The Productivity of Employees and Contractors: Evidence from the Emergency Department*

Diego Battiston† Jordi Blanes i Vidal‡

Elena Lucchese §

March, 2026

Abstract

Firms often use external contractors to fill key roles, yet little is known about their productivity. We compare permanent and contracted doctors in the Emergency Department of a large public hospital. Both perform identical tasks, and patients are quasi-randomly assigned, allowing causal inference. Contracted doctors work 35% slower, spend 8% more on tests and treatments, and do not achieve better outcomes. Crucially, these gaps reflect the causal effect of employment status rather than selection. We uncover an important mechanism: employees perform more unpaid overtime to discharge their patients promptly, but only when doing so helps their fellow employees.

JEL *classification*: J24, J41, D23, I11.

Keywords: Contractors, Labour Productivity, Emergency Department, Overtime.

*We thank David Chan, Bob Gibbons, Catherine Thomas, Jose Vasquez and audiences in various seminars and conferences. No financial support was received for this paper. All errors remain our own.

†School of Economics, University of Edinburgh. Email: diego.battiston@ed.ac.uk

‡Department of Management, London School of Economics and Center for Economic and Policy Research. Email: j.blanes-i-vidal@lse.ac.uk.

§University of Milano-Bicocca. Email: elena.lucchese@unimib.it.

1 Introduction

A sizeable minority of workers are independent contractors (Katz and Krueger, 2016).[1] For instance, Abraham et al. (2024) estimate that share at ∼15% of the US workforce. Contractors' importance is both increasing and particularly prevalent in high-skilled industries such as professional and business services, where they frequently work alongside permanent employees (Lim et al., 2019). The scale and growth of this segment of the labour force has stimulated an expanding literature on non-traditional work arrangements, examining workers' preferences for flexibility, job security, and working conditions more broadly (Bloom et al., 2015; Mas and Pallais, 2017; Ameriks et al., 2020).

In this paper, we provide causal evidence on a core question concerning contractors: how their productivity compares with that of permanent employees. Theory is ambiguous. Steeper incentives arising from contract-renewal risk (Shapiro and Stiglitz, 1984), together with a more diverse experience accumulated across institutions (Ozgen, 2021), could make contractors more productive than employees. Conversely, employees are likely to possess greater firm-specific human capital and be more committed to, and embedded within, the organisation (Akerlof and Kranton, 2005). Assessing performance differences between contractors and employees (along with the mechanisms behind these differences) is critical for understanding how contractual arrangements shape worker productivity, a core question in labour and organisational economics (Holmstrom and Milgrom 1994, Lazear 2000). Yet progress has been hampered by data limitations. Large administrative (such as matched employer-employee) datasets, for instance, are not well-suited to direct measurement of worker productivity (Eeckhout and Kircher 2011, Cornelissen et al. 2017); they make it difficult to hold constant the tasks and other conditions under which workers operate; and they do not track the contractual relations between the self-employed and the buyers of their services.

We make progress in this question by leveraging a rich dataset from a high-skilled, high-stakes organisation: the Emergency Department (henceforth, ED) of one of Europe's best and largest hospitals. The health sector is a good setting in which to study performance differences between employees and contractors, as both forms of labour coexist extensively within it (Doyle et al., 2024). A further advantage of our hospital setting is that we observe employee and contractor ED doctors performing the same job in the same location and treat-

[1] The precise legal definition of contractor varies across jurisdictions. Broadly speaking, a contractor is a self-employed worker operating under a contract for services, as opposed to an employment agreement. Employees and contractors typically differ in terms of their working conditions, legal and liability protections, tax regimes and benefits received.

ing identical patients. Access to detailed input and output variables allows us to measure performance at a granular level and investigate a range of mechanisms through which employees and contractors differ in their behaviour (Chan and Chen, 2022). Crucially, we also observe individual doctors transitioning in their status from contractor to employee, while continuing to do the exact same jobs. Within-doctor analyses then allow us to differentiate the *treatment* effect of contractor status from the *selection* effect of the individuals sorting into that status.

Our data covers more than 300,000 ED visits over a ten-year period. Upon arrival, patients are assessed by a triage nurse and assigned to a pod (i.e. a set of contiguous beds) and, within the pod, to one of the doctors on duty during that shift. Importantly, within-pod assignment to doctors is quasi-random: it depends solely on which doctor has the lowest caseload at the time of arrival, and is therefore orthogonal to both doctor and patient characteristics.[2] This feature of the institutional environment underpins our research design. Throughout the paper, we control for pod-by-shift fixed effects, allowing us to compare performance across doctors who work under identical conditions and treat, on average, the same types of patients.

We find that contractors use more hospital resources than employees, without delivering better patient outcomes. The length of stay of patients assigned to contractors upon arrival is 30% higher than for patients assigned to employees. The total cost of ED care is 7% higher for patients assigned to contractors.[3] Among other items, this includes a 43% increase in the cost of laboratory tests, a 10% increase in the likelihood of requesting an x-ray, and a 13% increase in the likelihood of requesting an electrocardiogram. In contrast, patients assigned to contractors are no more or less likely to be hospitalised, die in the ED, or return to the ED within 2, 5, 30 or 60 days after being discharged.

In our hospital, patients are not exclusively assigned to a single doctor for the duration of their stay. Instead, a doctor reaching the end of her official shift can finish work and hand over her patients to a doctor starting work in the same pod for the next shift. We find that patients initially assigned to a contractor are 25% more likely to be handed over to a doctor in the next shift. We show that these extra handovers are inefficient because, in addition to creating extra work for incoming doctors, they can generate unnecessary duplication of tests

[2]For studies leveraging the quasi-random assignment of patients to doctors, see Doyle et al. (2010), Chan (2016) and Chen (2021). We validate quasi-random assignment in our setting through a battery of falsification tests.

[3]Cost and length of stay are standard indicators of resource use in an ED setting. Length of stay involves the temporary use of a hospital bed, which matters because it affects the waiting times of the patients yet-to-be-seen in the ED. Waiting times in turn adversely affect both patient satisfaction and health (Thompson et al. 1996, Smalley et al. 2020, Nyce et al. 2021, Jones et al. 2022).

and treatments.

We begin to investigate mechanisms for these differences by exploiting within-doctor variation. We first find that the estimated differences in resource use (i.e. 7% for costs, 30% for length of stay) remain remarkably robust to including individual doctor fixed effects. We then estimate leads and lags around the transitions from contractor to employee. For length of stay, there is no discernible pre-trend in the lead estimates, a .21 log-point reduction immediately after the transition, and further reductions in the following years. For cost, the estimates are noisier but qualitatively similar. We interpret the within-doctor estimates as indicating that the average differences between employees and contractors reflect the treatment effect of the contractual arrangement, as opposed to the sorting of individuals into these arrangements (Lazear, 2000).

This treatment effect may operate through several mechanisms. For example, contractual status might influence doctors' propensity to engage in 'defensive medicine' (Frakes and Gruber, 2019), due to differences in liability insurance. Contractors and employees may also face different career incentives (Holmstrom, 1999). Finally, employees may be more likely to develop an identity that motivates them to exert effort on behalf of their fellow employees and the hospital more broadly (Akerlof and Kranton 2005, Chen and Li 2009).

We make further progress in the isolation of mechanisms by exploiting the fact that doctors can often avoid handing over patients by *overstaying* (i.e. working unpaid overtime to complete discharges after their official shifts have ended). In our setting, 3.7% of patients are discharged by doctors who overstay. Overstaying involves effort, as it requires giving up leisure time to support hospital throughput and to assist the colleague who would otherwise inherit the patient.[4] Consistent with the idea that departing doctors overstay, at least in part, to help incoming colleagues, we find that overstaying is more likely when the departing and incoming doctors share the same gender and are closer in age.[5]

Examining the decision to overstay helps to clarify the mechanisms through which employees use less resources than contractors. We find that employees are 30% more likely to overstay than contractors. Crucially, this difference also holds within doctors: the same individual becomes more likely to overstay after moving from contractor to employee status. By overstaying to discharge the patient, employees generate a lower length of stay and avoid the duplicated tests and treatments associated with handovers.

[4]In many cases, patients themselves would also prefer to be discharged rather than handed over to a new doctor who must familiarise themselves with the case.

[5]This finding resonates with Battiston et al. (2021), who study a setting in which workers communicate face-to-face to help their colleagues do their job better. Battiston et al. (2021) show that this help is higher among workers of the same gender and similar age.

We then investigate what motivates employees to overstay more than contractors. We find that employees are more likely to overstay *only* when the doctor who would inherit the case is also an employee. Therefore, employees' greater willingness to forgo leisure is not a general behavioural difference, but rather a targeted willingness to assist colleagues who share their employment status within the organisation.

Akerlof and Kranton (2005) and Chen and Li (2009) offer a compelling lens through which to view these findings. They distinguish between *insiders* and *outsiders* within the organisation, and argue that those who identify as insiders are inherently more motivated to exert effort for the collective good. A natural interpretation of our findings is that contractual status shapes doctors' social identity, effectively turning employees into insiders and contractors into outsiders. Social identity, in turn, motivates employee doctors to work more quickly, order fewer tests and undertake unpaid overtime to support the objectives of both the hospital and, as we document empirically, their fellow employee insiders.[6]

Much of the literature on non-traditional forms of labour focuses on measurement issues and on understanding the prevalence of these arrangements among different types of workers and firms (Katz and Krueger 2019, Ameriks et al. 2020, Mas and Pallais 2020, Bernhardt et al. 2023, Abraham et al. 2024). A related strand examines workers' preferences for flexibility in scheduling or work location (Mas and Pallais 2017, He et al. 2021, Maestas et al. 2023). Other contributions analyse the distributional consequences of outsourcing, for wage inequality (Goldschmidt and Schmieder 2017, Drenik et al. 2023), as well as the functioning and economic effects of online labour platforms (Cohen et al. 2016, Chen et al. 2019, Dube et al. 2020, Castillo 2025, and Stanton and Thomas 2025).

Despite this rich body of work, little is known about how worker productivity varies across contractual forms. A partial exception is Muralidharan and Sundararaman (2013), who experimentally evaluate the effect of *adding* a contract teacher to the existing roster of publicly employed teachers. Our approach differs in that we hold the number of workers fixed (i.e. at one doctor) and instead examine whether worker type, rather than workforce size, affects productivity.[7]

Non-traditional arrangements can differ from traditional employment both in terms of *conditions* (such as increased flexibility on location or schedule) and/or *legal status* (i.e.

[6] An alternative framework would posit that the higher employee/employee overstay is part of a reciprocal relational contract. This contract would consist of a doctor overstaying today in the expectation that the recipient of the favour provides a reciprocal favour in the future. However, we find no evidence that the patterns of shift rotations provide employees with more opportunities to exchange favours, relative to contractors.

[7] Hirsch and Mueller (2012) and Cappellari et al. (2012) use panel data methods to compare firm-level productivity across establishments with differing shares of temporary and permanent employees.

the contract linking worker and firm) (Mas and Pallais, 2020). A small set of papers vary conditions while holding status fixed. Bloom et al. (2015), for instance, compare productivity differences between workers who are all employees, but are based either in the office or at home (see also, Gibbs et al. 2023 and Emanuel and Harrington 2024). Likewise, Boltz et al. (2023) hold employment status constant while examining the productivity effects of schedule flexibility. By contrast, we show that contractual status itself matters for productivity, even among workers operating under identical conditions.

Our findings are particularly relevant for the health sector and contribute to the growing literature that uses the Emergency Department as a fertile setting for studying high-skilled work (Chan 2016, Silver 2021, Schwab and Singh 2024). The closest paper in this area is Chan and Chen (2022), who compare the productivity of doctors and nurse practitioners (NPs) working side-by-side in the ED. They show that NPs are 12% slower and generate 7% higher costs than doctors. By contrast, we identify similar or even larger effects *within individuals* who differ only in contractual status, rather than *across individuals* who differ in ability, human capital, and professional status. Another related contribution is Chan (2018), which demonstrates that decisions made towards the end of a shift disproportionately affect both length of stay and the cost of providing care. Although our institutional setting differs, we share the idea that the end of a shift forces doctors to trade off leisure against advancing the hospital's objectives.[8]

A central result in our paper is that employees are more willing to undertake unpaid overtime to discharge patients that would have been inherited by other employees. Willingness to cooperate in organisations has long been a fertile area of both theoretical (Kandel and Lazear 1992, Rotemberg 1994) and empirical (Bewley 1999, Costa and Kahn 2003, Bandiera et al. 2005) research. A particularly relevant study here is Battiston et al. (2021), who study help in organisations and find that it is higher within homogenous teams. Also related is Delfgaauw et al. (2022), who show that team incentives deliver higher performance when the groups are more socially cohesive. We contribute to this literature by highlighting the role of workers' contractual status in their willingness to cooperate with colleagues.

Section 2 describes the institutional setting and the dataset. Section 3 discusses the empirical strategy. In Section 4 we present the baseline results. In Section 5, we explore

[8]In Chan (2018), doctors are not permitted to finish work until all their assigned patients have been discharged. In our setting, doctors may leave immediately once their shift ends, and 36% of patients are handed over to the incoming doctor in the next shift. Thus, the end-of-shift decisions that Chan studies revolve around how many patients to admit and how much to spend on each patient to expedite their discharge. We instead study decisions over whether to undertake unpaid overtime to discharge patients rather than hand them over.

potential mechanisms. Section 6 discusses external validity and concludes.

2 Setting and Data

Italy's Health System Italy's public healthcare system (Servizio Sanitario Nazionale, or SSN) shares many characteristics with those of other European countries. The SSN provides universal coverage, and EDs serve all patients regardless of insurance or ability to pay. Access to ED care is available 24/7 and free for urgent cases. Non-urgent visits may incur a small fee.[9] National guidelines target waiting times and length of stay as the main performance measure by which hospitals are evaluated (Ministero della Salute, 2024).

Niguarda Hospital We study the ED of Grande Ospedale Metropolitano Niguarda in Milan. Niguarda is one of the largest and most renowned hospitals in Europe, recently ranking among the top 40 hospitals worldwide and second in Italy (Kayser, 2025).

The ED receives approximately 80,000 patient visits per year. It is organised as a system of pods, which are clusters of beds that are located in a specific area of the hospital. There are two types of pods: general and specialist. Specialist pods handle cases in specialised disciplines such as obstetrics, paediatrics, orthopaedics, psychiatry, and orthodontics. They are typically staffed by a single doctor whose primary training lies outside emergency medicine.

General pods, in contrast, are staffed by doctors specialised in emergency medicine and often host multiple doctors at a time. Throughout the sample period, six pods were used, though not all were in use at all times. The pods are called Medica, Chirurgica, M1, M2, M3 and M4. Some degree of specialisation exists within this group of pods. For example, patients expected to remain in the ED for extended periods are usually assigned to pods M2 or M4.[10]

Upon arrival, patients are assessed by a triage nurse, who assigns both a priority code (red, yellow, green, or white) and the main complaint. The nurse then allocates the patient

[9] ED visits incur a standard co-payment of €26, which covers all tests and treatments provided during the visit. This fee is waived for high-priority emergencies and for vulnerable groups such as children and pregnant women. This institutional setup creates a largely uniform treatment environment where financial considerations do not limit the care each patient receives.

[10] Table A1 reports estimates of predicted length of stay and predicted cost (computed on the basis of patient characteristics) on pod assignment. As the set of pods used has changed over our sample period, our empirical analysis always controls for the shift (i.e., the morning, afternoon or night shift within a specific date) during which the incident arrives to the ED. We find large differences in the sets of cases assigned to different pods, which indicates that assignment to pod cannot be regarded as exogenous.

to a pod and, within that pod, to a specific doctor. Pod assignment depends on the nature of the medical incident, but *doctor assignment within a pod does not*. Instead, triage nurses explicitly aim to equalise caseloads across doctors working in the same pod.[11] In Section 4, we use a variety of tests to validate the assumption that assignment of cases to doctors is orthogonal to patient and doctor characteristics.

The ED operates three shifts per day: 8am–2pm, 2pm–8pm, and 8pm–8am. The assignment of doctors to specific shifts and pods is determined several weeks in advance. Contrary to other settings (Chan, 2018), doctors are not required to remain on duty beyond their scheduled shift. Patients who have not been discharged at the end of a shift are typically handed over to an incoming doctor in the same pod. However, doctors occasionally choose to stay beyond their official hours to complete treatment and discharge their patients. As discussed in Section 5, this practice, though voluntary, is often seen as preferable to handing cases over to another doctor. In our setting, 3.7% of all incidents are discharged by an *overstaying* doctor (i.e. a doctor staying beyond their shift doing unpaid overtime).

Throughout the paper, we study the contractual status *of the doctor initially assigned to the incident*. While we argue that this initial assignment can be treated as exogenous, our setting is not well designed to assess the role of subsequent doctors who may take over the patient if the originally assigned doctor finishes their shift before discharge.

The Co-Existence of Employees and Contractors The health sector offers a useful setting in which to study performance differences between employees and contractors, as both forms of employment coexist extensively within it (Doyle et al., 2024). Traditionally, Italian public hospitals staffed their EDs with only full-time, permanent doctors (*medici strutturati*) employed exclusively by the SSN. In recent years, however, hospitals have increasingly relied on external contractors (*contrattisti*) to fill shifts, offering greater flexibility than public hiring procedures allow.

The hiring of contractors began during the public sector hiring freeze introduced by the Berlusconi IV government in 2010. Law 122/2010 sharply limited the replacement of departing public employees (including doctors) as part of a broader fiscal consolidation effort. Successive governments maintained these restrictions, which remained officially in force until at least 2018 (Law 205/2017), although hospitals sometimes managed to find loopholes around the restrictions. Nevertheless, the cumulative effect was a steady contraction of the

[11] This description is based on the authors' visits to Niguarda Hospital and interviews with triage nurses. During one visit, we observed the live allocation process: the nurse consulted the list of cases assigned to each doctor, assessed their relative caseloads, and determined which doctor would receive the next patient.

permanent medical workforce. By 2019, the SSN employed roughly 6,500 fewer specialist doctors than in 2009 (ANAAO Assomed, 2025).

Public hospitals responded by developing ad hoc mechanisms to sustain service provision. One such mechanism was the use of external contractors hired to cover clinical shifts. At Niguarda, ED contractors were engaged through annual contracts specifying both a ceiling on the number of shifts and a per-shift payment.[12] Contractors were integrated into the standard staffing pool and assigned shifts by the ED director alongside employee doctors.[13] Within shifts, both groups performed the same core clinical tasks: assessing patients, ordering diagnostic tests, and making treatment and discharge decisions.

The Niguarda ED also includes doctors who are gaining their qualification. We label these doctors as 'students' (*specializzando*), but they are equivalent to the US *residents* or the UK *junior doctors*. In our empirical analysis, we always include a student dummy, in order to be able to interpret the contractor dummy as the difference between contractors and employees.

Differences Between Employees and Contractors Despite functional equivalence on the ED, employment conditions differed. The first important difference is with respect to compensation. Employee doctors worked full-time with fixed annual hours and received a fixed monthly salary, while contractors were paid per shift worked. Implicit incentives also differed. Employees could pursue internal promotions and leadership roles within the hospital hierarchy, while contractors could not. Conversely, contractors faced renewal risk at the end of their term, whereas employed doctors held permanent tenure within the SSN. One important point of similarity however remained: neither group's pay depended directly on their individual contribution to ED resource use or on patient outcomes.

A second major difference concerned integration into the hospital organisation and culture. Employees were embedded in its collective routines: they attended training sessions, departmental meetings, and took part in shared decision-making. For example, they were invited to the monthly meetings where ED performance and complex cases from the pre-

[12]In other hospitals, contractors were hired directly on a per-shift basis, typically at substantially higher rates than Niguarda's contractors. These per-shift contractors are called *gettonisti* in an Italian context. The contractors in Niguarda were not gettonisti, as their contracts were longer-termed. The reason that Niguarda could attract contractors without having to pay the gettonisti rates is that it is regarded as the best hospital in Lombardy.

[13]By law, contractors could refuse the shifts assigned by the ED director. The absence of an employment relation implied that the ED director had no formal power to command contractors to accept a specific timetable. In practise, contractors in Niguarda rarely refused the shifts assigned. In any case, our identification strategy relies on the exogenous assignment of patients within a shift and pod. We do not rely on features of shift assignment for identification.

vious month were discussed. Contractors, by contrast, were excluded from these collective processes and had no formal obligations beyond their assigned shifts. Our discussions with Niguarda staff reveal that employees also self-organised team-building activities, in which contractors did not take part.[14]

Last, employees had a long-term stake in the institution, expecting to remain within the hospital for years. Contractors were regarded, and tended to see themselves, as temporary staff, even if there was a chance of transitioning into permanent posts.[15] Within Niguarda, these contrasts were believed to shape professional identity and team cohesion. Employees were more likely to develop an esprit de corps - a sense of collective responsibility and mutual loyalty within the department. Contractors, by contrast, were seen as maintaining looser attachments to both colleagues and the institution.

Transitions from Contractor to Employee During our sample period, hospitals in Lombardy occasionally had the opportunity to open vacancies for permanent employment, particularly after 2018. Crucially, hospitals could not decide which applicant would be filling a specific vacancy. The procedure operated as follows. Whenever a hospital opened a position, a nation-wide competition was initiated. Applicants were ranked based on rigid, objective criteria, including prior experience, educational qualifications, and academic publications. These criteria were not influenced by hospital preferences, but were instead regulated by national and regional laws (Laws 502/1992 and 165/2001). Softer measures, including reference letters or assessed job performance in previous employment, were not considered (Ministero della Salute, 2023). The hospital that had opened the vacancy was then required to hire the top-ranked candidate.

