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

Diego Battiston[†] Jordi Blanes i Vidal[‡]

Elena Lucchese [§]

April, 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 take at least 24% longer to treat and discharge patients and spend at least 7% more on tests and treatments, yet achieve no detectably better patient outcomes. These gaps largely reflect the causal effect of contractor status rather than selection into that status. We uncover an important mechanism: employees are willing to exert effort on behalf of their fellow employees, but not on behalf of contractors.

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

Keywords: Contractors, Labour Productivity, Emergency Department, Unpaid Overtime, Insider Econometrics.

*We thank David Chan, Bob Gibbons, Cristobal Otero, 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 this share at ∼15% of the US workforce. Contractors' importance is growing over time and particularly concentrated 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 (Bloom et al. 2015, Mas and Pallais 2017, Ameriks et al. 2020).

In this paper, we provide the first evidence on a central question concerning contractors: how their productivity compares with that of permanent employees. Theoretical predictions are mixed. Steeper incentives arising from contract-renewal risk (Shapiro and Stiglitz, 1984), together with the broader experience accumulated across multiple 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 in the organisation (Akerlof and Kranton, 2005). Assessing these performance differences (and the mechanisms behind them) 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 direct empirical evidence remains scarce. Large administrative datasets, such as matched employer-employee data, 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 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 treating identical patients. Access to detailed input and output variables allows us to measure

[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.

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 patient's time of arrival, and is therefore orthogonal to both doctor and patient characteristics.[2] This feature of the 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 27 log points higher than for patients assigned to employees. The total cost of ED care is 7 log points higher for patients assigned to contractors.[3] Among other items, this includes a 44 log points increase in the cost of laboratory tests, a 16% increase in the likelihood of requesting an x-ray, and a 8% 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 21% 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 and treatments.

[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).

We begin to investigate mechanisms for these differences by exploiting within-doctor variation. We first find that the estimated differences in resource use remain remarkably robust to including individual doctor fixed effects. We further find these differences to remain large even after controlling for time-varying characteristics, such as doctor age, tenure and experience. Overall we find that, across different specifications, the estimated coefficients range between 7 and 10 log points for cost, and between 21 and 31 log points for length of stay. 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 largely reflect the treatment effect of the contractual arrangement, as opposed to the sorting of more productive individuals into these arrangements (Lazear, 2000).

The contractual arrangement is a bundle of characteristics, and the treatment effect we estimate may therefore operate through several channels. For example, contractual status might influence doctors' propensity to engage in 'defensive medicine' (Frakes and Gruber, 2019), due to differences in who pays for liability insurance.[4] 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 the organisation and their fellow employees (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. 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]

We find that employees are 50% more likely to overstay than contractors. Crucially, this difference also holds within doctors: the same individual becomes more likely to overstay

[4]Section 2 discusses that contractors and employees are equally insured. The sole distinction is that the hospital covers employees' insurance for minor negligence, whereas contractors are responsible for this cost themselves.

[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.

after moving from contractor to employee status. We interpret employees' greater propensity to overstay as a window into their generally higher motivation to exert more effort so that their assigned cases are resolved faster and with fewer tests and treatments, a motivation that likely operates across all cases and not only those for which overstaying is a relevant option.[6]

We then investigate the mechanism motivating 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. We further find that employees' lower length of stay and costs are even lower when the doctor in the following shift is also an employee. Therefore, employees' greater willingness to exert effort is not only a general behavioural difference, but at least partly a targeted willingness to assist colleagues who share their employment status within the organisation.

The higher willingness of employees to cooperate with fellow employees, but not with contractors, can be micro-founded in two ways. First, employment status may demarcate an 'insider' group, inducing doctors to internalise the impact of their behaviour on fellow employees while discounting the welfare of contractors (Akerlof and Kranton 2005, Chen and Li 2009). Second, employees may reciprocate or exchange favours with one another in a relational contract sense, while excluding contractors from this reciprocal arrangement.[7] Distinguishing empirically between these micro-foundations is beyond the scope of this paper, but we regard it as a promising direction for future work. What our evidence establishes, taken together, is that contractors and employees differ meaningfully in productivity, that this difference is at least partly causal, and that it is shaped at least in part by the social structure of the workplace, specifically by whether the colleague standing to benefit shares one's employment status. We view each of these as a substantive step forward in understanding the productivity consequences of how firms choose to staff their workforce.

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

[6]Overstaying to discharge the patient may not, for instance, be a possibility for cases that are expected to last well into the next shift. However, even in these cases, the initially assigned doctor can exert effort to hand over a patient that is closer to resolution.

[7]We investigate and fail to find evidence that the patterns of shift rotations provide employees with more opportunities to exchange favours, relative to contractors. However, our tests are relatively weak, as they do not examine all the channels through which doctors could reciprocate with one another.

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.[8]

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. the contract linking worker and firm) (Mas and Pallais, 2020). A recent 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 et al. 2023). 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.[9]

[8]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.

[9]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 35% of patients are handed over to the incoming doctor in the next shift. Thus, the end-of-shift decisions that Chan studies

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 effects on ED cost and length of stay. In Section 5 we present the effects on other variables. In Section 6, we explore potential mechanisms. Section 7 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.[10]

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, having recently been ranked 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 into pods — clusters of beds 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 training is primarily outside emergency medicine.

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.

[10] 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. National guidelines target waiting times and length of stay as the main performance measure by which hospitals are evaluated (Ministero della Salute, 2024).

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. The general pods are not entirely homogeneous. For example, patients expected to remain in the ED for extended periods are usually assigned to pod M4.[11]

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 to a pod and, within that pod, to a specific doctor. Pod assignment depends on the characteristics 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.[12] In Sections 3 and 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. Unlike in some other settings (e.g., 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 sometimes choose to stay beyond their official hours to complete treatment and discharge their patients. As discussed in Section 6, this practice, though voluntary, is often considered 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 who stays beyond their scheduled shift to complete treatment unpaid).

