)	(0.121)
Panel D: Indicator for any NP on duty					
NP assignment	0.108	0.080	0.185	-0.048	0.211
	(0.049)	(0.030)	(0.643)	(0.121)	(0.130)
Full controls	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	4.608	6.483	16.625	1.247	1.234
S.D. dep. var.	1.161	0.878	37.230	11.099	11.041
Observations	1,110,798	1,108,961	1,118,836	1,118,836	1,118,836

Notes: Panel A reports results using an alternative measure of the number of NPs on duty as the instrument, which includes NPs with only one case in the analysis time window of an ED-day cell. Panel B reports results leaving out the index case in measuring the number of NPs on duty. Panels C uses the share of cases treated by NPs in the ED-day cell (leaving out the index case in calculating the share) as the instrument. Panel D uses an indicator for any NP on duty as the instrument. The outcomes in Columns 1-5 are, respectively, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, and 30-day preventable hospitalization. All estimations include the full set of controls described in the notes to Table 2. Standard errors clustered by provider are shown in parentheses.

Table A.10: Sample Restricted to ED-Days with $Z_i \in \{0, 1\}$

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

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

Table A.11: Hospital Admissions and Preventable Hospitalizations Outside VHA

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

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

Table A.12: Heterogeneous Effects by Provider Experience (Cases in 2018-)

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

Notes: This table reports heterogeneous effects of NPs by provider experience using cases visiting in 2018 or after. Panel A shows heterogeneity by provider specific experience in the case's condition, measured as the number of cases with the same three-digit primary diagnosis as the current case the provider has treated since the start of the study period to the day before the current case's visit. Panel B shows heterogeneity by provider general experience, measured as the number of cases (despite diagnoses) the provider has treated since the start of the study period to the day before the current case's visit. For ease of interpretation, both specific and general experience are standardized to have a mean of zero and a standard deviation of one for NPs and physicians separately. The outcome variables in Columns 1-8 are, respectively, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, 30-day preventable hospitalization, whether the patient receives formal consults in the ED visit, whether the patient receives CT scans in the ED visit, and whether the patient receives X-rays in the ED visit. All estimations include the full set of controls described in the notes to Table 2. Standard errors clustered by provider are shown in parentheses.

Table A.13: Heterogeneous Effects by Provider Experience (Prior-Year Experience)

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

Notes: This table reports heterogeneous effects of NPs by provider experience in the prior year. Panel A shows heterogeneity by provider specific experience in the case's condition, measured as the number of cases with the same three-digit primary diagnosis as the current case the provider has treated in the 365 days prior to the day of the current case's visit. Panel B shows heterogeneity by provider general experience, measured as the number of cases (despite diagnoses) the provider has treated in the 365 days prior to the day of the current case's visit. The sample is restricted to cases visiting in 2018 or after, to allow for at least a one-year look-back window for measuring experience in the prior 365 days. For ease of interpretation, both specific and general experience are standardized to have a mean of zero and a standard deviation of one for NPs and physicians separately. The outcome variables in Columns 1-8 are, respectively, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, 30-day preventable hospitalization, whether the patient receives formal consults in the ED visit, whether the patient receives CT scans in the ED visit, and whether the patient receives X-rays in the ED visit. All estimations include the full set of controls described in the notes to Table 2. Standard errors clustered by provider are shown in parentheses.

Table A.14: Heterogeneous Effects by Provider Experience (Measured in Days)

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

Notes: This table reports heterogeneous effects of NPs by provider general experience measured by the number of days the provider has worked since the start of the study period to the day before the current case's visit. For ease of interpretation, the experience measure is standardized to have a mean of zero and a standard deviation of one for NPs and physicians separately. The outcome variables in Columns 1-8 are, respectively, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, 30-day preventable hospitalization, whether the patient receives formal consults in the ED visit, whether the patient receives CT scans in the ED visit, and whether the patient receives X-rays in the ED visit. All estimations include the full set of controls described in the notes to Table 2. Standard errors clustered by provider are shown in parentheses.