Our sample spans a ten-year period from January 2014 to December 2023. During this time, 27 ED doctors obtained permanent employment at Niguarda. Of these, 13 were new recruits with no prior experience in the Niguarda ED, while 14 transitioned internally from contractor roles to employee status. Three main factors explain why most contractors in our dataset remained at Niguarda upon securing permanent posts. First, Niguarda is widely recognised as the leading hospital in Lombardy and is therefore a highly desirable employer. Consequently, contractors generally preferred to remain at Niguarda upon securing permanent employment. Second, Niguarda's contractors are typically highly qualified and capable physicians, well positioned to achieve top rankings whenever a new permanent post

[14]These included mountain walks, Christmas dinners and the bi-monthly parties organised by the nursing staff which permanent doctors (but not contractors) regularly attended.

[15]While the majority of contractors transitioned into permanent posts, there was ex ante significant legal and organisational uncertainty about whether these transitions might take place.

at Niguarda was opened. Last, Niguarda is a large employer, so it was relatively easy for contractors to find a position there. Importantly, however, prior employment or a strong performance record as a contractor at Niguarda did not influence an applicant's ranking in the national competition for permanent positions.

Data We use administrative data from Niguarda Hospital. For each ED visit, we observe detailed time stamps throughout the patient's stay, triage and diagnostic indicators, the treatments and tests ordered (along with their costs), the sequence of doctors assigned responsibility for the patient, and the discharge decision. We also observe demographic characteristics for both patients and doctors and, crucially for our analysis, the doctor's employment status at the exact time of the incident. We restrict the sample to cases assigned to a general pod and managed by a specialist in emergency medicine. The final dataset includes 307,875 observations.

Our main dependent variables are the patient's length of stay (defined as the time between assignment to a doctor and discharge) and the cost of care during the ED visit. These are standard measures of resource use in the ED setting (Silver 2021, Chan and Chen 2022). We also analyse the likelihood of hospitalisation (i.e., whether the patient was admitted to the hospital rather than discharged home). As indicators of patient outcomes, we use in-ED mortality and return visits to the ED within 2, 5, 30, and 60 days (Schwab and Singh, 2024).

Table 1 reports summary statistics for all doctors and separately for employees, contractors, and students.[16] On average, employees (42 years old) are only slightly older than contractors (39 years) and considerably older than students (30 years). Gender composition is balanced across groups. Employees work somewhat fewer days per month (13) compared with contractors (14.5) and students (13.7). Our sample includes 85 doctors in total, of whom 52 were employees at some point (alongside 18 contractors and 30 students).

3 Empirical Strategy

We exploit the exogenous assignment of incoming patients to doctors who are working within the same shift and pod combination.

[16] Because employment or contractor status can change over time within an individual, each incident is classified according to the status the doctor held at that specific moment.

Estimating Equation Our baseline estimating equation is:

$$y_j = \beta_1 contractor_{is(j)} + \beta_2 student_{is(j)} + \eta_{sp(j)} + \lambda \mathbf{X_j} + \epsilon_j$$

where j denotes an incident. The main independent variable is an indicator equal to one if the doctor i assigned to incident j was working as a contractor during shift s (i.e. the morning, afternoon, or night shift type on a given date). The key control variable is the combination of shift s in which the incident arrived and pod p to which the incident was allocated. We also control for student status, allowing β_1 to be interpreted as the difference between contractors and employees.[17] $\mathbf{X_j}$ denotes a vector of patient characteristics. In some specifications, we additionally include doctor fixed effects, thereby exploiting within-doctor variation in contractor status.

Our identification strategy rests on the assumption that triage nurses do not systematically assign patients with particular characteristics (such as expected cost or expected length of stay) to contractors relative to employees working at the same time in the same pod.

Balance Tests The inclusion of patient characteristics $\mathbf{X_j}$ primarily serves to improve the precision of our estimates by reducing standard errors. However, the validity of our identification strategy does not hinge on these controls. Instead, a key validation step is to verify that patient characteristics are balanced across contractor and employee status. We conduct such a test in this subsection.

$\mathbf{X_j}$ encompasses a wide range of patient characteristics, including age, gender, triage level, mode of arrival to the ED, insurance status, referral source (e.g. GP, self-referral, etc.), and the main complaint as recorded by the triage nurse. We regress normalised length of stay (Figure 1 Panel A) and normalised ED cost (Figure 1 Panel B) on these normalised characteristics. We find that these variables are highly predictive of resource use, both individually and jointly. For example, the F-statistics testing joint significance are 669 for length of stay (Panel A) and 5,269 for cost (Panel B). The coefficients on individual characteristics are often economically meaningful: over one-third of the coefficients in Panel B exceed .1, implying that a one-standard-deviation increase in the characteristic is associated with a 10%-standard-deviation change in cost.

By contrast, patient characteristics in $\mathbf{X_j}$ are largely uncorrelated with whether the as-

[17] In Section 5, we substitute $contractor_{is(j)}$ by $employee_{is(j)}$. The interpretation is identical but with the reversed sign. We make this substitution to examine the differential behaviour of employees as a function of the employee status of the doctors in the next shift and same pod.

signed doctor is a contractor (Figure 2). The F-statistic of joint significance in a regression of contractor status on patient characteristics is just 1.15. Moreover, nearly all coefficients are statistically indistinguishable from zero and economically negligible: all are below .02, and most fall below .01, implying that a one-standard-deviation increase in a patient characteristic is associated with at most a 1%-standard-deviation change in cost.

We interpret these balance tests in Figure 2 as providing strong support for our identification assumption. Nevertheless, we undertake two additional tests, presented in Figures 3 and 4, and discuss these in the following section.

4 Baseline Results

Average Differences in ED Cost and Length of Stay Table 2, Panel A reports the average resource use differences between contractors and employees working simultaneously within the same pod.[18] On average, contractors spend approximately 7% more financial resources per patient and discharge patients 31% more slowly.[19] Panel B shows that these differences in cost and length of stay remain virtually unchanged after controlling for all patient characteristics displayed in Figure 1.

Additional Tests of Identification Figure 3 expands on this comparison by examining the robustness of the estimated coefficients to the inclusion or exclusion of various subsets of patient characteristics (Chan and Chen, 2022). Specifically, we define seven subsets: gender, age (in 50 quantiles), triage level, main complaint, referral source, mode of arrival, and insurance status. These subsets are highly predictive of ED cost and length of stay (Figure 1). We estimate $2^7 = 128$ separate regressions, each controlling for a different combination of these subsets. The x-axis in Figure 3 represents the number of subsets included in each regression, and for each number, we plot the minimum, maximum, and average estimated coefficients for the effect of the contractor dummy on the dependent variable. The coefficients remain remarkably stable regardless of which subsets are included as controls. We interpret this robustness as supporting our identification assumption that, within a given pod and shift,

[18] Our dataset does not directly record which pod a doctor was assigned to during a given shift. Instead, we infer this assignment from the case-level data, where both the attending doctor and the pod are observed for each incident. In a small number of cases, the same doctor appears assigned to cases across multiple pods within a single shift. To address this, in Table A2 we re-estimate our baseline specifications after resolving these ambiguities by assigning each doctor a unique pod per shift, defined as the pod in which the doctor handled the majority of cases, with ties broken alphabetically. The resulting coefficients are very similar.

[19] The coefficients are .073 and .268. Throughout the paper, we transform the coefficients of variables in logs by using the formula $100 \times (exp^{\beta} - 1)$.

differential selection of incidents to contractor status is not strongly biasing our estimates (Altonji et al. 2005).

Figure 4 provides an additional way to assess our identification assumption and interpret the magnitude of the baseline effects. We first construct predicted length of stay and predicted ED cost using the regressions in Figure 1. We then regress these predicted outcomes on contractor status, and plot the results alongside the corresponding relationships for the actual outcomes.[20] The binned scatter plots and regression lines are displayed in Figure 4. In Panel A, predicted length of stay is uncorrelated with contractor status ($p = .879$), whereas actual length of stay is strongly correlated ($p = .000$). Panel B shows a similar pattern for predicted and actual ED cost ($p = .929$ and $p = .001$, respectively). These comparisons indicate that our baseline coefficients are not driven by systematic differences in the types of incidents assigned to contractors and employees (i.e., the predicted outcomes), but rather by differences in how these doctors manage the incidents to which they are assigned.

Accounting for Within-Shift Timing of Assigned Cases A residual concern is that employees and contractors may receive similar types of cases but at systematically different points within the shift. As we show below, cases handed over to the next shift are associated with higher cost and longer length of stay. If triage nurses disproportionately assigned cases to employees early in the shift and to contractors later, this pattern could confound our baseline estimates by increasing the likelihood of handover for contractor-treated patients.[21] To address this concern, we add the exact hour of assignment to the baseline specification in Table 2, Panel C. The resulting coefficients are very similar, and if anything slightly larger.

Exploiting Within-Doctor Variation in Contractor Status We now begin exploring the mechanisms behind the estimated average differences between employees and contractors. Specifically, we distinguish between *treatment* of contractor status and *selection* of individuals into contractors status. We disentangle these mechanisms empirically by leveraging the fact that fourteen contractors switched to employee status within our sample (see Table A1 for the full transition matrix). We augment our specification with individual fixed effects, which allows us to compare a doctor's performance across different contractual ar-

[20] Predicted outcomes, actual outcomes and contractor status are all residualised on shift–pod indicators.

[21] According to the hospital allocation procedure, this asymmetric assignment should not occur. In Table A3 we confirm that our results are robust to any potential asymmetry in timing. We augment the baseline with controls for the time remaining until the end of the shift and the number of cases previously assigned to the doctor within the shift. While the latter variable is endogenous to the speed at which doctors discharge patients, and therefore constitutes a bad control in the sense of Angrist and Pischke (2009), it is reassuring that even its inclusion leaves the estimated coefficients essentially unchanged.

rangements. The estimates of the contractor dummy, reported in Table 2 Panel D, are very similar to the Panel C estimates without doctor fixed effects, indicating that treatment effects dominate selection effects in our setting.

In Panel E, we further control for the doctor age, number of cases in the previous two years and tenure at Niguarda. We include these controls because switching doctors are older and more experienced during periods when they are observed as employees. While individuals fixed effects control adequately for time-invariant characteristics such as ability and education, doctors may still accumulate human capital over time, which could bias the within-doctor estimates from Panel D. However, even after adding these controls, we continue to find large and statistically significant differences in cost and length of stay.

Addressing Residual Differences in Working Conditions Our empirical strategy controls for the interaction of pod and shift, thereby comparing doctors working under identical conditions. Moreover, employees and contractors perform precisely the same tasks during an ED shift, so the transition from contractor to employee entails no change in responsibilities or working conditions within a shift.

A subtler concern is that systematic differences in the allocation of doctors to shifts and pods could indirectly affect productivity. For instance, if contractors were disproportionately assigned to night shifts, disrupting their sleep and impairing concentration, then even within-shift comparisons could partly reflect differences in rotation patterns rather than contractual status per se.

In Figure A2, we examine whether shift and pod assignments differ systematically by contractual status. We find statistically significant but economically negligible and inconsistent correlations. For example, contractors are unconditionally more likely to be assigned the night shift, yet this coefficient turns negative once doctor fixed effects are included; in both cases the magnitude is less than .03 standard deviations. Tables A4 and A5 show that recent rotation patterns have a negligible effect on our productivity measures. Consistent with this, the estimated effect of contractor status is essentially unchanged regardless of whether rotation controls are included in the baseline specification.

Economic Magnitude of the Differences in Resource Use In the previous subsections, we have found robust and large differences in cost and length of stay between employees and contractors. Our within-doctor estimates indicate that these differences reflect the causal effect of contractual status rather than differences in ability or human capital among doctors who select into a particular arrangement.

At the means of the dependent variables, the differences between employees and contractors correspond to a cost of around 11 euros per patient and an average delay of 77 minutes. To put these economic magnitudes in context, Table A6 reports the standard deviation of the estimated doctor fixed effects from a specification that includes shift/pod effects and patient characteristics. We present both the raw and the Bayes-shrunk estimates, the latter adjusting for differences in sample size across doctors (Chetty et al., 2014). The standard deviation of the Bayes-shrunk effects is .279 for length of stay. This implies that the estimated average difference between employees and contractors in Table 2 is roughly equivalent to replacing a median doctor with one who is a full standard deviation above the median in terms of length of stay. For cost, the Table 2 estimated difference is equivalent to replacing a median doctor with one who is three-fifths of a standard deviation above the median.

A second way to interpret these estimates is to benchmark them against the findings of Chan and Chen (2022), who compare doctors and nurse practitioners. They report that nurse practitioners use 7% more resources and take 12% longer to discharge patients. We therefore find equal or larger effects *within individuals* (as they switch from being contractors to being employees) than they find *across individuals* working in professions associated with very different pay and prestige.

Dynamic Effects around the Transition Events We further exploit the transitions from contractor to employee status in our sample to estimate event-study (lead and lag) effects around these events. Specifically, we estimate:

$$y_{je} = \sum_{\tau=-3}^{4} \pi_\tau \left(transition_{i(j),e} \times year_\tau(j,e) \right) + \eta_{i(j)} + \phi_e + \kappa_{sp(j)} + \phi_{h(j)} + \mathbf{X}_j' \lambda + \varepsilon_{je},$$

where $transition_{i,e} = 1$ if doctor i is the switching doctor in episode e, and $year_\tau(j,e) = 1$ indicates that observation j falls in event-time bin τ relative to the episode transition date (with the 0–1 year pre-transition bin omitted). $\eta_{i(j)}$, ϕ_e, and $\kappa_{sp(j)}$, $\phi_{h(j)}$ are doctor, episode, shift×pod and exact hour fixed effects, respectively, and $\mathbf{X}_j$ includes indicators for patient characteristics.

Staggered DiD estimators (Callaway and Sant'Anna 2021, Sun and Abraham 2021, Borusyak et al. 2024) require a sufficiently large and well-covered set of never-treated and/or not-yet-treated observations across the relevant fixed effect cells to form credible counterfactual trends. Our setting does not satisfy this requirement. Never-treated contractors number only four, and these appear sporadically across shift $\times$ pod cells relative to employee doctors, far too sparse to support these estimators once shift $\times$ pod fixed effects are absorbed. Not-yet-treated observations are equally limited, with thin coverage in pre-transition windows.

Identification in our within-doctor specification therefore relies on a different comparison group: always-employees and early switchers observed well after their transition. To ensure this group supports a clean comparison, we restrict attention to units whose transition occurred at least five years prior to the estimation window, a horizon over which treatment effects can plausibly be assumed to have stabilised. Figure 5 presents estimated effects under alternative sample restrictions, and Appendix B formalises the assumptions underpinning the validity of this approach and provides supporting evidence.

The corresponding estimates are displayed in Figure 5. For length of stay we observe a clear and economically meaningful pattern. Specifically, there is no discernible pre-trend in the lead estimates, followed by a sharp .21 log-point decline in the year immediately after doctors become employees. This reduction grows to .40 after five years. The cost estimates are less precise but display a similar qualitative pattern: no obvious pre-trend, a drop at the transition year, and persistently negative post-transition effects of around up to .10 after five years. Figure A3 provides formal tests of pre-trends (Borusyak et al., 2024), separately for different groups of employees. For all groups, we fail to reject the absence of pre-trends.

We conclude that, taken together, the results in Figure 5 are consistent with the shift from contractor to employee status producing an abrupt and sustained decrease in resource use, as reflected in lower ED cost and shorter patient stays.

Effects on Patient Outcomes Table 3 reports estimates from regressions of various patient outcomes on the contractor indicator. The first two outcomes are hospital admission (the omitted group here is home discharge) and patient death while in the ED. We also construct indicators for patients who return to the ED within 2, 5, 30 and 60 days.[22] The bottom row of Table 3 shows that these outcome variables exhibit significant variation, suggesting that our regressions are unlikely to suffer from low statistical power. For instance, approximately 19% of patients are hospitalised, and 13% return to the ED within thirty days.

All Table 3 empirical specifications include doctor fixed effects, and in Panel B we additionally control for doctor age and recent experience. Across all models, we find no robust evidence that the contractor indicator affects patient outcomes. Putting together

[22]These return indicators are designed to capture cases where the initial medical issue may not have been fully resolved. To qualify as a return, the patient must receive a new incident number; we therefore exclude planned follow-up visits that are part of the original episode.

Tables 2 and 3, we conclude that contractors spend more resources on their patients without achieving correspondingly better health results.

Why would contractors take longer and spend more money without improving patient health? A potential explanation for these findings is that ED doctors are operating in the 'flat of the curve' section of the health function linking treatment intensity with patient outcomes.[23] If this function has diminishing or even zero returns, a large increase in treatment intensity may not translate into an improvement in patient health.

An alternative explanation is that contractors exert lower diagnostic and treatment *personal* effort. If personal doctor effort (e.g., the cognitive labour required to achieve the correct diagnosis) and hospital resource use (e.g. diagnostic tests, observation while occupying a bed, etc.) are substitutes in the health production function, contractors may achieve similar outcomes with different combination of inputs. While we cannot directly observe cognitive effort, we provide evidence below that contractors are less willing to work unpaid overtime and sacrifice leisure, relative to employees.

Effects on Selected Cost Items The total cost of treating a patient in the ED reflects the sum of all procedures, consultations, and tests performed. Tables A7 and A8 list the thirty most important cost items. Table A7 presents the items most frequently used, which together account for 91% of all occurrences in the sample. Table A8 reports the items contributing most to total ED costs, jointly representing 81% of overall spending. Each item is also categorized by cost type (laboratory versus other) according to the Niguarda data system.

Laboratory tests are ordered relatively often, but their low unit cost means they are underrepresented among the items contributing most to overall spending. In contrast, brain CT scans, though less frequent, account for more than 10% of total ED expenditure due to their high per-unit cost. Overall, laboratory expenses represent about 29% of total ED costs.

The first two columns of Table 4 compare employees and contractors across the two main cost components: laboratory and non-laboratory (i.e. other) items. We report estimates from regressions both with and without doctor fixed effects. Contractors are associated with 38% higher spending on laboratory tests and 7% higher spending on other items. The fact that contractors disproportionately order relatively inexpensive but time-consuming laboratory tests helps explain why, in Table 2, we observe much larger effects on length of stay

[23]For recent studies of whether ED doctors operate in the 'flat of the curve', see Doyle (2011) and Silver (2021).

than on ED cost.

The remaining columns of Table 4 report results by specific test type, including blood tests, X-rays, CT scans, electrocardiograms, and ultrasounds. We find large and statistically significant differences for several of these categories. For example, contractors are 3.7 percentage points more likely to order an electrocardiogram, which is a 13% increase relative to the unconditional mean of 28%. Similarly, contractors are 6.6 percentage points more likely to order an X-ray, which is a 10% increase relative to the unconditional mean of 63%. These results indicate that the cost differences between contractors and employees are broad-based rather than concentrated in a small subset of procedures.

Effects on Handovers and Duplicated Cost Items We now examine the broader organisational implications of replacing employees with contractors. Specifically, we examine whether contractors are associated with: (a) a higher number of requests for specialist opinions, (b) more handover of cases to the next-shift doctor, and (c) more duplicated tests and treatments.

In Niguarda, ED doctors can request the opinion of doctors with other specialities. We find in the first column of Table 5 that contractors are not more likely to request that opinion. The estimates are not statistically significant and tiny in size.

As discussed in Section 2, doctors at Niguarda are not required to continue working beyond the end of their shift. Therefore, any patient who has not been discharged by the cut off times of 8am, 2pm or 8pm can be immediately transferred to an incoming doctor in the same pod and next shift. Unlike in other hospital settings, these transfers are routine events in Niguarda.[24] In our sample, 36% of incidents are handed over to a doctor in the next shift (Table 5).

Table 5 reports estimates from regressions of a handover indicator on whether the initially assigned doctor is a contractor, using specifications both with and without doctor fixed effects. In Panel B, contractors are 7.5 percentage points more likely to hand over a patient to the next-shift doctor, an economically large effect relative to a mean of 36%.[25] To understand the implications of this estimated difference, note that, although routine, the handover of patients across shifts entails non-trivial costs (Apker et al. 2007, Chan 2018).

[24] Chan (2018) argues that in the US hospital that he studies 'physicians are expected to complete work on any patient for whom they have assumed care (...) except in uncommon cases where the patient is expected to stay much longer in the ED.' This norm does not apply at Niguarda, where doctors may leave as soon as their shift ends.

[25] This result is related, though not identical, to the evidence on longer length of stay in Table 2. In principle, contractors could take longer with their patients yet still complete treatment before the end of their shift, generating no additional handovers.

First, communicating all relevant clinical information to the incoming doctor (whether face-to-face or electronically) requires time and coordination (Battiston et al., 2021). Second, information may be lost or distorted in the process.[26] Third, handovers impose an additional caseload on the receiving doctor, who must assume responsibility for another patient and consequently has less time for others, including those waiting to be seen. Given these costs and externalities, the literature generally views handovers as an efficiency loss and recommends minimising them when possible (Phillips et al., 2015).[27]

A potential fourth cost of handovers arises from the diagnostic behaviour of the incoming doctor. After taking over a case, the next-shift doctor may feel the need to collect additional evidence, either because some information was lost during the handover or to reassert diagnostic autonomy. While such additional testing may be clinically justified in some cases, it can also lead to the unnecessary duplication of diagnostic procedures, thereby increasing ED costs without improving patient outcomes (Kamat et al., 2013).