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 causal role of subsequent doctors who have been handed over the patient by the originally assigned doctor.

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

[11]Figure 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 further controls for the shift (i.e., the morning, afternoon or night shift within a specific date) during which the incident arrives at 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.

[12]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.

both forms of employment are widespread 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 allowed.

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 occassionally found ways around the restrictions. Nevertheless, the cumulative effect was a steady contraction of the 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.[13] Contractors were integrated into the standard staffing pool and assigned shifts by the ED director alongside employee doctors.[14] 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 within the ED, employment conditions differed. The first important difference concerns

[13] 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 one of the best hospitals in Italy.

[14] 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 practice, 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.

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 individual contributions to ED resource use or on patient outcomes. In fact, statistics on average resource use or patient outcomes by doctor were not even computed at all.

A second major difference concerned integration into the hospital organisation and culture. Employees were embedded in the hospital's collective routines: they attended 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 previous 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 revealed that employees also self-organised team-building activities, in which contractors did not take part.[15]

Third, employees and contractors differed in who paid for their liability insurance. Niguarda covers employees for civil liability for minor negligence, while contractors have to cover that cost themselves. Both types of doctors typically purchase additional insurance for cases of gross negligence.[16]

Last, employees had a long-term stake in the institution, expecting to remain at 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.[17] Within Niguarda, these contrasts were seen by staff as shaping 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 having looser attachments to both colleagues and the institution.

Transitions from Contractor to Employee During our sample period, hospitals in Lombardy could occasionally open vacancies for permanent employment, particularly after

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

[16] The annual premium for gross (respectively, minor) negligence is approximately €500 (respectively, €1,000).

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

2018. Crucially, hospitals had no say over which applicant would fill a given vacancy. The procedure operated as follows. Whenever a hospital opened a position, a nation-wide competition was held. Applicants were ranked based on rigid, objective criteria, including the performance in a nation-wide exam, prior experience, educational qualifications, and academic publications. These criteria were set by national and regional laws (Laws 502/1992 and 165/2001), with no role for hospital preferences.[18] Softer measures, including reference letters or in prior job performance assessments, 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 from January 2014 to December 2023. During this time, 14 doctors transitioned from contractor roles to employee status (Table A1). Three main factors explain why most contractors who secured permanent posts remained at Niguarda. 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 at Niguarda was opened. Last, Niguarda is a large employer, so it was relatively easy for contractors to find a position there. Crucially, 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 identity of the doctor first assigned responsibility for the patient, as well as the identity of the doctor discharging the patient. 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 analysis to cases assigned to a general pod and managed by a specialist in emergency medicine. The final dataset includes 305,221 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. Both

[18] Specifically, the criteria established a number of points assigned to each potential CV item. For instance, having done the military service as a medical doctor generated .5 points. An academic article or book chapter within the previous five years generated .2 points. Having an academic faculty contract provided .3 points. The CV part was worth at most 20 points. The exam contained three parts: a written test, an oral part and a practical component. The written test had 30 questions and each correct answer generated one point.

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 rather than discharged home). We use in-ED mortality and return visits to the ED within 2, 5, 30, and 60 days as indicators of patient outcomes (Schwab and Singh, 2024).

Table 1 reports summary statistics for all doctors and separately for employees, contractors, and students.[19] On average, employees (42 years) 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 who were contractors at some point and 30 who were students at some point, Table A1).

3 Empirical Strategy

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

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 for whether the doctor i *initially 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 shift-by-pod fixed effect $\eta_{sp(j)}$, capturing the combination of shift s 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.[20] $\mathbf{X_j}$ denotes a vector of patient characteristics. In some specifications, we include doctor fixed effects, thereby exploiting within-doctor variation in contractor status. Standard errors are two-way clustered at the doctor and shift-by-pod level.

Our identification strategy rests on the assumption that triage nurses do not system-

[19]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.

[20]In Section 6, 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.

atically 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 at the ED, 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 693 for length of stay (Panel A) and 4,874 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 contractor status (Figure 2). The F-statistic of joint significance in a regression of contractor status on patient characteristics is just 1.26. 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.

The balance tests in Figure 2 provide 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 Effects on Resource Use

Average Differences Table 2, Panel A reports average differences in resource use between contractors and employees, conditional on working in the same pod and shift.[21] On average,

[21] 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

contractors use approximately 7 log points more resources per patient and discharge patients 27 log points more slowly. Panel B shows that these differences in cost and length of stay remain virtually unchanged after controlling for the full set of patient characteristics used in Figure 1.

Additional Tests of Identification Figure 3 expands on the comparison between Panels A and B 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 and mode of arrival. These subsets are highly predictive of ED cost and length of stay, as shown in Figure 1. We estimate $2^6 = 64$ separate regressions, each controlling for a different combination of these subsets. The x-axis of 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, differential selection of incidents across contractor an employee status is not materially 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 reported in Figure 1. We then regress these predicted outcomes on contractor status and plot the results alongside the corresponding relationships for the actual outcomes.[22] 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 reflect differences in how contractors and employees manage their assigned incidents, rather than systematic differences in the types of incidents assigned to each group.

for each incident. In some 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 by choosing the most used pods first. The resulting coefficients are very similar.

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

Accounting for Within-Shift Timing of Assigned Cases A remaining 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 patients assigned to contractors.[23] 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 contractor status. We disentangle these mechanisms by exploiting the fourteen contractors who switched to employee status within our sample (see Table A1 for the full transition matrix). We augment our specification with individual fixed effects, allowing us to compare a doctor's performance across different contractual arrangements. The estimates of the contractor dummy, reported in Table 2 Panel D, are very similar to the Panel C estimates without doctor fixed effects.