Table A.15: Ten Most Common High-Mortality Diagnoses

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

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

Table A.16: Heterogeneous Effects by Patient Characteristics

	Dependent variable				
	Log length of stay (1)	Log ED cost (2)	Admission (3)	30-day mortality (4)	30-day prevent. hosp. (5)
Panel A: Elixhauser comorbidity count					
1st quartile	0.042	0.028	-0.077	-0.147	0.555
	(0.045)	(0.032)	(0.636)	(0.120)	(0.119)
2nd quartile	0.063	0.071	0.291	-0.041	0.438
	(0.044)	(0.030)	(0.642)	(0.126)	(0.132)
3rd quartile	0.117	0.082	-0.245	-0.099	0.250
	(0.048)	(0.031)	(0.761)	(0.178)	(0.186)
4th quartile	0.281	0.122	0.435	-0.203	-0.513
	(0.066)	(0.041)	(1.476)	(0.340)	(0.347)
Panel B: Diagnosis predicted 30-day mortality					
$<$ 95th percentile	0.080	0.064	-0.768	-0.077	0.361
	(0.044)	(0.029)	(0.573)	(0.110)	(0.118)
$\geq$ 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					
Stroke	1.863	0.651	72.609	3.373	-0.062
	(0.677)	(0.311)	(31.758)	(6.038)	(2.379)
AMI	0.806	1.780	123.007	-11.517	-3.219
	(0.562)	(0.655)	(62.695)	(9.684)	(7.593)
Sepsis	1.480	0.095	44.880	24.533	11.117
	(0.609)	(0.329)	(24.114)	(15.169)	(7.961)
Heart failure	1.125	0.088	20.263	1.262	-11.469
	(0.292)	(0.177)	(8.921)	(3.011)	(5.265)
Other	0.097	0.067	-0.343	-0.147	0.332
	(0.045)	(0.029)	(0.578)	(0.112)	(0.122)
Full controls	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	4.608	6.483	16.625	1.247	1.234
S.D. dep. var.	1.161	0.878	37.230	11.099	11.041
Observations	1,110,798	1,108,961	1,118,836	1,118,836	1,118,836

Notes: This table shows heterogeneous effects of NPs by patient characteristics described in Section 5.3. Panel A divides cases into quartiles by their total number of Elixhauser comorbidities, with higher quartiles indicating more complex cases. Panel B divides cases by whether condition severity measured by 30-day mortality of cases with the same three-digit ICD-10 primary diagnosis is equal to or above the 95th percentile of the sample. Panel C divides cases by their condition. The outcomes in Columns 1-5 are, respectively, log length of stay, log cost of the ED visit, hospital admission in the ED visit, 30-day mortality, and 30-day preventable hospitalization. All estimations include the full set of controls described in the notes to Table 2. Standard errors clustered by provider are shown in parentheses.

Table A.17: Variance of Provider Effects on Medical Spending

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

Notes: This table reports variance of provider effects on log total spending associated with the ED visit. Total spending associated with the ED visit is computed as the sum of the three main components of costs that we find significant NP effects: the cost of care at the ED, hospital admission, and 30-day preventable hospitalizations (we multiply the latter two components by the average cost of a hospital stay, $19,220). Row 1 reports variance of provider effects $\hat{\theta}_j$ estimated using Equations (A.4) and (A.5). To account for biases due to estimation noise in $\hat{\theta}_j$, Row 2 reports variance using a split-sample approach (details are described in Appendix A.4.2). Column 1 reports variance for NPs. Column 2 reports variance for physicians.

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

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

Notes: This table reports coefficients from regressions of provider-specific z-scores on associated standard errors. Columns 1 and 2 report results using z-scores and standard errors estimated in the full sample. Columns 3 and 4 randomly split cases for each provider into two approximately equal-sized partitions and regress z-scores from one partition on standard errors from the other partition, stacking the two partitions in the regressions. Columns 1 and 3 report results for NPs. Columns 2 and 4 report results for physicians. Standard errors clustered by ED are reported in parentheses. The number of unique providers in Columns 1 and 2 are smaller than those reported in Section 2.3 because our deconvolution includes only providers with least 150 cases (to restrict the inclusion of nosiy provider effect estimates). Columns 3 and 4 have smaller numbers of providers than those in Columns 1 and 2, respectively, because Columns 3 and 4 further drop providers with less than 150 cases in each split sample.