To assess this potential cost, we count the number of times each cost item appears within an incident and classify an item as *duplicated* if it occurs more than once. On average, an incident includes .72 duplicated cost items, and Table A9 lists the thirty most common of these. The presence of a duplicated item should not be automatically interpreted as wasteful, as it could reflect standard clinical practice.[28] However, the finding that duplicated items are more likely in incidents that are handed over could be interpreted as suggesting that these duplications likely represent a waste of resources.

Panel A of Table A10 reports estimates from regressions of the number of duplicated cost items on an indicator for whether the incident was handed over to the next shift. Incidents involving a handover exhibit 54% more duplicated costs. We also find large effects when we regress the number of duplications on the number of different doctors who take primary responsibility for the patient (i.e. a patient could be handed over more than once). Each additional doctor is associated with 32% more duplicated costs.

Panel B of Table A10 shows that handovers increase not only the absolute number of duplicated costs but also their importance relative to the total number of cost items. In this panel, the dependent variable is the share of duplicated items among all cost entries.

[26] The existence of standardised handover frameworks such as SBAR (UK) and I-PASS (US) illustrates the challenges of ensuring accurate communication.

[27] This conclusion also draws on the Table 3 result that patients treated by contractors do not experience better health outcomes. If contractors achieved superior outcomes, the longer stays and higher handover rates might be viewed as acceptable trade-offs.

[28] For example, the most common duplicated item in Table A9 is the Troponin test, which is routinely performed multiple times to monitor changes in cardiac enzyme levels. In this case, duplication likely reflects standard clinical practice rather than inefficiency.

We find that this share is 4.9 percentage points higher for incidents with handovers. This coefficient more than doubles the average share of 3%.

A natural explanation for the correlation between handovers and duplicated costs is that incidents involving handovers also tend to have longer lengths of stay. In turn, longer stays may reflect both greater clinical need and more opportunities to perform additional tests and treatments. Consistent with this intuition, Column 3 of Table A10 shows a strong positive association between the number and share of duplicated cost items and length of stay. However, columns 4 and 5 show that, even after controlling for length of stay, incidents with handovers or with a higher number of doctors continue to have more duplicated items, both in absolute terms and as a share of total costs. In other words, patients who are passed along between different doctors undergo more repeated tests and treatments, *even conditional on the duration of their stay in the ED*. We interpret this pattern as evidence that the additional duplications associated with handovers likely represent unnecessary, and therefore wasteful, resource use.

In column 2 Table 5 we found above that contractors generate more handovers. We now find in columns 3 of Panel A Table 5 shows that contractors generate .17 more duplicated cost items than employees, relative to a mean of .74. Column 4 shows that, among contractors, duplicated costs accounts for 3.9% of total costs, compared with 3% among employees. We interpret these differences as suggestive evidence of a less efficient use of resources in incidents initially assigned to contractors, partly operating through their higher propensity to hand over patients to the incoming doctors in the next shift.

Externalities on Employees In this subsection, we investigate whether contractors are assigned fewer cases over the course of a shift. The hypothesis follows from two findings: triage nurses assign cases based on the existing doctor caseload, and contractors' higher average length of stay means they clear cases more slowly. Using a doctor/shift-level dataset, we find in Table 6 that contractors are assigned 8–18% fewer cases. Columns 2 and 3 reweight cases by predicted length of stay and cost, respectively, and yield similar estimates.

The finding that contractors impose a negative externality on co-assigned employees raises a further question: do employees work faster or slower when sharing a pod with a contractor? Table A11 examines this using a sample restricted to incidents assigned to employees: all cases assigned to contractors or students are dropped from the baseline dataset. The main independent variable is the number of contractors sharing the employee's pod during the shift. Since this variable varies only at the shift/pod level, shift/pod fixed effects cannot be included, and the estimates should be interpreted with caution given that case assignment to pods is not exogenous. That said, the coefficients are small in magnitude and statistically insignificant, which we take as reassuring. We find no evidence that employees adjust their behavior depending on whether they share a pod with contractors or other employees.

5 Mechanisms

In Section 4 we documented three main findings: (a) contractors use more resources than employees, (b) these higher inputs do not translate into better patient outcomes, and (c) contractors generate more handovers and duplicated tests, which at least partly reflect inefficiency in resource use.

In this section, we consider two broad explanations for these performance differences: (a) differences in ability and (b) differences in motivation. We argue that differences in motivation provide a more plausible account of the Section 4 evidence. In particular, we argue that employees are more willing to exert effort toward supporting their fellow employees, and that this channel partly explains their superior performance relative to contractors.

Ability and Human Capital A natural explanation for the performance gap between contractors and employees is that contractors may have lower diagnostic ability, requiring more time and resources to achieve comparable patient outcomes. However, three pieces of evidence contradict this interpretation. First, Panel D of Table 2 shows that *the same doctor* behaves differently when working as a contractor compared to when working as a permanent employee, which is inconsistent with time-invariant differences in ability. Second, Panel E of Table 2 demonstrates that the results are robust to controlling for doctor age and experience. Third, Figure 5 shows that doctors reduce their length of stay *discontinuously* when transitioning from contractor to employee status. The latter two findings suggest that differences in time-variant human capital are unlikely to fully explain the baseline findings. In particular, because human capital cannot be acquired overnight, it cannot account for the immediate behavioural change documented in Figure 5.

We conclude that, while differences in ability and human capital may matter in other settings, they do not appear to be the primary mechanism behind the resource use differences observed in our setting.

Overstay as a Proxy for Effort A natural alternative explanation is that the observed performance differences stem from differences in the *willingness* or *motivation* to perform.

For instance, employees may be more willing than contractors to exert effort in support of the hospital's broad objectives, such as maintaining sufficient throughput or minimising unnecessary diagnostic tests.

Direct measures of motivation or effort are not available in our setting. We can, however, examine an indirect measure of discretionary effort: the decision to stay beyond scheduled hours to discharge a patient rather than hand the case over to the next-shift doctor. In a setting where doctors are free to leave at the end of their shift and handovers are considered routine, remaining on duty to complete treatment represents a voluntary decision that entails a personal cost in terms of foregone leisure. While costly to the doctor, such behaviour will often benefit colleagues, the hospital and, in many cases, the patients themselves.

We define *overstay* as an indicator equal to one when a patient is discharged by the doctor originally assigned to the case, but in the sixty minutes after that doctor's shift has ended. This dummy captures well the idea that the original doctor stays doing unpaid overtime to avoid handing the patient over. In our data, 3.7% of all incidents are discharged in this way, a small but non-negligible share. By contrast, 36% of all incidents are handed over to a doctor in the next shift. The remaining 60.3% of incidents are discharged by the initially assigned doctor and before the shift ends.

The next subsection examines differences in overstay behaviour between employees and contractors. In the remainder of this subsection, we validate empirically the use of overstay as a proxy for discretionary effort. Specifically, we show that the likelihood of overstaying is higher among homophilous groups, that is groups of doctors who share greater similarity in their age or gender. This pattern is consistent with the well-documented finding that individuals are more willing to exert effort to help those with whom they share observable traits (Baccara and Yariv 2013, Battiston et al. 2021, Békés and Ottaviano 2025).

We construct a difference in gender variable as follows. First, we define (a) a male dummy for the doctor initially assigned to the incident and (b) the average male dummy for the doctors working in the same pod during the following shift.[29] Our variable is the absolute difference between (a) and (b), so higher values indicate greater gender dissimilarity between the outgoing doctor and the group of doctors who might receive the patient. We construct a difference in age variable in an analogous way.

Column 1 of Table 7 reports estimates of the effect of gender and age dissimilarity on

[29] We use the average male dummy for next-shift doctors because we do not observe which specific doctor will inherit the patient if the outgoing doctor does not overstay. To the extent that the outgoing doctor anticipates this information, our variable will contain non-negligible measurement error. This measurement error in the independent variable is likely classical and therefore will bias our estimates towards zero.

the likelihood of overstaying. We use our baseline empirical specification, which includes controls for the shift/pod combination and patient characteristics. In addition, we include fixed effects for the identity of the outgoing doctor. We do not include indicators for the set of incoming doctors in the same pod and next shift, as these are already absorbed by the shift/pod fixed effects. This specification therefore identifies whether *the same doctor* exhibits different willingness to overstay depending on her gender and age dissimilarity to the doctors who might otherwise inherit the patient.

We find large and statistically significant effects. The same doctor is .3 percentage points more likely to overstay if she has the same gender as the doctors replacing her. The same doctor is .5 percentage points more likely to overstay if she has the same age as the doctors replacing her, relative to being ten years apart. These magnitudes are substantial relative to the unconditional mean of 3.7%. Overall, we conclude that the empirical patterns support interpreting the overstay variable as a valid proxy for discretionary effort by the outgoing doctor.

Average Differences in Overstay between Employees and Contractors In this subsection, we examine whether employees and contractors differ in their willingness to overstay and avoid handing patients over to the next-shift doctor. We employ our baseline specification, which includes shift/pod fixed effects and controls for patient characteristics. For ease of interpretation, we reverse the sign of the main independent variable so that it is an employee indicator, with contractors serving as the base group.[30]

Column 2 of Table 7 shows that employees are more likely to overstay than contractors. This difference could, in principle, reflect either the causal effect of employment status or selection into that status. Column 3 addresses this distinction by exploiting within-doctor variation. We find that *the same doctor* is more likely to overstay in periods when permanently employed, relative to periods in which she works as a contractor. This finding is inconsistent with explanations based on time-invariant differences in altruism or cost of effort and instead supports a treatment effect interpretation of the employment status coefficient. Column 4 further shows that this difference remains robust after controlling for the homophily variables introduced in the previous subsection. The magnitude of the effect is economically meaningful: the coefficient of .012 represents around a third of the unconditional likelihood of overstay (i.e., .037).[31]

[30] A small number of incidents are assigned to medical students (referred to as 'residents' in the U.S., 'trainee doctors' in the UK, and 'specializzandi' in Italy). All regressions control for student status. Therefore, when the contractor dummy is included, employees form the base group, and vice versa.

[31] The finding that employees overstay more than contractors is particularly remarkable given that em-

Evidence on Overstaying to Help Other Employees Having established that employee status increases the willingness to overstay, we now further narrow down on the mechanism through which employment status affects overstay.

Identifying the specific factor that causes employees to overstay is challenging because employee status includes a wide range of characteristics. For instance, employees attend the monthly meetings where hospital-wide objectives and performance are discussed, while contractors are not. These meetings may help to communicate hospital-wide objectives and instil a sense of community among fellow employees. Second, employees and contractors differ in their average weekly hours worked, which could influence the marginal cost of exerting effort during overstay episodes (although see Tables A4 and A5). Third, the two groups may face different monetary or career incentives.[32] Fourth, contractors and employees differ in the exposure to liability in case of a malpractice suit. Employees are covered by the hospital insurance, while contractors are required to have their own private insurance. Finally, employees typically have longer expected tenures at the hospital than contractors, which may strengthen their attachment to both institutional objectives and to their colleagues.

An important distinction among these mechanisms is on whether they depend on the identity of the incoming doctors. Explanations based on differences in malpractice liability, incentives, the marginal cost of effort or attachment to institutional objectives are unrelated to the contractual status of the next-shift doctor. By contrast, explanations rooted in a willingness to help colleagues predict that the likelihood of overstaying will be higher when the incoming doctor is a fellow employee rather than a contractor on a short-term arrangement.

In Column 5 of Table 7, we interact the employee indicator with the average employment status of the doctors working in the same pod during the following shift. We find that employees are *not* more likely than contractors to overstay when the incoming doctors are (on average) not fellow employees. In contrast, employees are substantially more likely to overstay when the incoming doctors are also employees. Columns 6 and 7 show that these results remain robust after controlling for outgoing doctor fixed effects and for the homophily variables from the previous subsections.

---

ployees are more likely to have discharged the incident before the end of the shift. As a result, the number of incidents on which doctors can overstay is smaller for employees than contractors. The samples on which Table 7 is estimated include all incidents, and therefore do not account for this selection effect. In Table A12, we replicate Columns 1-4 Table 7 but using only the incidents that are not discharged by the end of the shift. As a result, the sample size is 120,994 as opposed to 306,504. The average value of overstay is 9.1% in this subsample. Unsurprisingly, we estimate much larger coefficients.

[32] As discussed in Section 2, both employees and contractors receive fixed wages independent of patient outcomes or the number of incidents discharged. However, implicit incentives may differ: employees face promotion opportunities, while contractors may be concerned with contract renewal.

In Table A13 we confirm that the higher willingness to overstay by employees is really due to the fact that the next shift/pod is assigned to employees, as opposed to other characteristics of the shift/pod that may be correlated with the arrival of employees. Specifically, we control incrementally for the interactions between the employee dummy and a large number of pod/shift characteristics. We find that the coefficient associated with the employee/incoming employees is remarkably robust to controlling for a large number of interactions.

Overall, the estimates indicate that the additional effort undertaken by employees is not a general behavioural difference but rather reflects a targeted willingness to assist colleagues who share their employment status within the organisation.[33]

Evidence on Lower Resource Use to Help Other Employees We now bring our investigation of the mechanism full circle. We investigate what share of the Table 2 estimates can be accounted for by the differential behaviour of employees when the next-shift doctors who may inherit their incidents are also employees. One channel through which employees may reduce length of stay and cost when followed by employees is by overstaying, as documented above. This is not the only channel, however. For instance, an employee who knows that she is not likely to discharge the patient before having to go home, might still want to exert as much cognitive effort as possible in diagnosis and treatment, so that the next incoming employee is presented with a patient who is ready for discharge.

In columns 1 and 2 Table A15, we reproduce the baseline estimates from Panels C and D Table 2. The estimates are identical in absolute value across the two tables, but they have the opposite sign depending on whether the independent variable is an employee dummy (Table A15) or a contractor dummy (Table 2).

In column 3 Table A15, we expand the column 1 specification by interacting the employee dummy with the share of doctors in the next shift and pod who are also employees. In columns 4 and 5, we progressively control for doctor fixed effects and the homophily variables. We find that employees use substantially less resources, especially when they are followed by incoming employees in the same shift and pod. For instance, relative to con-

[33]In Table 7, we calculate the type of the next shift doctors by weighing by the number of cases assigned, to account for the probability that a case that is handed over ends up being inherited by a specific doctor type. In Panels A and B of Table A14, we find similar results when using the unweighted average or the type of the next shift doctor that handles the most cases. In Panel C, we use a 2-hour window instead of an 1-hour window to calculate the overstay variable. Occasionally, we find doctors taking their first case a few minutes before the end of a shift. We suspect that these observations capture doctors who started their shift before the official start time. In Panel D, we drop these observations from the regression. We find very similar results.

tractors, length of stay is 23% lower for employees who are followed by contractors, but 32% lower for employees who are followed by other employees. Similarly, relative to contractors, ED cost is 5.7% lower for employees who are followed by contractors, but 9.4% lower for employees who are followed by other employees.[34]

Overall, we conclude that part of the explanation why employees use less resources than contractors is that they want to help the fellow employees who will inherit their cases.

Relational Contracts The finding that employees overstay more when followed by other employees raises the possibility that this reflects a relational contract in which employees exchange favours with one another but not with contractors. We investigate this possibility by asking whether the structure of doctor rotations at Niguarda is consistent with sustaining such an arrangement through the overstay decision.

In the idealised case, two employees A and B would exchange favours by alternating who follows whom: A follows B in half their shared shifts, and B follows A in the other half. Figure A5 and A6 show that actual rotation patterns fall well short of this. For each incident, we identify the doctor in the next shift and pod and compute three overlap measures over the prior year: the forward overlap (the incoming doctor has previously followed the focal doctor), the backward overlap (the focal doctor has previously followed the incoming doctor), and the horizontal overlap (the two doctors have worked the same shift). We do this both for the previous year (Figure A5) and for the following year (Figure A6). The forward and backward overlap means are too small to sustain reciprocal arrangements: the probability that the incoming doctor has previously occupied the next shift behind the focal doctor is just 5%. Even the horizontal overlap, which is computed over a larger pool of co-workers, averages only 16%.

If relational contracts drove the employee–employee overstay pattern, we would expect overlap to be systematically higher among employee pairs than employee–contractor pairs. Tables A16–A18 show this is not the case. Overlap measures are no higher for employee pairs (Table A16), do not predict the overstay decision (Table A17), and controlling for them leaves the employee $\times$ incoming-employee coefficient in Table 7 virtually unchanged (Table A18).

We therefore conclude that reciprocal favour exchange does not explain the differential overstay behaviour documented in Table 7.

[34]The interactions in the within-doctor regressions of columns 4 and 5 are not statistically significant, although the interactions in the across-doctor regressions of column 3 are.

Identity, Insiders and Employment Status Akerlof and Kranton (2010) argue that organisations structure workers into groups, and that these groups differ in the extent to which individuals internalise organisational norms and a sense of community within their identities. Chen and Li (2009) find in a laboratory experiment that inducing group identity leads subjects to take more co-operative decisions when matched with a member of their own group. This framework provides a useful lens through which to interpret the results in this section. We find that, upon becoming an employee, the same individual becomes more likely both to advance the hospital's objectives (by reducing resource use without worsening patient outcomes) and to support her fellow employees (by overstaying more frequently). A natural interpretation is that employment status shapes social identity, casting contractors as outsiders and employees as insiders. Although identical in their clinical roles and tasks in the ED, this socialisation process may arise from the differential integration into the hospital's organisational culture documented in Section 2. These differing identities, in turn, lead them to behave differently depending on the identity of the doctor who will inherit their cases.

To be clear, we interpret our results as consistent with the mechanism above as opposed to conclusively generated by the mechanism above. It is possible, for instance, that employees want to gain a reputation for being cooperative among their fellow employees, in the expectation that this reputation later affects potential promotion decisions (Battiston et al., 2025). Alternatively, employees may use the overstay decision to sustain the exchange of favours that partly take place in other settings. In our data, we do not have a variable called 'identity' and neither are we able to manipulate it in an experimental setting. Nevertheless, the finding that employees behave differently towards the members of their own contractual group allows us to rule out a wide range of explanations for the differential overstay between employees and contractors.

6 External Validity and Concluding Remarks

We have shown that the productivity differences between employees and contractors are economically sizeable across a wide range of measures and behaviours. We have argued that the identity-based framework of Akerlof and Kranton (2005) offers a useful lense through which to interpret these findings. In this framework, permanent employees can be viewed as insiders who are willing to exert effort on behalf of both the organisation and other insiders. By contrast, the fact that contractors use more resources and are less willing to undertake unpaid overtime suggest that they perceive themselves as outsiders.

The richness of our analysis has been possible thanks to our focus on a single organisation, for which we have detailed administrative data and a deep understanding of the production technology and institutional setting. This granular perspective provides clear advantages in terms of causal identification and the breadth of insights that can be generated. At the same time, our 'insider econometrics' approach (Bandiera et al. 2011, Ichniowski and Shaw 2013) naturally raises questions about the extent to which our findings generalise beyond this setting. This is particularly relevant because our across-doctor and (typically identical) within-doctor results are based on a relatively small number of individuals (55 and 14, respectively). We discuss external validity in the remainder of this section.

The absence of meaningful selection effects indicates that, in our setting, doctors operating under a contractual status are not negatively selected. This is likely a consequence of the specific institutional environment that produced the rise and subsequent decline of contractor doctors in Italy.[35] Therefore, we do not believe that the findings of negligible selection will apply to other organisations. In other settings, independent contractors may be negatively (or perhaps even positively) selected.

We view the treatment effects we estimate as a 'proof of concept' that contractor status can affect productivity, independently of the working conditions typically bundled with it (Mas and Pallais 2020). Permanent public sector employees in Italy enjoy guaranteed lifetime employment and a strong esprit de corps, especially in prestigious hospitals like Niguarda. This may amplify their willingness to support their institution and colleagues and therefore enlarge the treatment effects we observe.[36] At the same time, it is noteworthy that these differences persist even in a context where contractors are not inherently less able and are often on the cusp of transitioning into permanent employment.

Because contractual status is a bundle of features, the specific mechanism that we document here may operate differently in other settings. In some environments, for example, employees and contractors may face much sharper differences in their explicit or implicit incentives. Likewise, employee status may not generate a strong sense of organisational identity in contexts where mission-oriented objectives are weaker (Besley and Ghatak, 2005). A useful comparison is therefore across settings that vary in the salience of mission, incentives, and institutional cohesion.

Ultimately, the extent to which our results replicate in other contexts is inherently difficult to establish, both in this paper and in the broader literature. Our aim has been

[35] Specifically, the Italian central government imposed a freeze on the hiring of public-sector employees between 2010 and 2018. Many able doctors who would otherwise have qualified for permanent employment were therefore forced to take independent contractor positions.

[36] Conversely, basic incentive theory predicts that workers will exert less effort when their continued employment is guaranteed (Shapiro and Stiglitz 1984).

to provide an initial look at this question and demonstrate the type of insights that can be obtained when detailed information on employment status and productivity is combined with a credible identification strategy. We leave to future research the task of examining whether similar effects and mechanisms arise in other organisations and institutional environments.