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 doctors who switch are older and more experienced during periods when they are observed as employees. While individual fixed effects absorb 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, statistically significant differences in cost and length of stay. Specifically, the coefficient for cost increases marginally from 9 to 10 log points, whereas the coefficient for length of stay decreases 8 log points but remains large at 24 log points. We therefore conclude that the estimated differences between employees and contractors are driven primarily by the treatment effect of contractual status rather than by selection or human capital accumulation.

[23] 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 the timing in which the case is assigned. We add controls for the time remaining in the shift and the number of cases previously assigned to the doctor. 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 nonetheless reassuring that even its inclusion leaves the estimated coefficients essentially unchanged.

Addressing Residual Differences in Working Conditions Our empirical strategy includes shift-by-pod fixed effects, thereby comparing doctors working under identical conditions. Moreover, employees and contractors perform precisely the same tasks, so the transition from contractor to employee entails no change in responsibilities or working conditions.

A residual 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 over the medium run, then even within-shift comparisons could partly reflect differences in recent rotation patterns rather than contractual status itself.

In Figure A2, we examine whether shift and pod assignments differ systematically by contractual status. We find statistically significant but economically negligible correlations, and these are inconsistent in sign. 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 prior 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 prior rotation patterns are included in the baseline specification.

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

Evaluated at the sample means, 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-by-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 .25 for length of stay. This implies that the estimated difference in Table 2 is roughly equivalent to replacing a median doctor with one who is around one standard deviation above the median in the length-of-stay distribution. For cost, the estimated difference in Table 2 is equivalent to replacing a median doctor with one who is three-fourths of a standard deviation above the median.

A second benchmark is provided by 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 larger effects *within the same 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 to estimate lead and lag effects around these transitions. Our baseline estimating equation is:

$$y_j = \sum_{\tau=-4}^{4} \pi_\tau \left(transition_{i(j)} \times year_\tau(i,j) \right) + \kappa_{i(j)} + \eta_{sp(j)} + \theta_{h(j)} + \mathbf{X}_j' \lambda + \varepsilon_j, \quad (1)$$

where $transition_{i(j)} = 1$ if doctor i is a transitioning doctor, and $year_\tau(i,j) = 1$ indicates that observation j falls in year τ relative to the date in which i transitioned. $\kappa_{i(j)}$, $\eta_{sp(j)}$ and $\theta_{h(j)}$ are doctor, shift-by-pod and exact hour fixed effects, respectively. $\mathbf{X}_j$ includes the baseline patient characteristics. Equation (1) is estimated on the baseline sample.

Equation (1) is potentially subject to the concerns about standard difference-in-differences (DiD) estimation raised by a recent and influential literature (e.g., Roth et al., 2023). The staggered-adoption estimators proposed in response (e.g., Callaway and Sant'Anna 2021, Sun and Abraham 2021, Borusyak et al. 2024) are not well suited to our setting, however, as they require sufficient overlap between switchers and a clean control group of never-treated or not-yet-treated units within the relevant fixed-effect cells. In our data, never-treated doctors (i.e., the always contractors in Table A1) number only four, do not span the full sample period, and yield very limited overlap with switchers once shift-by-pod fixed effects are absorbed. Not-yet-treated observations are similarly sparse in pre-transition windows.

To alleviate these concerns, we construct a stacked DiD estimator in which we select the control group to minimise the potential for contamination. The already-treated group comprises two types of doctor: those who were always employees and those who previously transitioned from contractor status. The latter are particularly problematic, as they may still be on a post-transition adjustment path, and their outcomes could therefore reflect an evolving treatment effect rather than a common time trend (Goodman-Bacon, 2021). We address this by restricting the control pool to always-employees and to doctors observed at least five years after their own transition, a group we call *stabilised* employees. The identifying assumption underlying this approach is that, beyond this horizon, post-transition adjustment is complete, and the outcomes of stabilised employees thus reflect common time shocks rather than their own continuing treatment dynamics. Appendix B examines the validity of this assumption directly, showing that: (a) re-estimating the event study with this stacked approach yields very similar results to the baseline specification in equation (1), and (b) stabilised employees and the small group of clean controls available in our sample behave similarly over time, suggesting that the stabilised group represents a reasonable substitute for the clean control group.[24]

Figure 5 reports estimates from the baseline specification (1) alongside two stacked alternatives that differ in the composition of the control group: one using only the stabilised employees described above, and one augmenting that group with the small number of clean controls available in the data. For length of stay we observe a clear and economically meaningful pattern: coefficients drop by about 21 log points in the year immediately after transition, an effect which grows to roughly 40 log points by years 3–5. The cost estimates display the same qualitative (albeit less precise) pattern, with effects of up to 10 log points. The estimated dynamics are essentially identical across the baseline and stacked specifications, suggesting that the baseline estimates are predominantly identified from the stabilised group.[25] Taken together, the evidence points to a fast and sustained reduction in resource use after doctors transition from contractor to employee status, reflected in both lower costs and shorter patient stays.[26]

5 Effects on Other Variables

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 in-ED mortality. We also construct indicators

[24]Specifically, Appendix B compares stabilized employees to the limited set of never- and not-yet-treated doctors over time periods in which both groups are observed. The pre-period trend differential between the two groups is small and jointly insignificant, supporting the use of stabilized employees as reasonable alternatives in those shift-by-pod cells in which clean observations are too sparse to contribute to identification.

[25]Although some lead coefficients are statistically different from zero Appendix B shows that when pre-trends are tested on the pre-period alone, as recommended by Borusyak et al. (2024), the lead coefficients are smaller in size and jointly insignificant (Figure A3).

[26]Appendix B reports two further robustness checks. First, augmenting the baseline specification with controls for log age and the log cumulative number of cases leaves the estimates essentially unchanged, addressing the concern that the estimated transition effect could capture life-cycle or human capital accumulation around the transition date (Figure A3). Second, a modified imputation estimator in the spirit of Borusyak et al. (2024), using always-employee observations and pre-transition observations of switchers as imputation donors, yields the same qualitative dynamic profile (Figure A6).

for patients who return to the ED within 2, 5, 30 and 60 days.[27] The bottom row of Table 3 reports the mean of each outcome variable, confirming 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 30 days. The exception is in-ED mortality, which is a rare outcome.

We present empirical specifications with and without doctor fixed effects. Across almost all models, we find no evidence of an effect of contractor status on patient outcomes. An exception is that contractors are associated with 1-2 percentage points higher likelihood of hospitalisation. Hospitalisation involves the use of additional hospital resources, and in this sense this finding is consistent with the Table 2 differences between employees and contractors.[28] Nevertheless, the relatively small size of the coefficients leads us to treat this finding with caution. Instead, we conservatively conclude that contractors spend more resources on their patients without achieving detectably better patient outcomes.

Why would contractors take longer and use more resources without improving patient health? A potential explanation for these findings is that ED doctors are operating in the 'flat of the curve' region of the health function linking treatment intensity with patient outcomes (Doyle 2011, Silver 2021). If this function has diminishing or even zero returns, a large increase in treatment intensity may not translate into measurably better patient outcomes.

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, additional observation time while the patient occupies a bed, etc.) are substitutes in the health production function, contractors may achieve similar outcomes with a different combination of inputs. While we cannot directly observe cognitive effort, we provide evidence below that contractors are less willing than employees to work unpaid overtime.

Effects on Selected Cost Items The total cost of treating a patient in the ED is the sum of all procedures, consultations, and tests performed. Tables A7 and A8 list the thirty most

[27]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.

[28]The finding that hospitalisation and in-ED resource use go hand-in-hand is consistent with previous studies. Handel et al. (2026), for instance, show that ED doctors that wish to exert lower cognitive effort (because they are more tired) order more tests and are more likely to hospitalise patients. Similarly, Chan (2018) finds that ED doctors that approach the end of their shift increase both hospitalisations and diagnostic tests.

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 classified as either laboratory or non-laboratory, following the Niguarda data system.

Laboratory tests are ordered frequently, but their low unit cost means they are under-represented 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. In Panel A, contractors are associated with 46 log points higher spending on laboratory tests and 10 log points 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 than on ED cost.

The remaining columns of Table 4 report results by selected test types, revealing large and statistically significant differences for several categories. In Panel A, contractors are 2.4 percentage points more likely to order an electrocardiogram, which is a 9% increase relative to the unconditional mean of 28%. Similarly, contractors are 10 percentage points more likely to order an X-ray, representing a 16% increase on the unconditional mean of 62%. 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, asking 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 slightly more likely to request that opinion, although the coefficient becomes non-statistically significant when controlling for doctor fixed effects.

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, such transfers are routine at

Niguarda.[29] In our sample, 35% of incidents are handed over to a doctor in the next shift (bottom row in 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 A, contractors are 7.6 percentage points more likely to hand over a patient to the next-shift doctor, an economically large effect relative to a mean of 35%.[30] To understand the implications of this estimated difference, note that handovers, though routine, entail non-trivial costs (Apker et al. 2007, Chan 2018). 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.[31] 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, the literature generally views handovers as an efficiency loss and recommends minimising them where possible (Phillips et al., 2015).[32]

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 because the incoming doctor wishes to reassert diagnostic autonomy. While such additional testing will be clinically justified in many cases, it can also reflect 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 .71 duplicated cost items, and Table A9 lists the thirty most common of these. The presence of duplicated items should not automatically be interpreted as wasteful,

[29] 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.

[30] 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.

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

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

as it could reflect standard clinical practice.[33] However, duplicated items being more common in handed-over incidents could be interpreted as reflecting unnecessary resource use.

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 have 44 log points more duplicated cost items. We also find large coefficients 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 28 log points more duplicated costs. Panel B of Table A10 shows that handovers are associated not just with a higher absolute number of duplicated costs but also with higher 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. We find that this share is 5 percentage points higher for incidents with handovers. This coefficient more than doubles the average share of 3%. Table A10 should not be interpreted as identifying causal effects, but instead aiming at establishing correlations between variables. Nevertheless, we interpret Table A10 as consistent with the additional duplications associated with handovers being the result of coordination frictions leading to unnecessary resource use.[34]

Column 2 Table 5 established that contractors generate more handovers. Column 3 of Panel A Table 5 now shows that contractors generate .17 more duplicated cost items than employees, relative to a mean of .71. Column 4 of Panel A shows that duplicated costs represent a .7% higher share of total costs for contractors, an economically significant difference relative to the unconditional mean of 3%. The estimates are even larger in the Panel B specifications with doctor fixed effects. We interpret these differences as suggestive evidence of a less efficient use of resources in incidents initially assigned to contractors, potentially operating through their higher propensity to hand over patients to the incoming doctors in the next shift.

[33] 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.

[34] A natural confounding factor 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.*

Externalities on Employees Sharing the Same Pod We now investigate whether contractors are assigned fewer cases over the course of a shift. The hypothesis follows naturally from two features of our setting: triage nurses assign cases based on the existing doctor caseloads, and contractors' longer average length of stay means they clear cases more slowly. Using a doctor-by-shift dataset, we find in Table 6 that contractors are assigned 8–17% fewer cases. Columns 2 and 3 reweigh cases by predicted length of stay and cost, respectively. The estimates are very similar.

We then investigate whether employees' resource use and length of stay are affected by sharing a pod with a contractor. Table A11 examines this using a sample restricted to incidents assigned to employees, excluding cases assigned to contractors or students. 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-by-pod level, shift-by-pod fixed effects cannot be included, and the estimates should be interpreted with caution, as case assignment to pods is not exogenous. That said, the coefficients are small in magnitude and statistically insignificant, which we take as consistent with the assumption that employees do not adjust their resource use depending on whether they share a pod with contractors rather than other employees.[35]

6 Mechanisms

In the previous sections, we documented three main findings: (a) contractors use more resources than employees, (b) these higher inputs do not translate into detectably better patient outcomes, and (c) contractors generate more handovers and duplicated tests, likely reflecting at least in part inefficiency in resource use.