REFERENCES

Abraham, K. G., Hershbein, B., Houseman, S. N., and Truesdale, B. C. (2024), "The Independent Contractor Workforce: New Evidence on Its Size and Composition and Ways to Improve Its Measurement in Household Surveys", *ILR Review*, 77(3): 336–365.

Akerlof, G. A., and Kranton, R. E. (2005), "Identity and the Economics of Organizations", *Journal of Economic Perspectives*, 19(1): 9–32.

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005), "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools", *Journal of Political Economy*, 113(1): 151–184.

Ameriks, J., Briggs, J., Caplin, A., Lee, M., Shapiro, M. D., and Tonetti, C. (2020), "Older Americans Would Work Longer If Jobs Were Flexible", *American Economic Journal: Macroeconomics*, 12(1): 174–209.

ANAAO Assomed (2025, March 4), "Dalla carenza di specialisti alla pletora: entro il 2032 60 mila medici in cerca di lavoro. Subito un piano assunzioni.", *ANAAO Assomed–Area Stampa – Studi/Indagini ANAAO*, retrieved from `https://www.anaao.it/content.php?cont=42365`

Apker, J., Mallak, L. A., and Gibson, S. C. (2007), "Communicating in the Gray Zone: Perceptions About Emergency Physician–Hospitalist Handoffs and Patient Safety", *Academic Emergency Medicine*, 14(10): 884–894.

Baccara, M., and Yariv, L. (2013), "Homophily in Peer Groups", *American Economic Journal: Microeconomics*, 5(3): 69–96.

Bandiera, O., Barankay, I., and Rasul, I. (2005), "Social Preferences and the Response to Incentives: Evidence from Personnel Data", *Quarterly Journal of Economics*, 120(3): 917–962.

Bandiera, O., Barankay, I., and Rasul, I. (2011), "Field Experiments with Firms", *Journal of Economic Perspectives*, 25(3): 63–82.

Bandiera, O., Barankay, I., and Rasul, I. (2013), "Team Incentives: Evidence from a Firm-Level Experiment", *Journal of the European Economic Association*, 11(5): 1079–1104.

Battiston, D., Blanes i Vidal, J., and Kirchmaier, T. (2021), "Face-to-Face Communication in Organizations", *Review of Economic Studies*, 88(2): 574–610.

Békés, G., and Ottaviano, G. I. (2025), "Cultural Homophily and Collaboration in Superstar Teams", *Management Science*, forthcoming.

Bernhardt, A., Campos, C., Prohofsky, A., Ramesh, A., and Rothstein, J.

(2023), "Independent Contracting, Self-Employment, and Gig Work: Evidence from California Tax Data", *ILR Review*, 76(2): 357–386.

Besley, T., and Ghatak, M. (2005), "Competition and Incentives with Motivated Agents", *American Economic Review*, 95(3): 616–636.

Bewley, T. F. (1999), "Why Wages Don't Fall During a Recession", Harvard University Press, Cambridge, MA.

Bloom, N., Liang, J., Roberts, J., and Ying, Z. J. (2015), "Does Working from Home Work? Evidence from a Chinese Experiment", *Quarterly Journal of Economics*, 130(1): 165–218.

Boltz, M., Ederer, K., Fuchs, J., and Hoelzl, M. (2023), "How Does Working-Time Flexibility Affect Workers' Productivity in a Routine Job? Evidence from a Field Experiment", *British Journal of Industrial Relations*, 61(3): 663–701.

Cappellari, L., Dell'Aringa, C., and Leonardi, M. (2012), "Temporary Employment, Job Flows and Productivity: A Tale of Two Reforms", *Economic Journal*, 122(562): F188–F215.

Castillo, J. C. (2025), "Who Benefits from Surge Pricing?", *Econometrica*, 93(5): 1811–1854.

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

Chan, D. C. (2018), "The Efficiency of Slacking Off: Evidence from the Emergency Department", *Econometrica*, 86(3): 997–1030.

Chan, D. C., and Chen, S. (2022), "The Productivity of Professions: Evidence from the Emergency Department", *NBER Working Paper* 30608.

Chen, M. K., Rossi, P. E., Chevalier, J. A., and Oehlsen, E. (2019), "The Value of Flexible Work: Evidence from Uber Drivers", *Journal of Political Economy*, 127(6): 2735–2794.

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

Chen, Y., and Li, S. X. (2009), "Group Identity and Social Preferences", *American Economic Review*, 99(1): 431–457.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014), "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood", *American Economic Review*, 104(9): 2633–2679.

Cohen, P., Hahn, R., Hall, J., Levitt, S., and Metcalfe, R. (2016), "Using Big Data to Estimate Consumer Surplus: The Case of Uber", *Working Paper*.

Costa, D. L., and Kahn, M. E. (2003), "Cowards and Heroes: Group Loyalty in the American Civil War", *Quarterly Journal of Economics*, 118(2): 519–548.

Delfgaauw, J., Dur, R., Nijkamp, R., and van den Bossche, S. (2022), "Team Incentives, Task Allocation, and Performance: Evidence from a Field Experiment", *Management Science*, 68(11): 7961–7979.

Doyle, J. J. (2011), "Returns to Local-Area Healthcare Spending: Evidence from Health Shocks to Patients Far from Home", *American Economic Journal: Applied Economics*, 3(3): 221–243.

Doyle, J. J., Ewer, S. M., and Wagner, T. H. (2010), "Returns to Physician Human Capital: Evidence from Patients Randomised to Physician Teams", *Journal of Health Economics*, 29(6): 866–882.

Doyle, J., Nightingale, M., Dawney, J., Whitmore, M., van Stolk, C., and Hofman, J. (2024), "Effective contracting of employment and health services: Evidence review", *RAND Europe*, Cambridge, UK.

Drenik, A., Jäger, S., Plotkin, P., and Schoefer, B. (2023), "Paying Outsourced Labor: Direct Evidence from Linked Temp Agency–Worker–Client Data", *Review of Economics and Statistics*, 105(1): 206–216.

Dube, A., Jacobs, J., Naidu, S., and Suri, S. (2020), "Monopsony in Online Labor Markets", *American Economic Review: Insights*, 2(1): 33–46.

Eeckhout, J., and Kircher, P. (2011), "Identifying Sorting — in Theory", *Review of Economic Studies*, 78(3): 872–906.

Emanuel, N., and Harrington, E. (2024), "The Power of Proximity to Coworkers: Training for Tomorrow or Productivity Today?", *American Economic Journal: Applied Economics*, 16(3): 154–190.

Frakes, M. D., and Gruber, J. (2019), "Defensive Medicine: Evidence from Military Immunity", *American Economic Journal: Economic Policy*, 11(3): 197–231.

Gallup (2017), "Gallup's Perspective on the Gig Economy and Alternative Work Arrangements", Gallup, Washington, DC.

Gibbons, R., and Roberts, J. (2013), "Economic Theories of Incentives in Organizations", in *The Handbook of Organizational Economics*, Princeton University Press.

Gibbs, M., Mengel, F., and Siemroth, C. (2023), "Work from Home & Productivity: Evidence from Personnel & Analytics Data on IT Professionals", *Journal of Political Economy: Microeconomics*, 1(1): 1–45.

Goldschmidt, D., and Schmieder, J. F. (2017), "The Rise of Domestic Outsourcing and the Evolution of the German Wage Structure", *Quarterly Journal of Economics*, 132(3):

1165–1217.

He, H., Neumark, D., and Weng, Q. (2021), "Do Workers Value Flexible Jobs? A Field Experiment", *Journal of Labor Economics*, 39(3): 709–738.

Hirsch, B., and Mueller, S. (2012), "The Productivity Effect of Temporary Agency Work", *Economic Journal*, 122(562): F216–F243.

Holmström, B. (1999), "Managerial Incentive Problems: A Dynamic Perspective", *Review of Economic Studies*, 66(1): 169–182.

Holmström, B., and Milgrom, P. (1994), "The Firm as an Incentive System", *American Economic Review*, 84(4): 972–991.

Ichniowski, C., and Shaw, K. (2013), "Insider Econometrics: Empirical Studies of How Management Matters", in Gibbons, R. and Roberts, J. (eds.), *The Handbook of Organizational Economics*, Princeton University Press.

Jones, S., Wells, L., Harper, R., et al. (2022), "Association Between Delays to Patient Admission from the Emergency Department and All-Cause 30-Day Mortality", *Emergency Medicine Journal*, 39(3): 168–173.

Kamat, A. B., Midgley, S., and Kimbrell, K. (2015), "Duplication of Radiology Imaging Studies in the Emergency Department: What Is the Cost?", *Emergency Medicine Journal*, 32(2): 144–148. *(Online first 2013.)*

Kandel, E., and Lazear, E. P. (1992), "Peer Pressure and Partnerships", *Journal of Political Economy*, 100(4): 801–817.

Kayser, A. (2025), "World's Best Hospitals 2025 – Top 250", *Newsweek*, 26 February.

Katz, L. F., and Krueger, A. B. (2019), "The Rise and Nature of Alternative Work Arrangements in the United States, 1995–2015", *ILR Review*, 72(2): 382–416.

Lazear, E. P. (2000), "Performance Pay and Productivity", *American Economic Review*, 90(5): 1346–1361.

Lim, K., Miller, A., Risch, M., and Wilking, E. (2019), "Independent Contractors in the US: New Trends from 15 Years of Administrative Tax Data", *Unpublished Working Paper*, available at `https://www.irs.gov/pub/irs-soi/19rpindcontractorinus.pdf`

Mas, A., and Pallais, A. (2017), "Valuing Alternative Work Arrangements", *American Economic Review*, 107(12): 3722–3759.

Mas, A., and Pallais, A. (2020), "Alternative Work Arrangements", *Annual Review of Economics*, 12: 631–658.

Maestas, N., Mullen, K. J., Powell, D., von Wachter, T., and Wenger, J. (2023), "The Value of Working Conditions in the United States and Norway", *Review of Economics and Statistics*, 105(2): 385–402.

Ministero della Salute (2023), “Il reclutamento del personale dirigente del Servizio Sanitario Nazionale”, Rome: Ministero della Salute.

Ministero della Salute (2024), “Linee di indirizzo nazionali per lo sviluppo del Piano di gestione del sovraffollamento in Pronto Soccorso”, Rome: Ministero della Salute.

Nyce, A., Gandhi, S., Freeze, B., Bosire, J., Ricca, T., Kupersmith, E., and Rachoin, J. S. (2021), “Association of Emergency Department Waiting Times with Patient Experience in Admitted and Discharged Patients”, *Journal of Patient Experience*, 8: 1–7.

Ozgen, C. (2021), “The Economics of Diversity: Innovation, Productivity and the Labour Market”, *Journal of Economic Surveys*, 35(4): 1168–1216.

Palagashvili, L., and Sharfuddin, R. (2024), “A Fresh Look at the Independent Workforce with New BLS Data”, *Labor Market Matters*, 26 November. Available at: `https://www.labormarketmatters.com/p/a-fresh-look-at-the-independent-workforce`.

Phillips, A. W., Malamet, P., and Williams, S. R. (2015), “Standardized Sign-Outs: An Opportunity to Improve Patient Safety”, *Emergency Medicine*, 47(3): 125–130.

Rotemberg, J. J. (1994), “Human Relations in the Workplace”, *Journal of Political Economy*, 102(4): 684–717.

Schwab, S. D., and Singh, M. (2024), “How Power Shapes Behaviour: Evidence from Physicians”, *Science*, 384(6697): 802–808.

Shapiro, C., and Stiglitz, J. E. (1984), “Equilibrium Unemployment as a Worker Discipline Device”, *American Economic Review*, 74(3): 433–444.

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

Smalley, C. M., Simon, E. L., Meldon, S. W., Muir, M. R., Briskin, I., Crane, S., . . . and Fertel, B. S. (2020), “The Impact of Hospital Boarding on the Emergency Department Waiting Room”, *JACEP Open*, 1(5): 1052–1059.

Stanton, C. T., and Thomas, C. (2025), “Who Benefits from Online Gig Economy Platforms?”, *American Economic Review*, 115(6): 1857–1895.

Thompson, D. A., Yarnold, P. R., Williams, D. R., and Adams, S. L. (1996), “Effects of Actual Waiting Time, Perceived Waiting Time, Information Delivery, and Expressive Quality on Patient Satisfaction in the Emergency Department”, *Annals of Emergency Medicine*, 28(6): 657–665.

TABLES

TABLE 1
SUMMARY STATISTICS

	All	Employees	Contractors	Students
	(1)	(2)	(3)	(4)
Panel A: Averages in Baseline Sample				
Doctor Characteristics:				
Doctor Age (years)	41	42	39	30
Doctor Male Dummy	.53	.53	.53	.52
Outcomes:				
ED (Total) Cost (€)	163	160	164	219
Length of Stay (minutes)	288	283	299	364
Hospitalisation Dummy	.19	.19	.19	.19
Return to ED 30 Dummy	.13	.13	.13	.13
Death Dummy	.0009	.0010	.0006	.0000
Cost of Laboratory Tests (€)	46	45	48	60
Other Costs (€)	113	111	114	149
Specialist Consultation Dummy	.35	.35	.30	.58
Handover to Next Shift Dummy	.36	.35	.37	.51
Duplicated Costs Items (units)	.74	.73	.82	.89
Share Duplicated Costs	.03	.03	.03	.03
Overstay Dummy	.037	.040	.021	.011
Panel B: Other Statistics				
Total Number of Doctors	85	52	18	30
Total Number of Incidents	306,635	262,437	30,491	13,707
Average Number of Incidents Per Doctor	3,066	5,047	1,694	457
Average Number of Shifts Per Doctor	500	799	288	109
Average Number of Hours Worked (per month)	93.7	92.8	99.1	97.8
Average Number of Days Worked (per month)	13.2	13.0	14.5	13.7
Average Number of Incidents per Hour	.75	.76	.71	.58

Panel A displays means of selected variables in the estimating baseline sample, separately for all doctors combined, employees, contractors and students. Panel B displays other selected statistics. 14 doctors switched throughout our sample period from contractors to employees, and one doctor switched from student to employee (see Table A1). As a result, the total number of doctors in the sample is 85=52+18+30-14-1.

TABLE 2
BASELINE RESULTS

	ED Cost	Length of Stay
Panel A: Controlling Only for Shift/Pod		
Contractor	.073***	.268***
	(.020)	(.036)
Panel B: Adding Patient Characteristics		
Contractor	.073***	.263***
	(.016)	(.039)
Panel C: Adding Exact Hour		
Contractor	.100***	.313***
	(.018)	(.048)
Panel D: Adding Doctor Fixed Effects		
Contractor	.080***	.297***
	(.019)	(.034)
Panel E: Adding Doctor Characteristics		
Contractor	.085***	.207***
	(.020)	(.038)
Log Doctor Age	-.671	-3.029*
	(.849)	(1.653)
Log Doctor Past Cases	-.008	-.027*
	(.008)	(.014)
Log Doctor Tenure at Niguarda	.023**	.000
	(.010)	(.015)

This table displays regressions of length of stay and total ED cost on contractor status. Cost and length of stay are in logs. All regressions control for whether the assigned doctor is a student, and for the interaction of shift (i.e. the morning, afternoon or night shift within a specific date) and pod. The patient characteristics added in Panels B-E are age, gender, triage, main complaint, mode of arrival, referral party and insurance status. The exact hour added in Panels C-E is the interaction of year, month, day of month and hour of day. Doctor past cases is the number of cases in the previous two years. Tenure at Niguarda is the number of days since the doctor first appeared the sample. Both variables are left censored, in which case we give the variable a value of zero and it is observed by the doctor fixed effect. Standard errors two-way clustered at the doctor and shift/pod interaction level. The number of observations is 306,635.

TABLE 3
OTHER PATIENT OUTCOMES

	Hospitalisation	Death in ED	Return to ED 2	Return to ED 5	Return to ED 30	Return to ED 60
Panel A: With Doctor Fixed Effects but Without Doctor Characteristics						
Contractor	.0150**	-.0003	.0001	-.0005	.0002	.0000
	(.0067)	(.0006)	(.0032)	(.0045)	(.0061)	(.0069)
Panel B: Adding Doctor Characteristics						
Contractor	.0007	.0002	.0029	.0056	.0029	.0033
	(.0078)	(.0006)	(.0039)	(.0055)	(.0071)	(.0083)
Log Doctor Age	-.8850***	.0135	.0216	.2323	.1428	.1255
	(.2764)	(.0128)	(.1049)	(.1460)	(.1904)	(.1824)
Log Doctor Past Cases	.0020	.0001	.0025*	.0006	.0022	.0012
	(.0036)	(.0002)	(.0014)	(.0019)	(.0025)	(.0031)
Log Doctor Tenure at Niguarda	.0015	.0001	-.0001	.0006	-.0022	-.0004
	(.0040)	(.0002)	(.0015)	(.0021)	(.0030)	(.0031)
Mean Variable	.1877	.0009	.0363	.0649	.1327	.1740

All regressions control for whether the assigned doctor is a student, the interaction of shift (i.e. the morning, afternoon or night shift within a specific date) and pod, exact hour and patient characteristics (age, gender, triage, main complaint, mode of arrival, referral party and insurance status). Hospitalisation is a dummy for whether the patient is admitted to hospital as opposed to discharged home. Death in ED is a dummy variable for whether the patient dies in the Emergency Department. Return to ED 2 is a dummy variable for whether the patient returns to the Emergency Department generating a different incident within two days. Return to ED 5, 30 and 60 are generated equivalently. Doctor past cases is the number of cases in the previous two years. Tenure at Niguarda is the number of days since the doctor first appeared the sample. Both variables are left censored, in which case we give the variable a value of zero and it is observed by the doctor fixed effect. Standard errors two-way clustered at the doctor and shift/pod interaction level. The number of observations is 306,540.

TABLE 4
EFFECTS ON SELECTED COST ITEMS

	COST TYPES		SELECTED TESTS				
	Cost of Laboratory Tests	**Other Costs**	**Blood Test**	**X-Ray**	**CT Scan**	**Electro-Cardiogram**	**Ultra-Sound**
Panel A: Not Controlling for Doctor Fixed Effects							
Contractor	.437***	.091***	.058***	.098***	.01	.023*	.013**
	(.063)	(.018)	(.007)	(.014)	(.007)	(.013)	(.006)
Panel B: Adding Doctor Fixed Effects							
Contractor	.355***	.073***	.04***	.066***	.003	.037**	.016
	(.077)	(.023)	(.009)	(.017)	(.011)	(.017)	(.013)
Mean DV	Logs	Logs	.59	.63	.26	.28	.1

The cost of laboratory tests and the other costs variables are in logs. All regressions control for whether the assigned doctor is a student, the interaction of shift (i.e. the morning, afternoon or night shift within a specific date) and pod, exact hour and patient characteristics (age, gender, triage, main complaint, mode of arrival, referral party and insurance status). Standard errors two-way clustered at the doctor and shift/pod interaction level. The number of observations is 306,540.

TABLE 5
EFFECTS ON CONSULTATIONS, HANDOVERS AND DUPLICATED COSTS

	(1) Required Specialist Consultation	(2) Handover to Next Shift	(3) Total Duplicated Costs	(4) Share Duplicated Costs
Panel A: Not Controlling for Doctor Fixed Effects				
Contractor	.013***	.073***	.167***	.006***
	(.005)	(.011)	(.04)	(.001)
Panel B: Adding Doctor Fixed Effects				
Contractor	.01	.075***	.219***	.009***
	(.009)	(.011)	(.051)	(.002)
Mean Dep.	.35	.36	.74	.03

Standard errors two-way clustered at the doctor and shift/pod interaction level. All regressions control for whether the assigned doctor is the student, the interaction of shift (i.e. the morning, afternoon or night shift within a specific date) and pod, exact hour and patient characteristics (age, gender, triage, main complaint, mode of arrival, referral party and insurance status). The number of observations is 306,540.

TABLE 6
EFFECTS ON NUMBER OF CASES DURING THE SHIFT

	(1) Unweighted	(2) Weighted by LOS	(3) Weighted by ED Cost
Panel A: Controlling Only for Shift/Pod			
Contractor	-.079***	-.102***	-.104***
	(.025)	(.029)	(.028)
Panel B: Adding Doctors Fixed Effects			
Contractor	-.173***	-.210***	-.206***
	(.020)	(.025)	(.025)

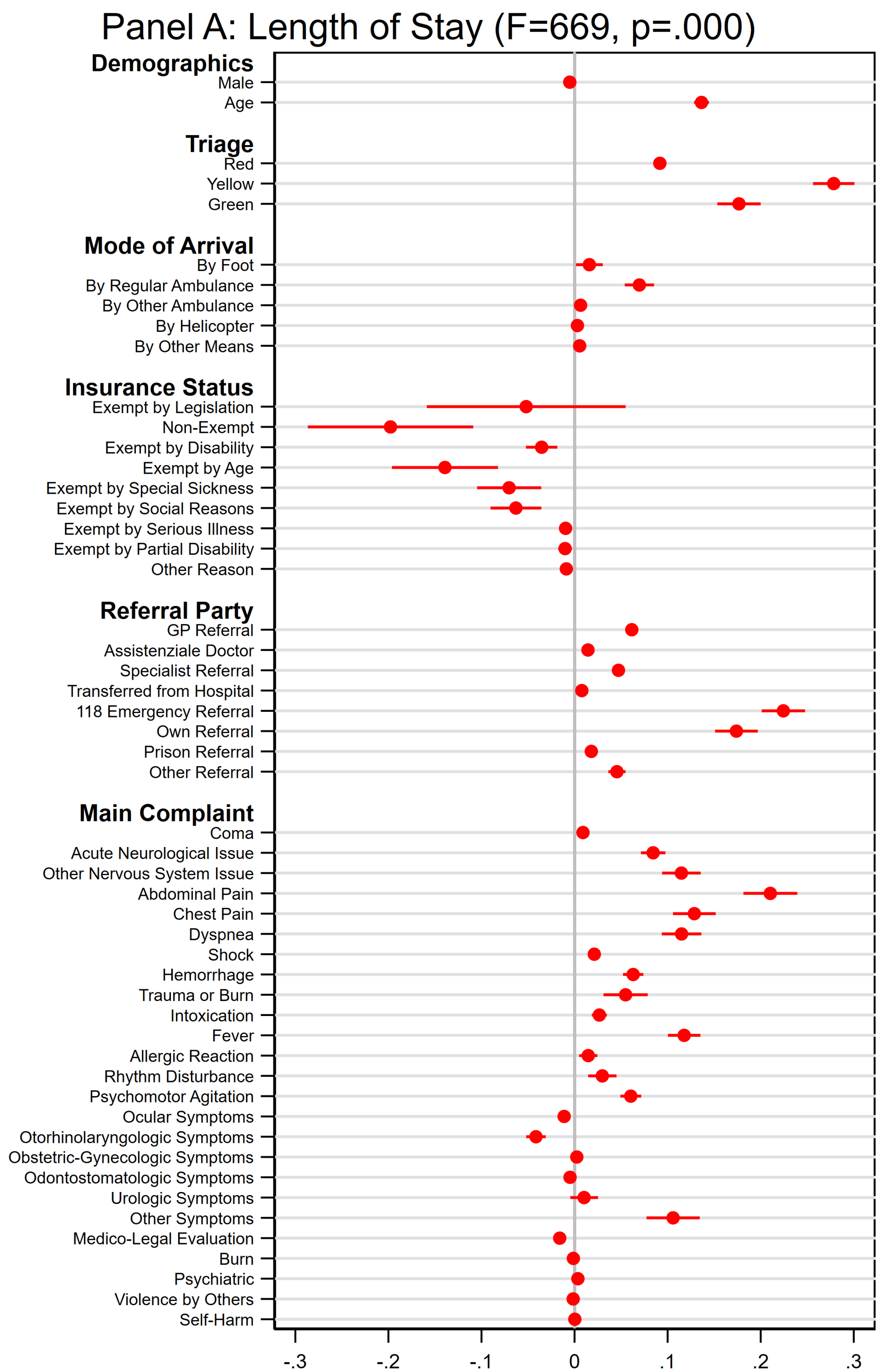
This table displays regressions of the number of incidents assigned to a doctor during a shift, on whether the doctor is a contractor. The sample is a sample of shift/pods and doctors. In the first column the dependent variable is the number of incidents assigned. In the second and third columns, the cases assigned to a doctor are weighted by the predicted length of stay and ED cost, respectively. The predictions are on the basis of case characteristics. All dependent variables are in logs. Standard errors are two-way clustered at the shift/pod and doctor level. The number of observations is 82,115.

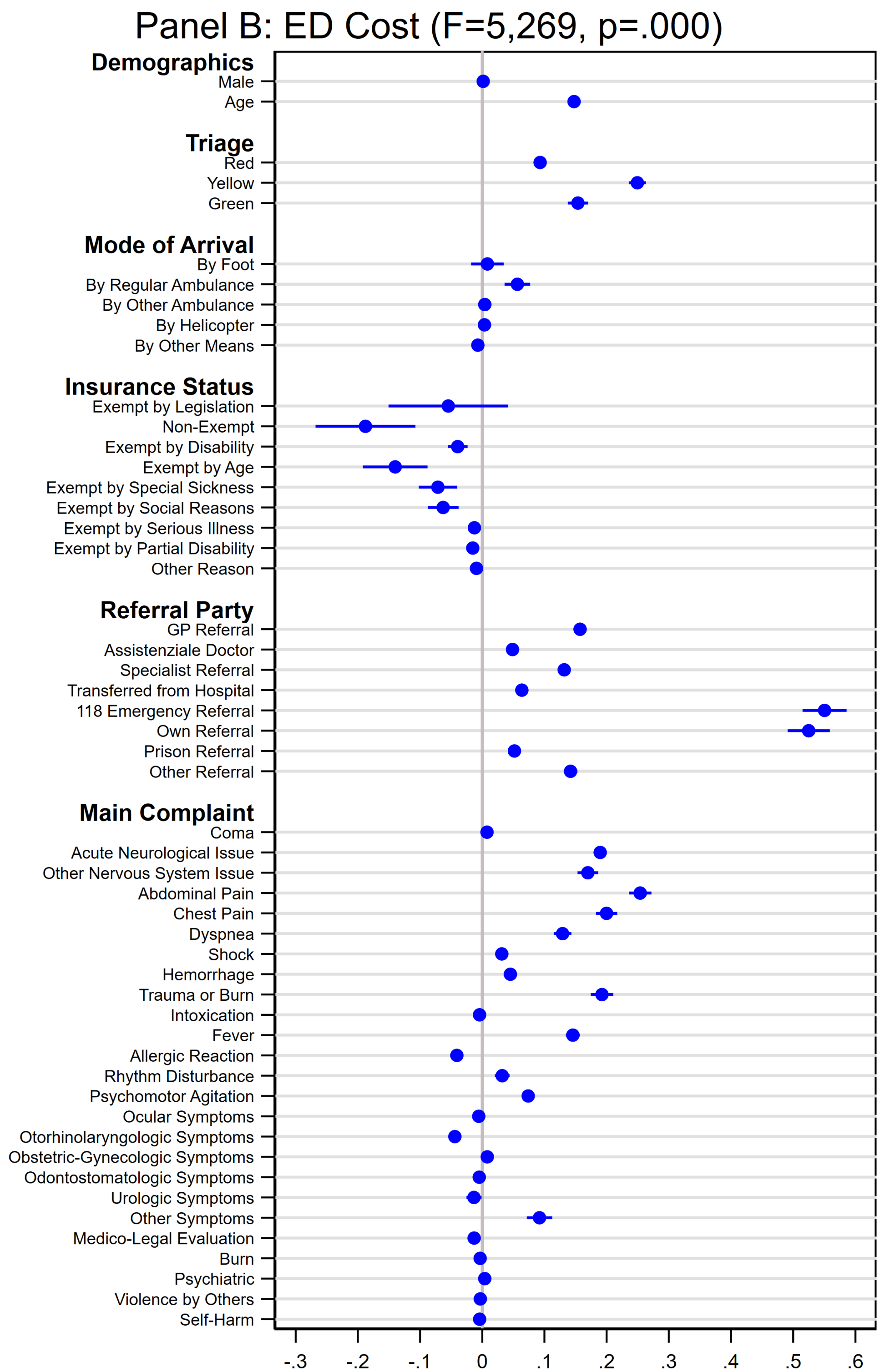
TABLE 7
MECHANISM: EFFECTS ON OVERSTAY VARIABLE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Employee		.019***	.014***	.012***	.005	.000	.001
		(.003)	(.005)	(.004)	(.006)	(.006)	(.006)
Employee × Incoming Employees					.016***	.016***	.013**
					(.006)	(.006)	(.006)
Difference in Gender	-.003*			-.003*			-.003*
	(.002)			(.002)			(.002)
Difference in Age / 10	-.005***			-.004***			-.004***
	(.002)			(.002)			(.001)
P-value 1+2					.000	.002	.005
Doctor Fixed Effects	Yes	No	Yes	Yes	No	Yes	Yes

Standard errors two-way clustered at the doctor and shift/pod level. Overstay takes value one when a patient is discharged by the doctor originally assigned to the case, but in the sixty minutes after that doctor's shift has ended. Difference in gender is the absolute value of the difference between a male dummy for the doctor initially assigned to the incident, and the average male dummy for the doctors working in the same pod and during the following shift. Difference in age variable is defined in an analogous way. Incoming Employee is the average of the employee dummy for the doctors working in the same pod and during the following shift. All regressions control for the interaction of shift (i.e. the morning, afternoon or night shift within a specific date) and pod, and patient characteristics (age, gender, triage, main complaint, mode of arrival, referral party and insurance status). The regressions in Columns 2-7 control for whether the assigned doctor is a student. The mean of the dependent variable Overstay is .037. The number of observations is 306,504.

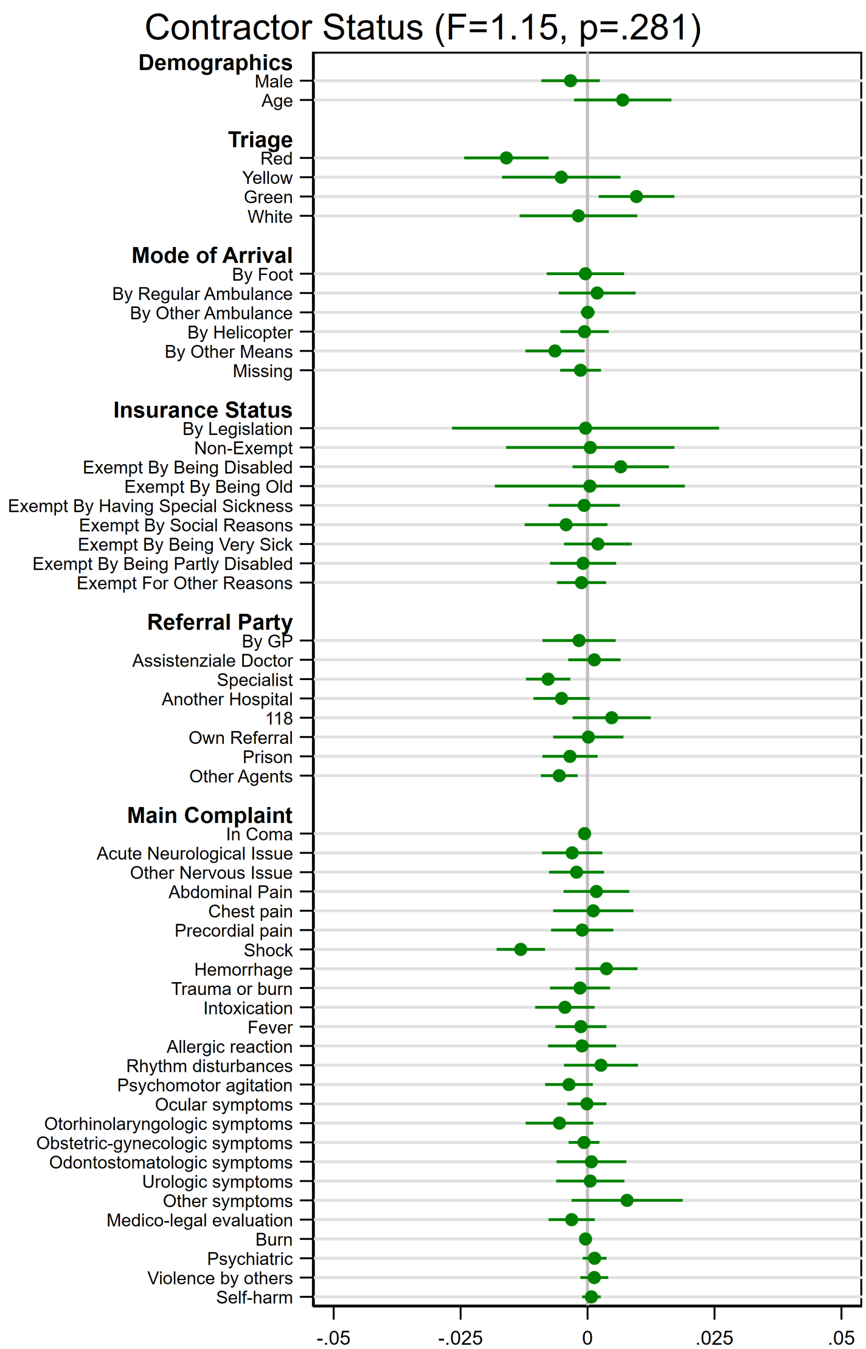
FIGURE 1: PREDICTIVE POWER OF PATIENT CHARACTERISTICS





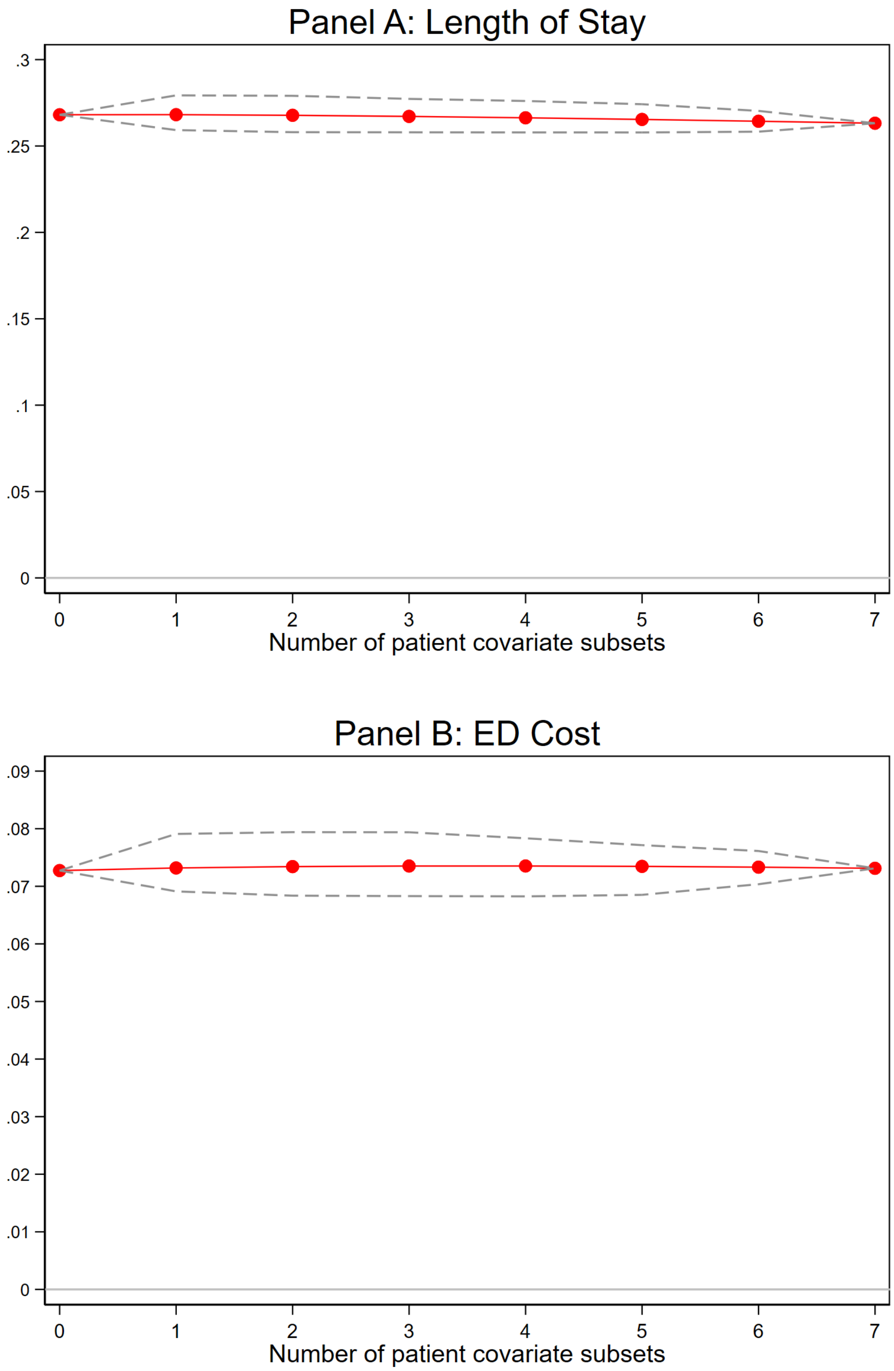
This figure displays coefficients and 95% confidence intervals from regressions of length of stay and ED cost on patient characteristics, controlling for the shift/pod indicators. All variables are standardised. Standard errors are clustered two-way at the doctor and shift/pod level. The F-statistics are from tests that all the patient characteristics are jointly equal to zero. p are the p-values of the corresponding tests.

FIGURE 2: BALANCE TESTS



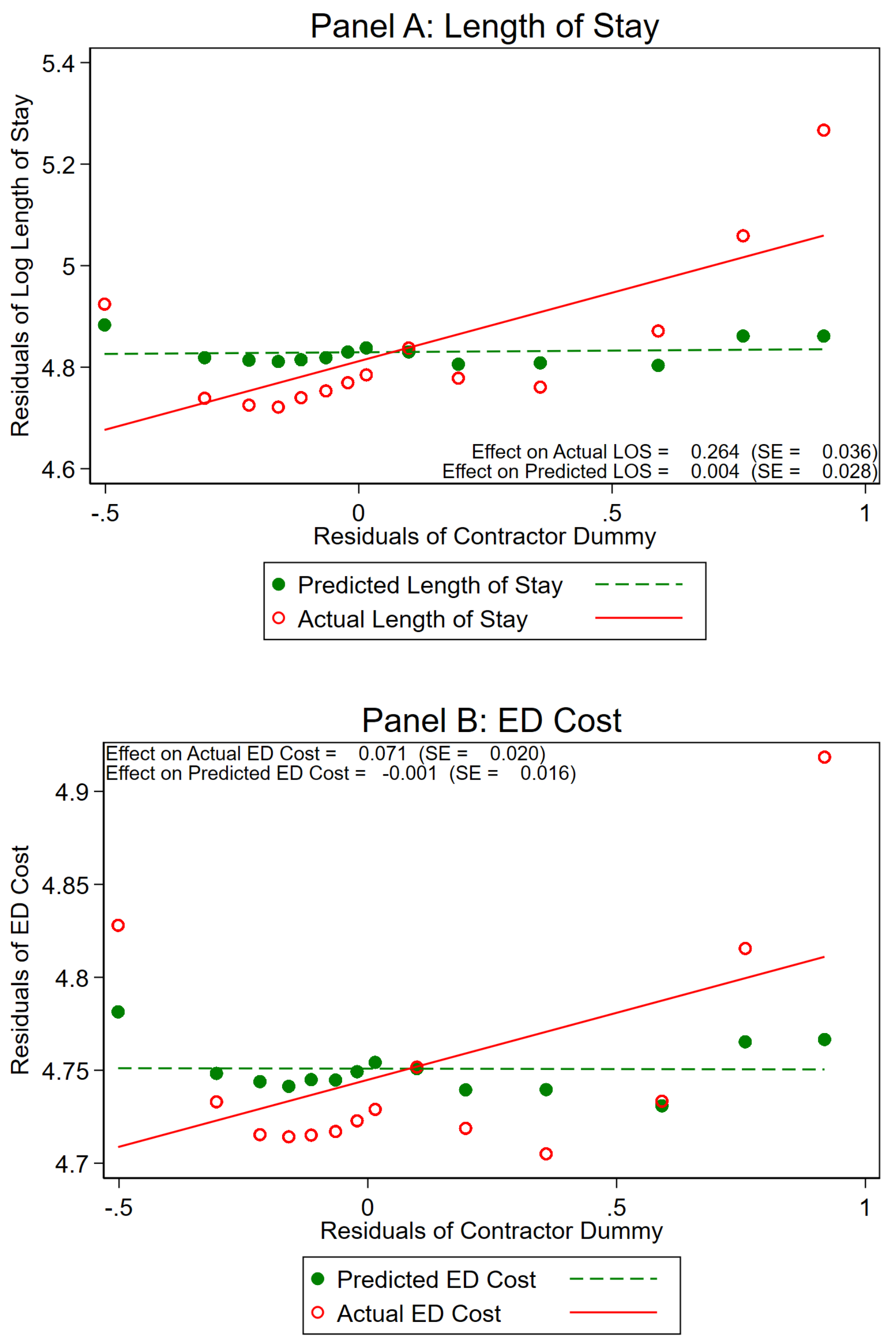
This figure displays coefficients and 95% confidence intervals from regressions of patient characteristics on contractor status, controlling for the shift/pod indicators. All variables are standardised. Standard errors are clustered two-way at the doctor and shift/pod level. The F-statistic at the top of the figure is from a reversed regression of contractor status on all the patient characteristics together. The F-statistic is from a test of whether the patient characteristics are jointly equal to zero. p is the p-value of the corresponding test.