In this section, we consider two broad explanations for these performance differences: differences in ability and differences in motivation. We argue that motivation is the primary driver of the gap. We then uncover evidence in support of a specific mechanism: employees exert more effort on behalf of their fellow employees (including but not exclusively through overstaying), and this helps explain their superior performance relative to contractors.

Ability and Human Capital A natural explanation for the performance gap is that contractors may have lower diagnostic ability and therefore require more time and resources to achieve comparable patient outcomes. Three pieces of evidence suggest, however, that

[35]Table A11 is consistent with the SUTVA assumption that potential outcomes for employee-treated patients are unaffected by the treatment assigned to other patients.

ability differences cannot account for the documented patterns. 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 remain economically and statistically significant even after controlling for doctor age, experience and tenure. Third, Figure 5 shows that doctors reduce their length of stay *discontinuously* when transitioning from contractor to employee status. The fact that human capital cannot be acquired overnight implies that it cannot account for the immediate behavioural change documented in Figure 5. We conclude that differences in ability and human capital play at most a minor role in explaining the resource use differences we document, and that the primary explanation must lie elsewhere.

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 operational objectives, such as maintaining patient 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 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 within the sixty minutes after that doctor's shift has ended. This indicator captures the idea that the originally assigned doctor remains on duty beyond their scheduled hours 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, 35% of all incidents are handed over to a doctor in the next shift. The remaining 61% of incidents are discharged by the initially assigned doctor 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, doctor pairs with greater similarity in age or gender. This pattern is consistent with the well-documented finding that individuals are more willing to exert effort on behalf of those with whom they share observable characteristics (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 (i.e., the *focal* outgoing doctor) and (b) the average male dummy for the doctors working in the same pod during the following shift (i.e., the *incoming* doctors who will inherit the patient if the focal doctor hands it over instead of overstaying).[36] The resulting variable is the absolute difference between (a) and (b), with higher values indicating greater gender dissimilarity between the focal doctor and the incoming group. 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 the likelihood of overstaying. We use our baseline empirical specification, which includes controls for the shift-by-pod combination and patient characteristics. In addition, we include fixed effects for the identity of the focal 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-by-pod fixed effects. This specification therefore identifies whether *the same doctor* exhibits different willingness to overstay depending on their 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 when sharing the same gender as the incoming doctors, and .5 percentage points more likely when being the same age, relative to being ten years apart. These magnitudes are substantial relative to the unconditional mean of 3.7%. Overall, the empirical patterns are consistent with the overstay variable being a valid proxy for discretionary effort.

Average Differences in Overstay between Employees and Contractors We now examine whether employees and contractors differ in their willingness to overstay. We employ our baseline specification, which includes shift-by-pod fixed effects and controls for patient characteristics. For ease of interpretation, we use an employee indicator as the main independent variable, with contractors as the base group.[37]

[36] We use the average male dummy for next-shift doctors because we do not observe which specific doctor will inherit the patient if the focal doctor does not overstay. To the extent that the focal 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.

[37] A small number of incidents are assigned to medical students. All regressions control for student status. Therefore, when the contractor dummy is included, employees form the base group, and vice versa.

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 employment 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 they work as a contractor. This finding is inconsistent with explanations based on time-invariant differences in altruism or cost of effort and instead supports interpreting the employment status coefficient as a treatment effect. Column 4 further shows that this difference remains robust to controlling for the homophily variables introduced in the previous subsection. The magnitude of the effect is economically meaningful: the coefficient of .013 represents around a third of the overstay rate of .037.[38]

Evidence on Overstaying to Help Other Employees We now examine more closely the mechanism through which employment status affects overstay.

Identifying the specific factor that causes employees to overstay is challenging because employee status bundles together a wide range of characteristics. For instance, employees attend the monthly meetings where hospital-wide objectives and performance are discussed, while contractors do not. These meetings serve to communicate hospital-wide objectives and instil a sense of community among employees. Second, employees and contractors differ in their average weekly hours worked (Table 1), which could influence the marginal cost of exerting effort during overstay episodes (although Tables A4 and A5 suggest this channel is unlikely to be quantitatively important). Third, the two groups may face different monetary and career incentives.[39] Fourth, contractors and employees differ in who pays for liability insurance for cases of minor negligence. Although both types of doctors are covered, contractors (but not employees) have to pay for that cost themselves. 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 whether they depend on the

[38]The finding that employees overstay more than contractors is particularly striking given that employees are more likely to have discharged the patient 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 specifications in Table 7 use 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 119,584 as opposed to 305,221. The average value of overstay is 9.1% in this subsample. Unsurprisingly, we estimate much larger coefficients, both in absolute terms and relative to the average likelihood of overstaying.

[39]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.

contractual status of the incoming doctors. Explanations based on differences in the cost of liability insurance, incentives, the marginal cost of effort, or attachment to institutional objectives are unrelated to the contractual status of the doctors potentially inheriting the patient. 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. We have, indeed, already found in Column 1 Table 7 that focal doctors are more likely to overstay when they share gender and age with incoming doctors. We now investigate whether homophily also matters along the employee status dimension.

In Column 5 of Table 7, we interact the employee indicator with the share of employees among 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 predominantly contractors. 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 focal doctor fixed effects and for the gender and age homophily variables.

In Table A13 we confirm that the higher willingness to overstay among employees is driven by the contractual status of the incoming doctors, as opposed to other shift-by-pod characteristics that may be correlated with incoming doctor composition. Specifically, we control incrementally for the interactions between the employee dummy and a large number of shift-by-pod characteristics. We find that the coefficient on the interaction between employee status and the share of 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 fellow employees.[40]

Evidence on Lower Resource Use to Help Other Employees We now return to the baseline resource use estimates to assess how much of the performance gap can be attributed to the differential behaviour of employees when the incoming doctors are also employees.