FIGURE 3: STABILITY OF COEFFICIENTS



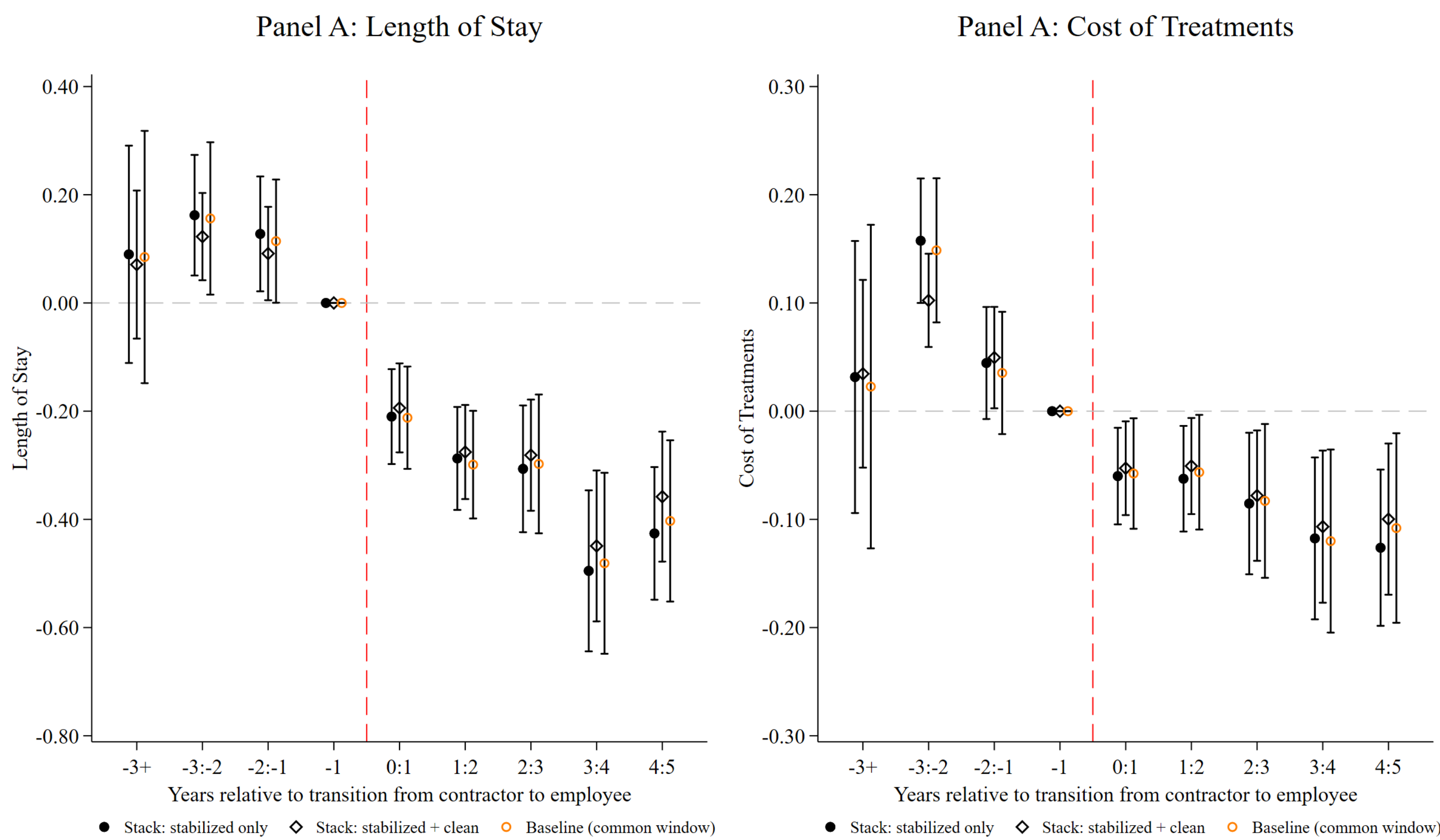
This figure displays the robustness of the coefficients to the inclusion or exclusion of different subsets of patient characteristics. Our subsets of patient characteristics are gender, age (in 50 quantiles), triage, main complaint, referral party, mode of arrival and insurance status. We run separate regressions that control for each of the $2^7 = 128$ different combinations of patient characteristics. We display the number of subsets included in the x-axis. For each number, we plot the maximum, mean and minimum of the estimated coefficients for the effect of the contractor dummy on the dependent variable, using all possible combinations with the number of subsets. The circles and solid lines represent the mean coefficients. The dashed lines represent the maximum and minimum of the estimated coefficients.

FIGURE 4: BASELINE COEFFICIENT AND BALANCE IN PREDICTED DEPENDENT VARIABLES



This figure displays binned scatter plots of actual and predicted outcomes on the contractor dummy, residualised on the shift/pod indicators. In Panel A we display the log length of stay. In Panel B we display the log cost. The solid circles and lines represent the actual outcomes. The hollow circles and dashed lines represent predicted outcomes based on the baseline patient characteristics (age and gender, triage, main complaint, mode of arrival, referral party and insurance status). These patient characteristics are both jointly and separately highly statistically significant, as displayed in Figure 1. The estimated coefficients and the standard errors (clustered two-way at the doctor and hour/pod level), of the regressions on actual and predicted outcomes are reported inside the graphs.

FIGURE 5: LEADS AND LAGS AROUND THE TRANSITION



This figure displays the seven coefficients π_y from estimating: $y_j = \sum_y \pi_y \Big(transition_{i(j)} \times year_{y(j)} \Big) + \eta_{i(j)} + \kappa_{sp(j)} + \lambda \mathbf{X_j} + \epsilon_j$

where $transition_{i(j)} = 1$ if the doctor i assigned to incident j experienced a transition at some point in the sample period. $year_{y(j)} = 1$ if the incident j occurred in year y relative to the transition. $\eta_{i(j)}$ and $\kappa_{sp(j)}$ are doctor and hour/pod fixed effects and $\mathbf{X_j}$ includes the standard patient characteristics (i.e. age, gender, triage, mode of arrival to the ED, insurance status, referral source and main complaint). Standard errors are two-way clustered at the doctor and shift/pod level.

APPENDIX A: TABLES AND FIGURES

TABLE A1
TRANSITIONS BETWEEN
CONTRACTOR AND EMPLOYEE STATUS

	STATUS UPON JOINING THE SAMPLE		
	Employee	Contractor	Student
STATUS UPON LEAVING THE SAMPLE			
Employee	37	14	1
Contractor	0	4	0
Student	0	0	29

This Table displays the transitions between status among the doctors in our sample. The column dimension captures the doctor status when first observed in our sample. The row dimension captures the doctor status when last observed in a sample. The diagonal of the table depicts the doctors that do not switch status.

TABLE A2
BASELINE RESULTS
REASSIGNING OBSERVATIONS TO POD IN WHICH DOCTOR APPEARS MOST

	ED Cost		Length of Stay	
	Original	**Reassigned**	**Original**	**Reassigned**
Panel A: Without Doctor Fixed Effects				
Contractor	.100***	.100***	.313***	.323***
	(.018)	(.018)	(.048)	(.041)
Panel B: With Doctor Fixed Effects				
Contractor	.080***	.058***	.297***	.218***
	(.019)	(.019)	(.034)	(.035)

This table is identical to Table 2 - Baseline Estimates, with the exception that the pod in which the doctor works is calculated differently. In the raw data, sometimes a doctor appears in more than one pod during a shift. In these cases, we reassign the pod of the doctor to the pod that in the data that appears the most. Ties regarding what is the most-common pod are broken using the following priority: M2, M3, M1, M4, Medica and Chirurgica. The number of observations is 291,664.

TABLE A3
BASELINE ESTIMATES
CONTROLLING FOR WITHIN-SHIFT ASSIGNMENT VARIABLES

	ED Cost			Length of Stay		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Panel A: No Doctor Fixed Effects						
Contractor	.100***	.100***	.091***	.313***	.313***	.296***
	(.018)	(.018)	(.018)	(.048)	(.048)	(.046)
Hours to End of Shift		.010	-.008		.007	-.028***
		(.007)	(.006)		(.013)	(.012)
Cases Previously Assigned in Shift			-.010***			-.021***
			(.002)			(.004)
Panel B: With Doctor Fixed Effects						
Contractor	.080***	.075***	.075***	.297***	.285***	.285***
	(.019)	(.019)	(.019)	(.034)	(.034)	(.034)
Hours to End of Shift		-.002	-.002		-.017	-.017
		(.006)	(.006)		(.012)	(.012)
Cases Previously Assigned in Shift			-.007***			-.016***
			(.001)			(.003)

This table displays regressions of log ED cost and log length of stay on contractor status, controlling for within-shift-level variables. All regressions control for patient characteristics (age, gender, triage, main complaint, mode of arrival, referral party and insurance status), and the interaction of shift and pod. Panel B further controls for doctor fixed effects and exact hour (year $\times$ month $\times$ day $\times$ hour). Standard errors two-way clustered at the doctor and shift/pod level. The number of observations is 306,540.

TABLE A4
BASELINE CONTROLLING FOR RECENT ROTATION PATTERNS

	ED Cost		Length of Stay	
	(1)	(2)	(3)	(4)
Contractor	.100***	.099***	.313***	.307***
	(.018)	(.017)	(.048)	(.039)
Total Shifts		-.004		.004
		(.007)		(.015)
Morning Shifts		-.009**		-.024***
		(.004)		(.009)
Afternoon Shifts		.001		-.002
		(.004)		(.011)
M1 Shifts		.009**		.001
		(.004)		(.009)
M2 Shifts		.004		-.013
		(.004)		(.009)
M3 Shifts		.000		-.009
		(.005)		(.010)
M4 Shifts		.003		-.008
		(.005)		(.011)
Medica Shifts		.007*		.010
		(.004)		(.009)
Monday Shifts		.001		-.003
		(.005)		(.010)
Tuesday Shifts		.004		.016
		(.006)		(.012)
Wednesday Shifts		-.000		-.001
		(.006)		(.013)
Thursday Shifts		.008		.011
		(.006)		(.011)
Friday Shifts		.004		.026*
		(.006)		(.014)
Saturday Shifts		.009		.018
		(.008)		(.015)
Doctor Fixed Effects	No	No	No	No

This table displays regressions of log ED cost and log length of stay on contractor status and the doctor's shift rotation patterns in the previous 30 days. Shift rotation variables capture the number of shifts worked in each shift type (morning, afternoon), pod (M1–Medica), and day of week (Monday–Saturday) in the prior 30 days. The omitted groups are night shifts, Chirurgica, and Sunday. All regressions control for whether the assigned doctor is a student, and for the interaction of shift and pod, the exact hour of the visit (the interaction of year, month, day of month and hour of day), and patient characteristics (age, gender, triage, main complaint, mode of arrival, referral party and insurance status). Standard errors two-way clustered at the doctor and shift/pod level. The number of observations is 306,540.

TABLE A5
BASELINE CONTROLLING FOR RECENT ROTATION PATTERNS

	ED Cost		Length of Stay	
	(1)	(2)	(3)	(4)
Contractor	.080***	.080***	.297***	.272***
	(.019)	(.019)	(.034)	(.033)
Total Shifts		-.008		-.018*
		(.005)		(.011)
Morning Shifts		.001		.008
		(.003)		(.005)
Afternoon Shifts		.003		.012*
		(.003)		(.006)
M1 Shifts		.005		-.002
		(.004)		(.008)
M2 Shifts		-.000		-.009
		(.003)		(.006)
M3 Shifts		.002		-.005
		(.004)		(.007)
M4 Shifts		-.000		-.005
		(.005)		(.008)
Medica Shifts		.005		.010
		(.004)		(.006)
Monday Shifts		.001		-.001
		(.004)		(.010)
Tuesday Shifts		.007		.017
		(.005)		(.011)
Wednesday Shifts		.002		.002
		(.005)		(.011)
Thursday Shifts		.004		.004
		(.005)		(.011)
Friday Shifts		.001		.007
		(.005)		(.010)
Saturday Shifts		.010*		.017
		(.006)		(.011)
Doctor Fixed Effects	Yes	Yes	Yes	Yes

This table displays regressions of log ED cost and log length of stay on contractor status and the doctor's shift rotation patterns in the previous 30 days. Shift rotation variables capture the number of shifts worked in each shift type (morning, afternoon), pod (M1–Medica), and day of week (Monday–Saturday) in the prior 30 days. The omitted groups are night shifts, Chirurgica, and Sunday. All regressions control for whether the assigned doctor is a student, and for the interaction of shift and pod, the exact hour of the visit (the interaction of year, month, day of month and hour of day), and patient characteristics (age, gender, triage, main complaint, mode of arrival, referral party and insurance status). Doctor fixed effects are absorbed. Standard errors two-way clustered at the doctor and shift/pod level. The number of observations is 306,540.

TABLE A6
STANDARD DEVIATION OF DOCTOR FIXED EFFECTS

	Total Cost	**Length of Stay**	**Hospita-lisation**	**Death in ED**	**Return to ED 2**	**Return to ED 5**	**Return to ED 30**	**Return to ED 60**
S.D. Raw Fixed Effects	.1407	.2985	.03	.001	.0127	.0159	.0248	.0275
S.D. Bayes Shrunk Fixed Effects	.1309	.2796	.0208	.0003	.0037	.0048	.0068	.0059
Mean Variable	Logs	Logs	.1877	.0009	.0363	.0649	.1327	.174

This table displays regressions of a set of doctor performance measures on individual doctor indicators. We display the standard errors of the estimated fixed effects, both raw and Bayes-Shrunk using the method of Chetty et al. (2014). Cost and length of stay are in logs. Hospitalisation is a dummy for whether the patient is admitted to hospital as opposed to discharged home. Death in ED is a dummy variable for whether the patient dies in the Emergency Department. Return to ED 2 is a dummy variable for whether the patient returns to the Emergency Department generating a different incident within two days. Return to ED 5, 30 and 60 are generated equivalently. All regressions control for the interaction of shift (i.e. the morning, afternoon or night shift within a specific date) and pod, exact hour and patient characteristics (age, gender, triage, main complaint, mode of arrival, referral party and insurance status). The number of observations is 306,540.

TABLE A7
LIST OF 30 MOST COMMON COST ITEMS, BY NUMBER OF COUNTS

Cost Item	Percentage	Item Type
Palliative care consultation (first visit)	7.33	Other
Urinary potassium	5.14	Laboratory
Complete blood count with differential	5.14	Laboratory
Urinary sodium	5.06	Laboratory
Creatinine	5.06	Laboratory
Urea (blood urea nitrogen)	5.05	Laboratory
Pre-prandial glucose	5.04	Laboratory
Transaminase GPT (ALT)	5.02	Laboratory
Creatine kinase (CPK or CK)	4.99	Laboratory
Total calcium	4.82	Laboratory
Venous blood draw	4.55	Laboratory
C-reactive protein: quantitative assay	4.52	Laboratory
Total bilirubin reflex, incl. possible determination of fractionated bilirubin	4.5	Laboratory
Prothrombin time (PT)	3.54	Laboratory
Partial thromboplastin time (PTT)	3.53	Laboratory
Injection or infusion of specific drugs	3.11	Other
Chest X-ray	2.61	Other
Amylase	2.61	Laboratory
Electrocardiogram (ECG/EKG)	2.25	Other
Troponin T	1.68	Laboratory
CT scan of the brain (without contrast)	1.05	Other
Microbiological specimen collection	.84	Laboratory
Abdominal X-ray (without contrast)	.78	Other
Neurology consultation (first visit)	.54	Other
Viral nucleic acid detection in biological material by hybridization following PCR	.52	Laboratory
Blood culture (hemoculture)	.51	Laboratory
Systemic arterial blood gas analysis	.48	Laboratory
Cardiology consultation (first visit)	.38	Other
Complete abdominal ultrasound	.36	Other
Left hip X-ray	.27	Other

This table displays the thirty cost items that appear most often in the Niguarda ED cost system. We report the percentage of the total number of appearances that each cost item represents. The Item Type refers to the category under which they appear in the Niguarda data system (Laboratorio versus Ambulatorio).

TABLE A8
LIST OF 30 MOST COMMON COST ITEMS, BY TOTAL COST

Cost Item	Percentage	Item Type
Palliative care consultation (first visit)	16.95	Other
CT scan of the brain (without contrast)	10.7	Other
Chest X-ray	4.66	Other
CT scan of the abdomen, complete, with and without contrast	3.44	Other
Viral nucleic acid detection in biological material by hybridization following PCR	3.27	Laboratory
Troponin T	3.21	Laboratory
Injection or infusion of specific drugs	3.03	Other
C-reactive protein: quantitative assay	2.69	Laboratory
Electrocardiogram (ECG/EKG)	2.69	Other
Complete abdominal ultrasound	2.65	Other
CT scan of the chest with and without contrast	2.22	Other
CT scan of the brain with and without contrast	2.19	Other
Complete blood count with differential	2.14	Laboratory
Abdominal X-ray (without contrast)	1.71	Other
CT scan of the abdomen, complete, without contrast	1.62	Other
CT scan of the lumbosacral spine (3 vertebrae and 2 interspaces) without contrast	1.49	Other
CT scan of the chest (without contrast)	1.45	Other
Blood culture (hemoculture)	1.44	Laboratory
Venous blood draw	1.36	Laboratory
Upper abdominal ultrasound	1.31	Other
Neurology consultation (first visit)	1.24	Other
CT scan of the facial bones (without contrast)	1.23	Other
Echocardiography	1.21	Other
Creatine kinase (CPK or CK)	1.18	Laboratory
Partial thromboplastin time (PTT)	1.05	Laboratory
Prothrombin time (PT)	.95	Laboratory
Urinary potassium	.9	Laboratory
Urinary sodium	.88	Laboratory
Creatinine	.88	Laboratory
Urea (blood urea nitrogen)	.88	Laboratory
Pre-prandial glucose	.88	Laboratory

This table displays the thirty cost items that contribute the most to the ED cost in the Niguarda ED cost system. We report the percentage of the total ED cost that each cost item represents. The Item Type refers to the category under which they appear in the Niguarda data system (Laboratorio versus Ambulatorio).

TABLE A9
LIST OF 30 MOST COMMON DUPLICATED COST ITEMS

Cost Item	Percentage	Item Type
Troponin T	7.03	Laboratory
Urinary potassium	6.19	Laboratory
Blood culture	5.81	Laboratory
Complete blood count with differential	5.8	Laboratory
Creatine kinase (CPK or CK)	5.73	Laboratory
Transaminase GPT (ALT)	5.62	Laboratory
Urinary sodium	4.72	Laboratory
Creatinine	4.72	Laboratory
Urea (blood urea nitrogen)	4.63	Laboratory
Pre-prandial glucose	4.57	Laboratory
C-reactive protein (CRP): quantitative measurement	4.14	Laboratory
Total bilirubin reflex, incl. possible fractionated bilirubin determination	3.91	Laboratory
Total calcium	3.77	Laboratory
Venous blood draw	2.97	Laboratory
Prothrombin time (PT)	2.84	Laboratory
Partial thromboplastin time (PTT)	2.81	Laboratory
Electrocardiogram (ECG/EKG)	2.29	Other
Amylase	2.12	Laboratory
Palliative care consultation (first visit)	2.03	Other
Drugs of abuse: heroin metabolites in urine	1.52	Laboratory
Chest X-ray	1.39	Other
Left hip X-ray	1.19	Other
Neurology consultation (first visit)	1.16	Other
CT scan of the brain without contrast	1.01	Other
Left wrist X-ray	.91	Other
Orthopedic-traumatology consultation (first visit)	.69	Other
Right arm X-ray	.6	Other
Right tibiotarsal X-ray	.6	Other
Left clavicle X-ray	.54	Other
Left leg X-ray	.54	Other

This table displays the thirty cost items that appear most often as duplicated in the Niguarda ED cost system. We report the percentage of the total number of duplications that each cost item represents. The Item Type refers to the category under which they appear in the Niguarda data system (Laboratorio versus Ambulatorio).

TABLE A10
CORRELATIONS BETWEEN HANDOVERS TO THE NEXT SHIFT AND DUPLICATED COST ITEMS

	(1)	(2)	(3)	(4)	(5)
Panel A: Dep. Var. = Log Number of Duplicated Cost Items					
Handover to Next Shift	.434***			.292***	
	(.006)			(.009)	
Total Number of Doctors		.278***			.229***
		(.004)			(.004)
Length of Stay			.133***	.082***	.068***
			(.004)	(.004)	(.003)
Panel B: Dep. Var. = Share of Duplicated Cost Items					
Handover to Next Shift	.049***			.031***	
	(.001)			(.001)	
Total Number of Doctors		.034***			.028***
		(.001)			(.001)
Length of Stay			.016***	.011***	.008***
			(.000)	(.000)	(.000)

Standard errors two-way clustered at the doctor and shift/pod level. The average number of duplicated cost items is .74. The average share of duplicated cost items within all cost items is .03. Length of stay is in logs. All regressions control for the interaction of shift (i.e. the morning, afternoon or night shift within a specific date) and pod, exact hour and patient characteristics (age, gender, triage, main complaint, mode of arrival, referral party and insurance status). The number of observations is 306,540.

TABLE A11
SPILLOVERS OF CONTRACTORS ON EMPLOYEES' PRODUCTIVITY

	ED Cost	Length of Stay
Panel A: Controlling Only for Shift		
Number of Contractors in Pod	.002	-.018
	(.008)	(.018)
Panel B: Adding Patient Characteristics		
Number of Contractors in Pod	-.001	-.021
	(.007)	(.018)
Panel C: Adding Doctor Fixed Effects		
Number of Contractors in Pod	-.002	-.023
	(.006)	(.016)

This table displays regressions of length of stay and total ED cost on the number of contractors in the pod. The sample includes only employees' observations. A contractor is regarded as present in a pod if the contractor takes at least one case in that pod. Cost and length of stay are in logs. All regressions control for whether the total number of doctors, the total number of students, and for the shift (i.e. the morning, afternoon or night shift within a specific date) and, separately, for the pod. The patient characteristics added in Panels B-C are age, gender, triage, main complaint, mode of arrival, referral party and insurance status. Standard errors two-way clustered at the doctor and shift/pod interaction level. The number of observations is 262,357.