[40]In Table 7, we calculate the share of incoming employees 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 throughout.

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. An employee who anticipates not discharging the patient before the end of their shift may nonetheless exert additional cognitive effort in diagnosis and treatment, so that the incoming doctor inherits a case that is closer to resolution.

In columns 1 and 2 Table A15, we reproduce the baseline estimates from Panels C and D Table 2. The estimates are identical in magnitude but opposite in 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 employees among the doctors working in the same pod during the following shift. In columns 4 and 5, we progressively control for doctor fixed effects and the homophily variables. We find that employees use substantially fewer resources, especially when they are followed by incoming employees in the same shift and pod. For instance, relative to contractors, length of stay is 20 log points lower for employees followed by contractors, but 34 log points lower for employees followed by employees. Similarly, relative to contractors, ED cost is 6 log points lower for employees followed by contractors, but 11 log points lower for employees followed by employees.[41]

Overall, we conclude that part of the reason employees use fewer resources than contractors is that they exert additional effort on behalf of the fellow employees who will inherit their cases, both by overstaying more often and by handing off patients who are closer to discharge.

Interpretation of Mechanism Table 7 shows that employees overstay more than contractors, *but only when followed by other employees*. Similarly, Table A15 shows that resource use is lower for employees, *but especially when followed by other employees*. This finding is inconsistent with a wide range of natural explanations for the differential behaviour of employees and contractors, including (a) differences in liability insurance coverage, (b) differences in monetary and career incentives, (c) differences in the marginal cost of effort, and (d) differences in the overall commitment to the hospital's objectives. In their natural formulations, these factors do not operate through a channel that depends on the contractual

[41] The interactions in the within-doctor regressions of columns 4 and 5, while economically large, are not or only marginally statistically significant, although the interactions in the across-doctor regressions of column 3 are.

status of the incoming doctors.[42]

Instead, our findings suggest that becoming an employee increases a doctor's willingness to cooperate with fellow employees but not with contractors. We now discuss two potential micro-foundations for this increased willingness: (a) the adoption of an insider social identity (Akerlof and Kranton, 2005) and (b) reciprocity and favour exchange among employees (Neilson, 1999).

A large theoretical and experimental literature on social identity has argued that individuals identifying with a group are willing to exert costly effort to benefit fellow members (Charness and Chen, 2020). For instance, Akerlof and Kranton (2005) posit that workers adopt one of two identities within organisations, with insiders experiencing a disutility from deviating from a high-effort norm that outsiders do not face. Chen and Li (2009) find in a laboratory experiment that cooperative behaviour is higher towards members of an artificially generated in-group than towards members of the out-group. An important open question is what determines, within organisations, identification with a specific social group (Akerlof, 2007).[43] One interpretation of Table 7 is that *employment status shapes social identity*, casting contractors as outsiders and employees as insiders. Although employees and contractors perform identical clinical tasks in the ED, this socialisation process may arise from their differential integration into the hospital's organisational culture, as documented in Section 2. These differing identities, in turn, may lead doctors to behave differently depending on the contractual status of the doctor who will inherit their cases.

A related but somewhat different micro-foundation is that employees reciprocate or exchange favours with one another, while excluding contractors from this pattern of mutual cooperation. To understand this mechanism, consider an idealised case in which two doctors, A and B, alternate in who follows whom across shifts: A follows B in half of their overlapping shifts, and B follows A in the other half. In this ideal scenario, A and B might find it relatively easy to sustain a bilateral cooperation equilibrium, in which each doctor overstays or generally exerts higher effort in the expectation that the other will reciprocate when roles are reversed. In Niguarda, differences in overstay might reflect the fact that employees have rotation patterns that are more conducive to the sustained cooperation described above.

[42]In a model with sufficient degrees of freedom, one can hypothesise more sophisticated mechanisms in which career incentives depend on the contractual status of the incoming doctors. For instance, employees may want to gain a reputation for being cooperative among their fellow employees, expecting this reputation to somehow influence future promotion decisions. For a recent paper exploring empirically the interaction between peer and managerial pressure, see Battiston et al. (2025).

[43]Evidence from laboratory settings suggests both that natural identities such as gender, ethnicity or race are powerful predictors of behaviour (Bernhard et al. 2006, Goette et al. 2006) and that identity is malleable and can be primed relatively easily (Chen and Li, 2009).

We do not find evidence of these differential rotation patterns. The likelihood of overlapping shifts is not higher for employees than for contractors, and measures of overlap are not correlated with the decision to overstay.[44] However, we do not regard this as conclusive evidence that reciprocity plays no role in explaining the higher propensity to overstay by employees. First, it is possible that at least some favours traded between employees take place outside of the overstay decision. Second, favour exchange might not be bilateral but might instead operate across the whole in-group (Neilson, 1999). Last, testing theories of reciprocity in long-term relations is notoriously difficult, as key variables such as the value of the relationship and off-equilibrium payoffs are typically unobserved (Gil and Zanarone, 2017).

We conclude this section by noting that these two micro-foundations are not necessarily independent of each other. Gibbons and Henderson (2013), for instance, argue that many organisational attributes that underpin high performance, such as high-effort norms, organisational culture and the social identity of workers, can be understood through the lens of relational contracts. Disentangling the specific micro-foundations of the enhanced willingness of employees to cooperate with other employees is beyond the scope of this paper, though we regard it as a fertile area of future research.

7 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 these differences at least partially reflect the higher motivation of employees to cooperate with one another.