TABLE A12
EFFECTS ON OVERSTAY VARIABLE
ONLY OBSERVATIONS THAT SPILL OVER NEXT SHIFT

	(1)	(2)	(3)	(4)
Employee		.0643***	.0469***	.0418***
		(.0093)	(.0106)	(.0104)
Difference in Gender	-.0056			-.0055
	(.0058)			(.0057)
Difference in Age	-.0015***			-.0011**
	(.0005)			(.0005)
Doctor Fixed Effects	Yes	No	Yes	Yes

Standard errors two-way clustered at the doctor and shift/pod level. Overstay takes value one when a patient is discharged by the doctor originally assigned to the case, but in the sixty minutes after that doctor's shift has ended. Difference in gender is the absolute value of the difference between a male dummy for the doctor initially assigned to the incident, and the average male dummy for the doctors working in the same pod and during the following shift. Difference in age variable is defined in an analogous way. All regressions control for the interaction of shift (i.e. the morning, afternoon or night shift within a specific date) and pod, exact hour and patient characteristics (age, gender, triage, main complaint, mode of arrival, referral party and insurance status). The regressions in Columns 2-4 control for whether the assigned doctor is a student. The mean of the dependent variable Overstay is .091. The number of observations is 120,994.

TABLE A13
ROBUSTNESS OF EMPLOYEE × INCOMING EMPLOYEES COEFFICIENT TO CONTROLLING FOR INTERACTIONS OF EMPLOYEE AND SHIFT/POD CHARACTERISTICS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Employee × Incoming Employees	.016***	.015***	.015***	.015***	.018***	.017***	.020***	.018***
	(.006)	(.006)	(.006)	(.006)	(.006)	(.006)	(.007)	(.006)
Doctor F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Employee × Pod F.E.	No	Yes	No	No	No	No	No	Yes
Employee × Shift F.E.	No	No	Yes	No	No	No	No	Yes
Employee × Year F.E.	No	No	No	Yes	No	No	No	Yes
Employee × # Cases Deciles F.E.	No	No	No	No	Yes	No	No	Yes
Employee × # Cases per Doctor Deciles F.E.	No	No	No	No	No	Yes	No	Yes
Employee × # Doctors Next Shift F.E.	No	No	No	No	No	No	Yes	Yes

Standard errors two-way clustered at the doctor and shift/pod level. Overstay takes value one when a patient is discharged by the doctor originally assigned to the case, but in the sixty minutes after that doctor's shift has ended. Incoming Employee is the average of the employee dummy for the doctors working in the same pod and during the following shift. All regressions control for the interaction of shift and pod, patient characteristics (age, gender, triage, main complaint, mode of arrival, referral party and insurance status), and doctor fixed effects. The mean of the dependent variable Overstay is .037. The number of observations is 305,798.

TABLE A14
EFFECTS ON OVERSTAY VARIABLE - ROBUSTNESS

	(1)	(2)	(3)
Panel A: Unweighted Average Among Next-Shift Doctors			
Employee	.008	.004	.004
	(.006)	(.006)	(.006)
Employee × Incoming Employees	.012*	.012*	.009
	(.007)	(.007)	(.007)
Panel B: Using the Next-Shift Doctor with Most Cases			
Employee	.010**	.005	.005
	(.005)	(.006)	(.006)
Employee × Incoming Employees	.010**	.010**	.008**
	(.004)	(.004)	(.004)
Panel C: Overstay Calculated Over 2-Hour Window			
Employee	.004	-.002	-.001
	(.006)	(.007)	(.007)
Employee × Incoming Employees	.016**	.016**	.013*
	(.007)	(.007)	(.007)
Panel D: Dropping Shifts That Doctor Starts Too Early or Ends Too Late			
Employee	.005	-.003	-.003
	(.006)	(.007)	(.007)
Employee × Incoming Employees	.014**	.015**	.014**
	(.007)	(.007)	(.007)
Doctor Fixed Effects	No	Yes	Yes

This table is equivalent to Columns 5-7 in Table 7, with the following differences. In Panel A, we compute the variable Incoming Employees by weighing all incoming doctors equally, independently of how many incidents they deal with during the shift. In Panel B, we calculate this variable by using only the doctor that receives the highest number of incidents during the shift. Instead, in Table 7 and in the current Panels C-D, we weigh all incoming doctors by the number of incidents that they deal with during the shift. In Panel C, we compute the overstay variable by using a 2-hour window. The variable takes value one when a patient is discharged in the two hours after a shift started (i.e., 8am-10am, 2pm-4pm, or 8pm-10pm), was assigned to the initial doctor in the previous shift, and was discharged by that initial doctor. In Panel D, we compute the overstay variable by using a sixty minutes window but drop observations in which the doctor initially assigned to the case might have started their shift earlier than on the official time. Specifically, we identify doctors that we observe in at least one incident across two consecutive shifts. We then drop observations for the shift in which the doctor is observed in less incidents.

TABLE A15
MECHANISM
HETEROGENEITY BY EMPLOYEES FOLLOWING EMPLOYEE

	(1)	(2)	(3)	(4)	(5)
Panel A: Dependent Variable = Length of Stay					
Employee	-.313***	-.297***	-.197***	-.228***	-.231***
	(.048)	(.034)	(.064)	(.047)	(.047)
Employee × Incoming Employees			-.132***	-.079*	-.065
			(.046)	(.045)	(.046)
Difference in Gender					.023*
					(.013)
Difference in Age / 10					-.015
					(.014)
P-value 1+2			.000	.000	.000
Doctor Fixed Effects	No	Yes	No	Yes	Yes
Panel A: Dependent Variable = ED Cost					
Employee	-.100***	-.080***	-.056**	-.057**	-.057**
	(.018)	(.019)	(.025)	(.028)	(.029)
Employee × Incoming Employee			-.050**	-.027	-.024
			(.022)	(.021)	(.023)
Difference in Gender					.007
					(.007)
Difference in Age / 10					-.009
					(.007)
P-value 1+2			.000	.000	.000
Doctor Fixed Effects	No	Yes	No	Yes	Yes

Standard errors two-way clustered at the doctor and shift/pod level. Difference in gender is the absolute value of the difference between a male dummy for the doctor initially assigned to the incident, and the average male dummy for the doctors working in the same pod and during the following shift. Difference in age variable is defined in an analogous way. Incoming Employee is the average of the employee dummy for the doctors working in the same pod and during the following shift. In Panel A, the dependent variable is length of stay and columns 1 and 2 are equivalent to column 2 in Table 2 Panels B and C. In Panel B, the dependent variable is ED cost and columns 1 and 2 are equivalent to column 1 in Table 2 Panels B and C. Length of stay and ED cost are in logs. Incoming employee is the employee dummy for the doctors working in the same pod and during the following shift, weighted by the number of incidents that they deal with. All regressions control for whether the assigned doctor is the student, the interaction of shift (i.e. the morning, afternoon or night shift within a specific date) and pod, and patient characteristics (age, gender, triage, main complaint, mode of arrival, referral party and insurance status). The number of observations is 303,466.

TABLE A16
CORRELATION BETWEEN OVERLAP MEASURES AND EMPLOYEE STATUS

	Horizontal Overlap (Past)	Backward Handover (Past)	Forward Handover (Past)	Horizontal Overlap (Future)	Backward Handover (Future)	Forward Handover (Future)
Panel A: Without Doctor Fixed Effects						
Employee	-.009*	.002	.000	-.010**	.006***	.001
	(.005)	(.004)	(.002)	(.004)	(.002)	(.002)
Panel B: With Doctor Fixed Effects						
Employee	-.001	-.003	-.004*	-.001	.003	-.001
	(.004)	(.004)	(.002)	(.002)	(.002)	(.001)
Mean Variable	.16	.053	.053	.16	.047	.047

This table displays regressions of overlap measures on the employee dummy. The sample is the baseline sample. All regressions control for whether the assigned doctor is a student, the interaction of shift and pod, exact hour and patient characteristics (age, gender, triage, main complaint, mode of arrival, referral party and insurance status). Horizontal Overlap (Past) is the share of focal doctor shifts in the past 365 days where the focal doctor coincided with the incoming doctors. The formula is $Horizontal_{jt} = \frac{\sum_{n=1}^{N}\sum_{k=1}^{K} horizontal_{jkn}}{N \times K}$, where j is the focal doctor, k is one of the incoming doctors, n is a focal doctor shift in the past 365 days, and $horizontal_{jkn} = 1$ if k worked also in that shift. Backward Handover is the share of focal doctor shifts where the incoming doctor followed the focal doctor in the same pod. Forward Handover is the share of focal doctor shifts where the focal doctor followed the incoming doctor in the same pod. The Future variables are defined accordingly, but computed over the future 365-day window. Standard errors two-way clustered at the doctor and shift/pod interaction level. The number of observations is 303,466.

TABLE A17
OVERLAP VARIABLES AND OVERSTAY

	(1)	**(2)**
Horizontal Overlap (Past)	.028	.020
	(.019)	(.018)
Backward Handover (Past)	-.095***	-.065***
	(.026)	(.024)
Forward Handover (Past)	-.032	-.010
	(.033)	(.030)
Horizontal Overlap (Future)	.033	.031*
	(.021)	(.016)
Backward Handover (Future)	-.073	-.020
	(.045)	(.035)
Forward Handover (Future)	.028	.050
	(.035)	(.032)
Doctor F.E.	No	Yes

This table displays regressions of the overstay dummy on overlap measures. The sample is the baseline sample. All regressions control for the interaction of shift and pod, exact hour and patient characteristics (age, gender, triage, main complaint, mode of arrival, referral party and insurance status). Horizontal Overlap (Past) is the share of focal doctor shifts in the past 365 days where the focal doctor coincided with the incoming doctors. The formula is $Horizontal_{jt} = \frac{\sum_{n=1}^{N}\sum_{k=1}^{K} horizontal_{jkn}}{N \times K}$, where j is the focal doctor, k is one of the incoming doctors, n is a focal doctor shift in the past 365 days, and $horizontal_{jkn} = 1$ if k worked also in that shift. Backward Handover is the share of focal doctor shifts where the incoming doctor followed the focal doctor in the same pod. Forward Handover is the share of focal doctor shifts where the focal doctor followed the incoming doctor in the same pod. The Future variables are defined accordingly, but computed over the future 365-day window. Standard errors two-way clustered at the doctor and shift/pod interaction level. The number of observations is 287,604.

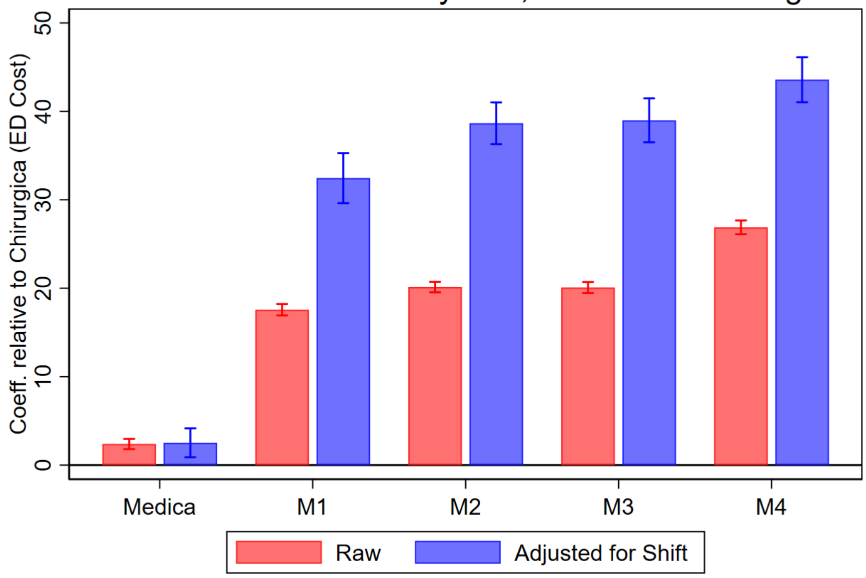
TABLE A18
MECHANISM: EFFECTS ON OVERSTAY VARIABLE CONTROLLING FOR OVERLAP VARIABLES

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Employee		.020***	.015***	.013***	.006	.000	.001
		(.003)	(.004)	(.004)	(.006)	(.006)	(.006)
Employee × Incoming Employees					.017***	.017***	.015***
					(.006)	(.006)	(.006)
Difference in Gender	-.003			-.003			-.003
	(.002)			(.002)			(.002)
Difference in Age / 10	-.005***			-.004**			-.003*
	(.002)			(.002)			(.002)
P-value 1+2					.000	.000	.001
Doctor Fixed Effects	Yes	No	Yes	Yes	No	Yes	Yes
Past Overlap Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Future Overlap Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes

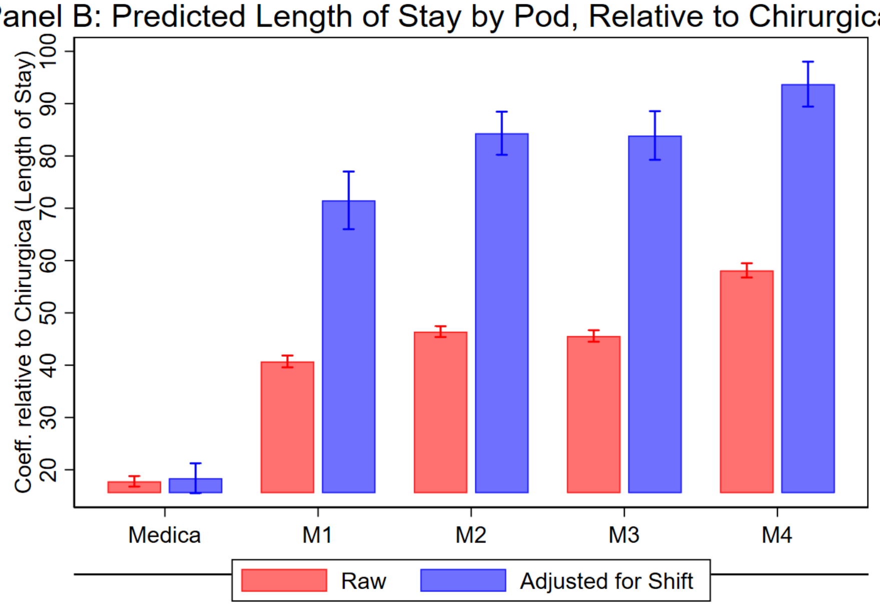
Standard errors two-way clustered at the doctor and shift/pod level. Overstay takes value one when a patient is discharged by the doctor originally assigned to the case, but in the sixty minutes after that doctor's shift has ended. Difference in gender is the absolute value of the difference between a male dummy for the doctor initially assigned to the incident, and the average male dummy for the doctors working in the same pod and during the following shift. Difference in age variable is defined in an analogous way. Incoming Employee is the average of the employee dummy for the doctors working in the same pod and during the following shift. All regressions control for the interaction of shift (i.e. the morning, afternoon or night shift within a specific date) and pod, and patient characteristics (age, gender, triage, main complaint, mode of arrival, referral party and insurance status). The regressions in Columns 2-7 control for whether the assigned doctor is a student. The mean of the dependent variable Overstay is .037. The number of observations is 287,604.

FIGURE A1: AVERAGE PREDICTED COST AND PREDICTED LENGTH OF STAY, BY POD

Panel A: Predicted ED Cost by Pod, Relative to Chirurgica

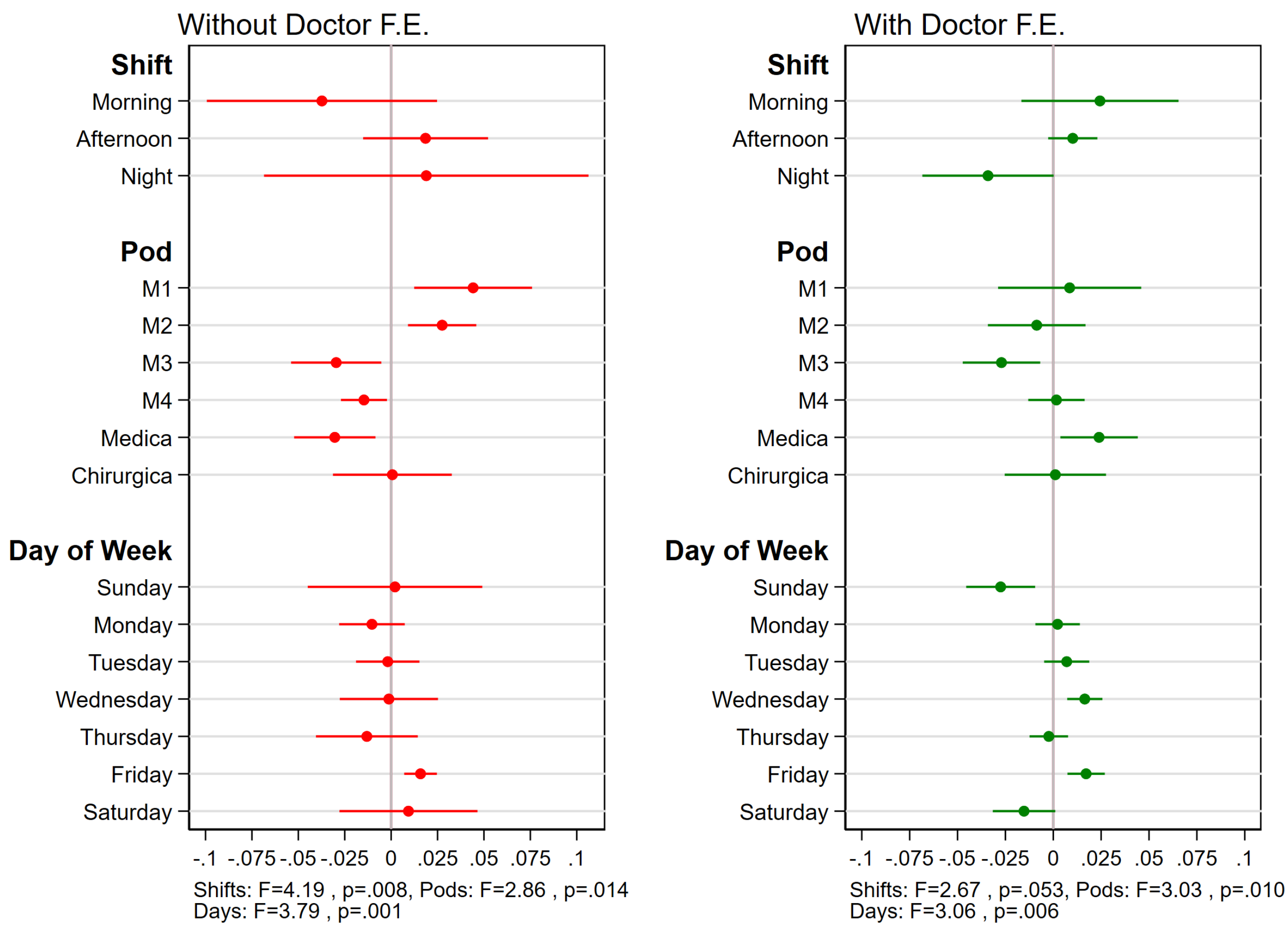


Panel B: Predicted Length of Stay by Pod, Relative to Chirurgica



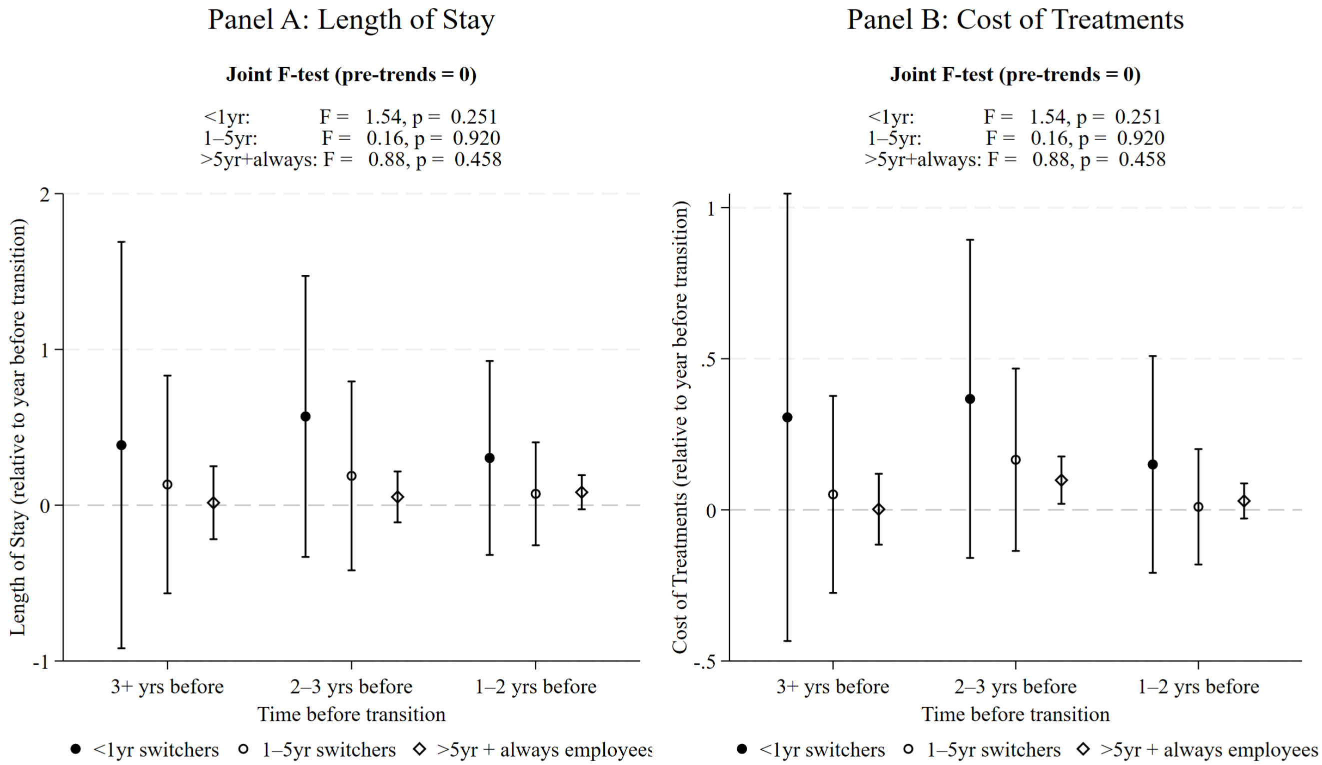
This figure displays regression coefficients of predicted cost and predicted length of stay on the pod to which the incident is assigned. The predictions are done based on patient characteristics (age, gender, triage, main complaint, mode of arrival, referral party and insurance status). Cost and length of stay are in logs. The regressions in Panel B control for the exact shift of the incident (i.e., the morning, afternoon or night shift within a specific date). 95% confidence intervals displayed. Standard errors two-way clustered at the doctor and shift/pod interaction level. The sample is the baseline sample. The number of observations is 306,635.