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

[44] We compute measures of scheduling overlap for the focal doctor and the incoming doctor(s). *Forward overlap* measures the frequency with which the incoming doctor follows the focal doctor in the focal doctor's other shifts. *Backward overlap* measures the frequency with which the focal doctor follows the incoming doctor in the focal doctor's other shifts. *Horizontal overlap* measures the two doctors working the same shift in the focal doctor's other shifts. We compute these three measures using the focal doctor's shifts in both the preceding and the following year. Figures A7 and A8 display the distribution of overlap measures, for the preceding and following year respectively. The forward and backward overlap measures have an average of 5%, while the horizontal overlap, which is computed over a larger pool of co-workers, averages 16%. If bilateral favour-exchange 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), mostly do not predict the overstay decision (Table A17), and controlling for them leaves the employee × incoming-employee coefficient in Table 7 virtually unchanged (Table A18).

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 very similar in size) within-doctor results are based on a relatively small number of individuals (56 and 14, respectively). We discuss external validity in the remainder of this section.

The modest role of selection effects indicates that, in our setting, doctors operating under contractor status are not particularly negatively selected. In Table 2, for instance, the coefficients barely change when doctor fixed effects are added. This likely reflects the specific institutional circumstances that drove the rise and subsequent decline of contractor doctors in Italy. Specifically, the fact that the Italian central government imposed a freeze on public-sector hiring between 2010 and 2018. Many able doctors who would otherwise have easily qualified for permanent employment were consequently forced into independent contractor positions. We therefore caution against generalising this finding: in other settings, independent contractors may be more negatively — or in some cases even positively — selected.

We view the treatment effects we estimate as a 'proof of concept' that contractor status can affect productivity, independently of selection and 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.[45] At the same time, it is noteworthy that these differences persist even in a context where contractors are not significantly less able and are often on the cusp of transitioning into permanent employment.

Because contractual status bundles together multiple 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). Comparisons across settings that vary in the salience of mission, incentives, and institutional cohesion would therefore be informative.

Ultimately, the extent to which our results replicate in other contexts is inherently

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

difficult to establish, both in this paper and in the broader literature. Our aim has been 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. (2007), "The Missing Motivation in Macroeconomics", *American Economic Review*, 97(1): 5–36

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.

Battiston, D., Blanes i Vidal, J., Kirchmaier, T., and Szemeredi, K. (2025), "Peer Pressure and Manager Pressure in Organisations", *Working Paper*.

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

Bernhard, H., Fehr, E., and Fischbacher, U. (2006), "Group Affiliation and Altruistic Norm Enforcement", *American Economic Review*, 96(2): 217–221.

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.

Borusyak, K., Jaravel, X., and Spiess, J. (2024), "Revisiting Event-Study Designs: Robust and Efficient Estimation", *Review of Economic Studies*, 91(6): 3253–3285.

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

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.

Charness, G., and Chen, Y. (2020), "Social Identity, Group Behavior, and Teams", *Annual Review of Economics*, 12(1): 691–713.

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., Harrington, E., and Pallais, A. (2023), "The Power of Proximity to Coworkers: Training for Tomorrow or Productivity Today?", NBER Working Paper No.

31880.

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

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.

Gil, R., and Zanarone, G. (2017), "Formal and Informal Contracting: Theory and Evidence", *Annual Review of Law and Social Science*, 13: 141–159.

Goette, L., Huffman, D., and Meier, S. (2006), "The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups", *American Economic Review*, 96(2): 212–216.

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.

Goodman-Bacon, A. (2021), "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 225(2): 254–277.

Handel, B. R., Heizlsperger, L., Knecht, J., Kolstad, J. T., Malmendier, U., and Matějka, F. (2026), "Thinking versus Doing: Cognitive Capacity, Decision Making and Medical Diagnosis", NBER Working Paper No. 35034.

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.

Neilson, W. S. (1999), "The Economics of Favors", *Journal of Economic Behavior and Organization*, 39(4): 387–397.

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.

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.

Roth, J., Sant'Anna, P. H. C., Bilinski, A., and Poe, J. (2023), "A Guide to Event Studies and Difference-in-Differences with Heterogeneous Treatment Effects," *Journal of Econometrics*, 235(2): 2–20.

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.

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

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 (€)	162	159	163	217
Length of Stay (minutes)	280	274	292	352
Hospitalisation Dummy	.19	.19	.18	.18
Return to ED 30 Dummy	.13	.13	.13	.13
Death Dummy	.0008	.0009	.0006	.0000
Cost of Laboratory Tests (€)	46	45	48	59
Other Costs (€)	112	110	113	148
Specialist Consultation Dummy	.35	.35	.30	.58
Handover to Next Shift Dummy	.35	.35	.37	.50
Duplicated Costs Items (units)	.71	.70	.80	.85
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	305,221	261,239	30,370	13,612
Average Number of Incidents Per Doctor	3,052	5,024	1,687	454
Average Number of Shifts Per Doctor	499	797	288	108
Average Number of Hours Worked (per month)	93.5	92.6	99.0	97.4
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-by-pod		
Contractor	.073***	.269***
	(.020)	(.036)
Panel B: Adding Patient Characteristics		
Contractor	.075***	.266***
	(.017)	(.036)
Panel C: Adding Exact Hour		
Contractor	.105***	.323***
	(.018)	(.045)
Panel D: Adding Doctor Fixed Effects		
Contractor	.090***	.316***
	(.018)	(.032)
Panel E: Adding Doctor Characteristics		
Contractor	.100***	.236***
	(.020)	(.038)
Log Doctor Age	-.280	-2.356
	(.733)	(1.465)
Log Doctor Past Cases	-.003	-.016
	(.007)	(.014)
Log Doctor Tenure at Niguarda	.016*	-.013
	(.009)	(.016)

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 and referral party. 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 absorbed by the doctor fixed effect. Standard errors two-way clustered at the doctor and shift-by-pod interaction level. The number of observations is 305,221.