FIGURE A2: ASSIGNMENT OF SHIFTS AND PODS BY CONTRACTOR STATUS



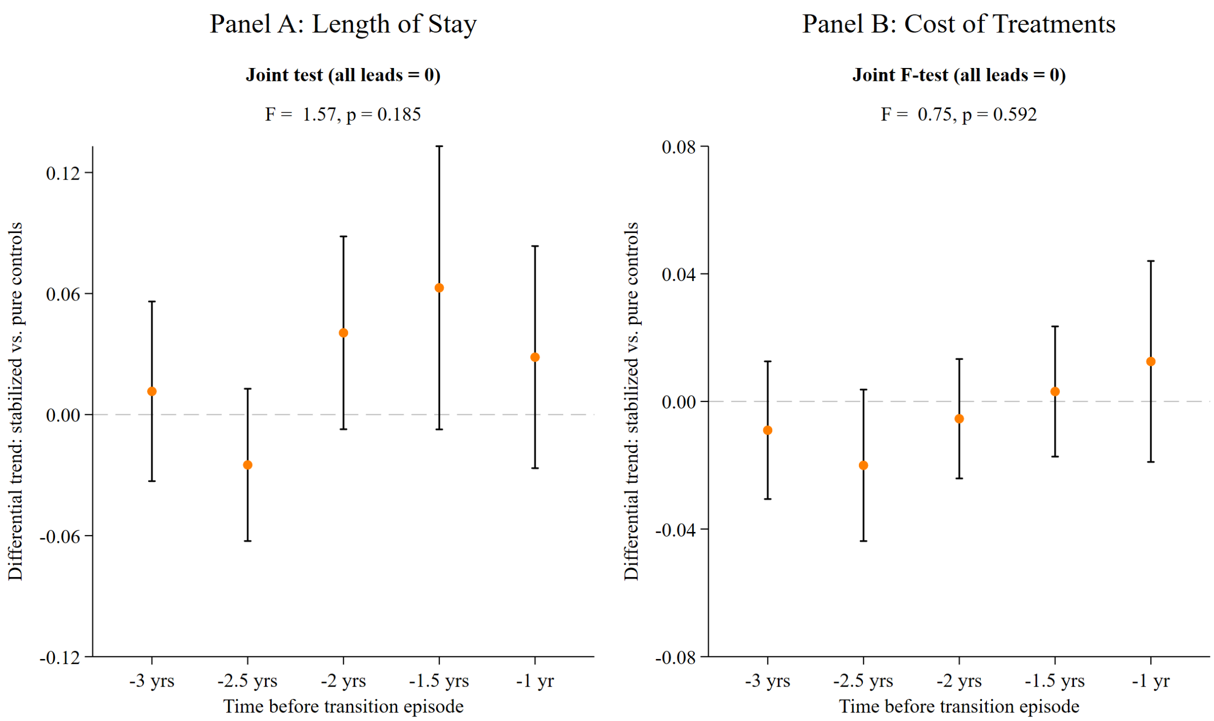
This figure displays regression coefficients of shift and pod characteristics on contractor status. The sample is a sample of shift/pod and doctors. Each row displays the coefficient of a separate regression with the row characteristic on the left hand side and contractor status on the right hand side. All regressions control for student dummy. The shift regressions control for exact date. The pod regressions control for the exact shift. The day of week regressions control for the interaction of year and week. F-statistics of tests that all the shift (alternatively, pod or day of week) are jointly equal to zero are displayed. Standard errors are two-way clustered at the shift and doctor level.

FIGURE A3: FORMAL TESTS OF PRE-TRENDS BY SAMPLE RESTRICTION



This figure displays regression coefficients of shift and pod characteristics on contractor status. The sample is a sample of shift/pod and doctors. Each row displays the coefficient of a separate regression with the row characteristic on the left hand side and contractor status on the right hand side. All regressions control for student dummy. The shift regressions control for exact date. The pod regressions control for the exact shift. The day of week regressions control for the interaction of year and week. F-statistics of tests that all the shift (alternatively, pod or day of week) are jointly equal to zero are displayed. Standard errors are two-way clustered at the shift and doctor level.

FIGURE A4: COMPARING TRENDS BETWEEN STABILIZED EMPLOYEES VS. CLEAN CONTROLS

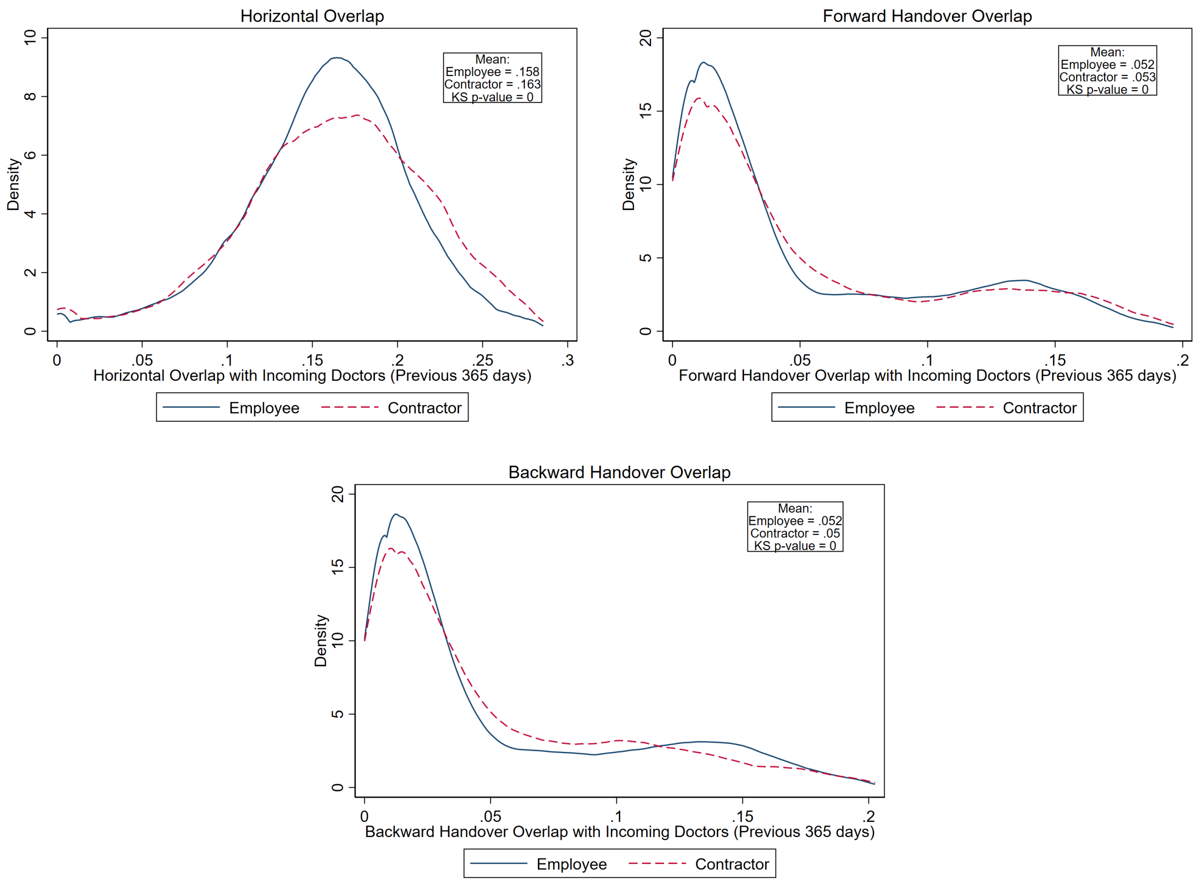


This figure displays the lead coefficients δ_s from estimating:

$$y_{je} = \sum_{s=2}^{6} \gamma_s \, \mathbb{1}\{S_{je} = s\} + \sum_{s=2}^{6} \delta_s \left(\mathbb{1}\{S_{je} = s\} \times \textit{stabilized}_{i(j)} \right) + \eta_{i(j)} + \phi_e + \kappa_{h \times p(j)} + \mathbf{X}_j' \lambda + \varepsilon_{je},$$

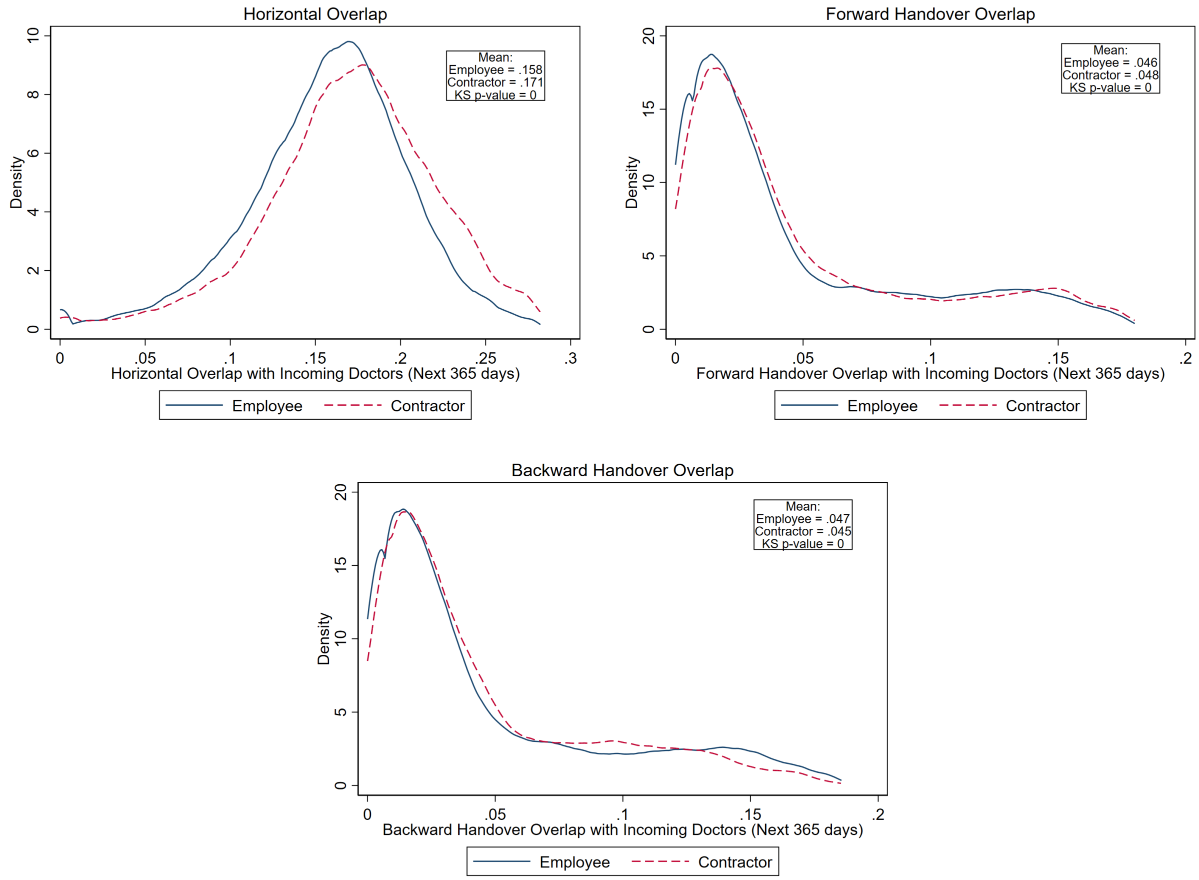
where e indexes transition episodes (one per switching doctor), and the data are stacked by constructing, for each episode e, a pre-transition window of three years prior to the episode transition date. S_{je} indexes six-month bins before the episode transition (with the 0–6 months bin omitted), and $\textit{stabilized}_{i(j)} = 1$ for doctors who are always observed as employees or for switchers observed more than five years after their own transition into employee status. The omitted category is the 0–6 months pre-transition bin, so the plotted coefficients δ_s capture differential pre-period movements of stabilized employees relative to the clean control group (never-employees and not-yet-employees, defined using each doctor's own transition date) within episode windows. $\eta_{i(j)}$, ϕ_e, and $\kappa_{h \times p(j)}$ denote doctor, episode, and hour×pod fixed effects (hdateamb), and $\mathbf{X}_j$ includes patient controls (sex, age, triage, condition, arrival mode, and other baseline covariates). Standard errors are two-way clustered by doctor and shift-date. Observations are weighted by the inverse of the number of episode windows in which they appear. The joint F-test reported above each panel tests $H_0 : \delta_s = 0$ for all displayed leads.

FIGURE A5: DISTRIBUTION OF OVERLAP BETWEEN FOCAL DOCTOR AND INCOMING DOCTOR(S) (PREVIOUS 365 DAYS)



These figures display measures of overlap between the focal doctor and the doctors that are incoming in the same pod and next shift. The horizontal overlap measure is the number of shifts in which the focal doctor and the incoming doctors have coincided over the previous year, divided by the product of: (a) the number of shifts that the focal doctor has had and, (b) the number of incoming doctors. The forward handover overlap is the number of shifts in which the focal doctor has been followed by the incoming doctors, over the previous year, divided by the same denominator as the horizontal overlap measure. The backward handover overlap is similar to the forward handover overlap, but with the focal doctor following the incoming doctors. The dataset is the baseline dataset of incidents, and the measures are computed separately for employees and contractors.

FIGURE A6: DISTRIBUTION OF OVERLAP BETWEEN FOCAL DOCTOR AND INCOMING DOCTOR(S) (NEXT 365 DAYS)



These figures display measures of overlap between the focal doctor and the doctors that are incoming in the same pod and next shift. The horizontal overlap measure is the number of shifts in which the focal doctor and the incoming doctors will coincide over the next year, divided by the product of: (a) the number of shifts that the focal doctor will have and, (b) the number of incoming doctors. The forward handover overlap is the number of shifts in which the focal doctor will be followed by the incoming doctors, over the next year, divided by the same denominator as the horizontal overlap measure. The backward handover overlap is similar to the forward handover overlap, but with the focal doctor following the incoming doctors. The dataset is the baseline dataset of incidents, and the measures are computed separately for employees and contractors.

Appendix B: Event Study Robustness

Recent staggered difference-in-differences (DiD) estimators—such as Callaway and Sant'Anna (2021), Sun and Abraham (2021), and Borusyak, Jaravel, and Spiess (2024)—typically require a sufficiently large and well-covered set of never-treated and/or not-yet-treated observations to derive credible counterfactual trends. In our setting, however, we observe only four never-treated contractors, who work substantially fewer shifts than employees and appear only sporadically across shift-by-pod cells. Not-yet-treated observations are likewise limited and sparsely distributed in pre-transition windows. Because our baseline specifications absorb shift-by-pod fixed effects, identification requires transitioning doctors and comparison observations to coincide within the same shift-by-pod cells; this overlap occurs in relatively few cells, making staggered DiD implementations infeasible or severely underpowered.

Empirically, once we absorb granular shift-by-pod or hour-by-pod fixed effects, the scarcity of clean controls implies that most of the effective comparison variation is driven by doctors who are always observed as employees or those who transitioned to employees many years before. This group is the most prevalent and provides dense coverage across shift-by-pod cells, yielding the overlap necessary for estimation in the presence of thousands of fixed effects.

Figure 5 shows that the estimated dynamic effects are similar when we explicitly restrict the control group to doctors with a long tenure as employees. We verify this using a stacked-episode implementation of the baseline event study, defining one episode around each contractor-to-employee transition. For each episode, we retain the transitioning doctor's event-time observations and restrict the comparison pool to doctors who are always observed as employees and switchers observed at least five years after their own transition into employee status. We also report a specification that augments this stabilized pool with "clean controls" (never-employed and not-yet-employed doctors, when observed).[1]

Therefore, the identifying assumption is a stable-treatment-effects (steady-state) condition: after sufficient tenure in employee status, the effect of becoming employee on outcomes has converged, so subsequent outcome dynamics for stabilized employees reflect common shocks and time-varying confounders that are shared with contractors. In this framework, level differences—potentially including permanent components of the treatment effect—are absorbed by doctor fixed effects (and other controls), while identification comes from differential changes over time between switchers and stabilized employees. Figure 5 suggest that after transitioning to employee status, doctors converge relatively quickly to a new steady state and effects do not continue to increase in the long run.

The stabilization restriction mitigates the "forbidden comparisons" concern that can arise in staggered adoption designs when already-treated units are used as controls while their treatment effects are still adjusting. If the post-transition path is not yet in steady state, differences in pre-period coefficients may reflect heterogeneous treatment dynamics (and the choice of baseline) rather than violations of parallel trends. Restricting the comparison set to workers plausibly in a steady state—always-employees and switchers observed sufficiently long after their own transition—reduces this source of bias and makes pre-period diagnostics more informative. As in any event-study design, this approach still relies on standard, untestable conditions: no confounding shocks on coincident with transition timing and no spillovers across doctors.

The key question, therefore, is whether stabilized employees provide valid counterfac-

[1] The baseline series restricts the lead/lag window to match the stacked specifications, and we exclude students throughout. In the stacked specifications, we also exclude switchers with fewer than five years of tenure in employee status from the comparison pool.

tual trends for contractors transitioning to employees. We assess this in two complementary ways. First, Figure A3 reports lead coefficients from event-study regressions estimated separately for comparison groups defined by employee tenure (These diagnostic regressions are estimated separately from the main effects, in the spirit of Borusyak, Jaravel, and Spiess, 2024.). If treatment effects stabilize with tenure, longer-tenure groups should exhibit flatter pre-trends. Because recent and mid-tenure switchers are sparse in the relevant hour-by-pod cells once we absorb high-dimensional fixed effects, these leads are estimated with low power and wide confidence intervals. By contrast, as tenure increases—and especially for the stabilized group—the lead coefficients become smaller and much more precisely estimated; for stabilized employees, they are close to zero and jointly insignificant.

Second, we directly compare pre-period dynamics for stabilized employees and the "clean control" group defined by contractors that never transitioned to employees and yet-to-transition contractors (Figure A4). This comparison is feasible because stabilized employees are observed in a large share of shift-by-pod cells, so the subset of cells in which they co-occur with clean controls is larger than the subset in which clean controls co-occur with near-transition switchers. Put differently, there remains non-trivial overlap between clean controls and stabilized employees working simultaneously in the same shift-pod, whereas the overlap between clean controls and switchers is much more limited.

Specifically, for each of the 14 transition episodes we retain observations in the three years preceding the transition date, stack these windows, and estimate

$$y_j = \sum_{s=2}^{6} \gamma_s \mathbf{1} S_j = s + \sum_{s=2}^{6} \delta_s \mathbf{1} S_j = s \cdot \mathit{stabilized}_{i(j)} + \eta_{i(j)} + \phi_{e(j)} + \kappa_{sp(j)} + \boldsymbol{\lambda}' \mathbf{X} j + \epsilon j,$$

where S_j indexes 6-month bins prior to the episode transition (the 0–6 months bin is omitted), $\phi_{e(j)}$ are episode fixed effects, and the remaining controls follow the baseline specification. The interaction coefficients δ_s capture differential pre-period movements between stabilized employees and clean controls; under parallel trends, $\delta_s = 0$ for all s. Figure A4 shows that the δ_s estimates are small and statistically indistinguishable from zero.

Together, these exercises support the identification strategy. Sparse clean-control coverage makes standard staggered DiD implementations impractical in our setting. Instead, after conditioning on our granular fixed effects, the effective identifying comparisons are driven primarily by always-employees and by switchers observed several years after their contractor-to-employee transition. Their pre-period dynamics closely track those of clean controls in the windows where overlap exists, and the baseline estimates are very similar when the comparison group is restricted to this set of employee observations.

References

Callaway, Brantly, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." Journal of Econometrics 225(2): 200–230.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." Journal of Econometrics 225(2): 175–199.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2024. "Revisiting Event-Study Designs: Robust and Efficient Estimation." The Review of Economic Studies 91(6): 3253–3285.