TABLE 3
OTHER PATIENT OUTCOMES

	Hospita-lisation	Death in ED	Return to ED 2	Return to ED 5	Return to ED 30	Return to ED 60
Panel A: Without Doctor Fixed Effects						
Contractor	.0105**	-.0003	.0008	-.0003	-.0006	-.0019
	(.0048)	(.0002)	(.0020)	(.0025)	(.0037)	(.0043)
Panel B: With Doctor Fixed Effects						
Contractor	.0196***	-.0002	-.0004	-.0008	.0001	-.0002
	(.0061)	(.0005)	(.0033)	(.0047)	(.0062)	(.0070)
Mean Variable	.1854	.0008	.0364	.0650	.1328	.1741

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 and referral party). 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. Standard errors two-way clustered at the doctor and shift-by-pod interaction level. The number of observations is 305,205.

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	.457***	.096***	.06***	.1***	.01	.024*	.013**
	(.065)	(.018)	(.007)	(.014)	(.007)	(.014)	(.007)
Panel B: Adding Doctor Fixed Effects							
Contractor	.393***	.081***	.044***	.072***	.005	.039**	.016
	(.072)	(.022)	(.009)	(.017)	(.011)	(.017)	(.013)
Mean DV	Logs	Logs	.59	.62	.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 and referral party). Standard errors two-way clustered at the doctor and shift-by-pod interaction level. The number of observations is 305,205.

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	.014***	.076***	.171***	.007***
	(.005)	(.01)	(.038)	(.001)
Panel B: Adding Doctor Fixed Effects				
Contractor	.013	.082***	.234***	.009***
	(.009)	(.011)	(.051)	(.002)
Mean Dep.	.35	.35	.71	.03

Standard errors two-way clustered at the doctor and shift-by-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 and referral party). The number of observations is 305,205.

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-by-pod			
Contractor	-.078***	-.101***	-.103***
	(.025)	(.030)	(.028)
Panel B: Adding Doctors Fixed Effects			
Contractor	-.172***	-.208***	-.204***
	(.020)	(.025)	(.025)

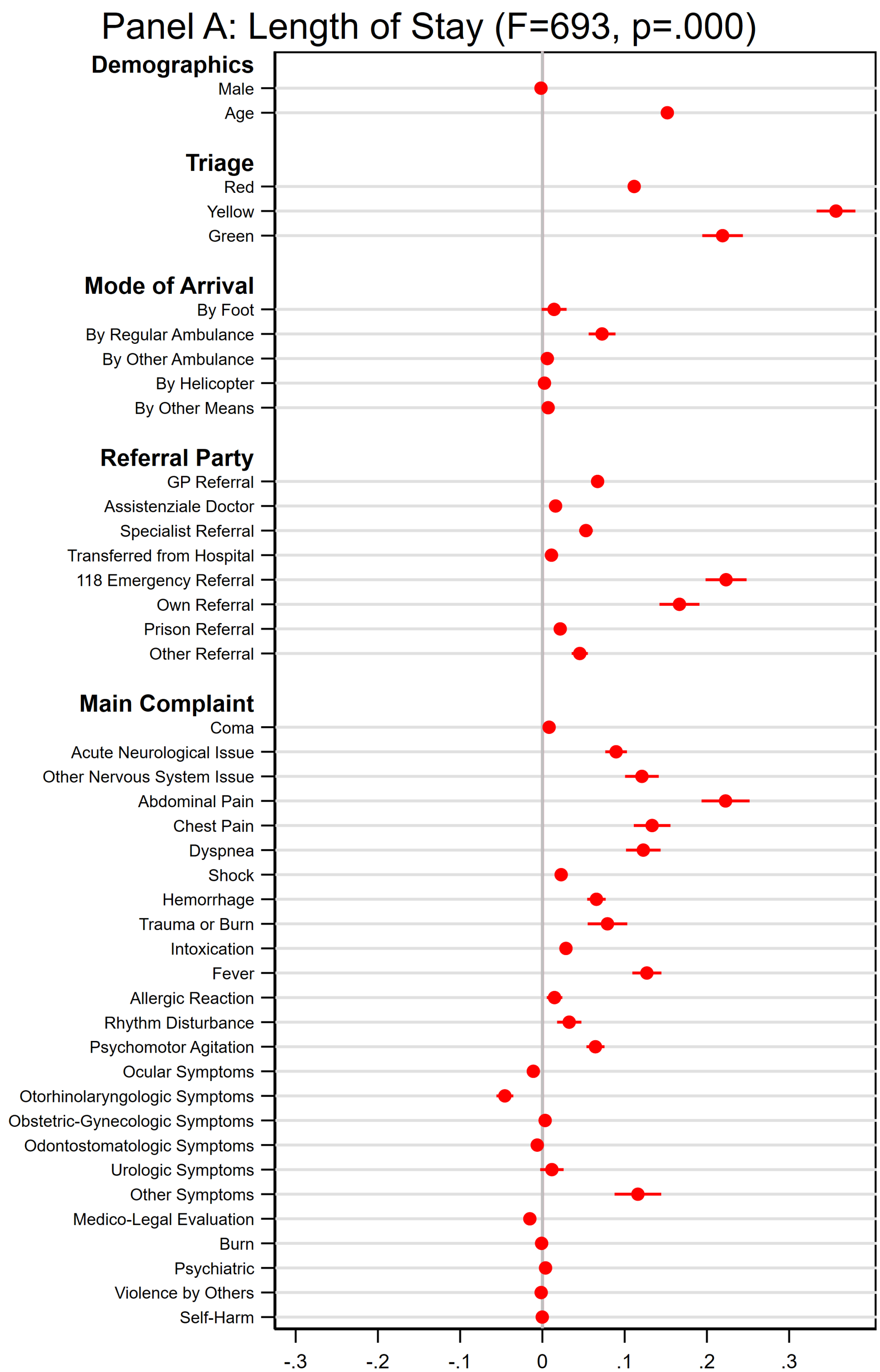
This table displays regressions of the number of incidents assigned to a doctor during a shift, on whether the doctor is a contractor. The sample is a sample of shift-by-pods and doctors. In the first column the dependent variable is the number of incidents assigned. In the second and third columns, 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-by-pod and doctor level. The number of observations is 81,681.

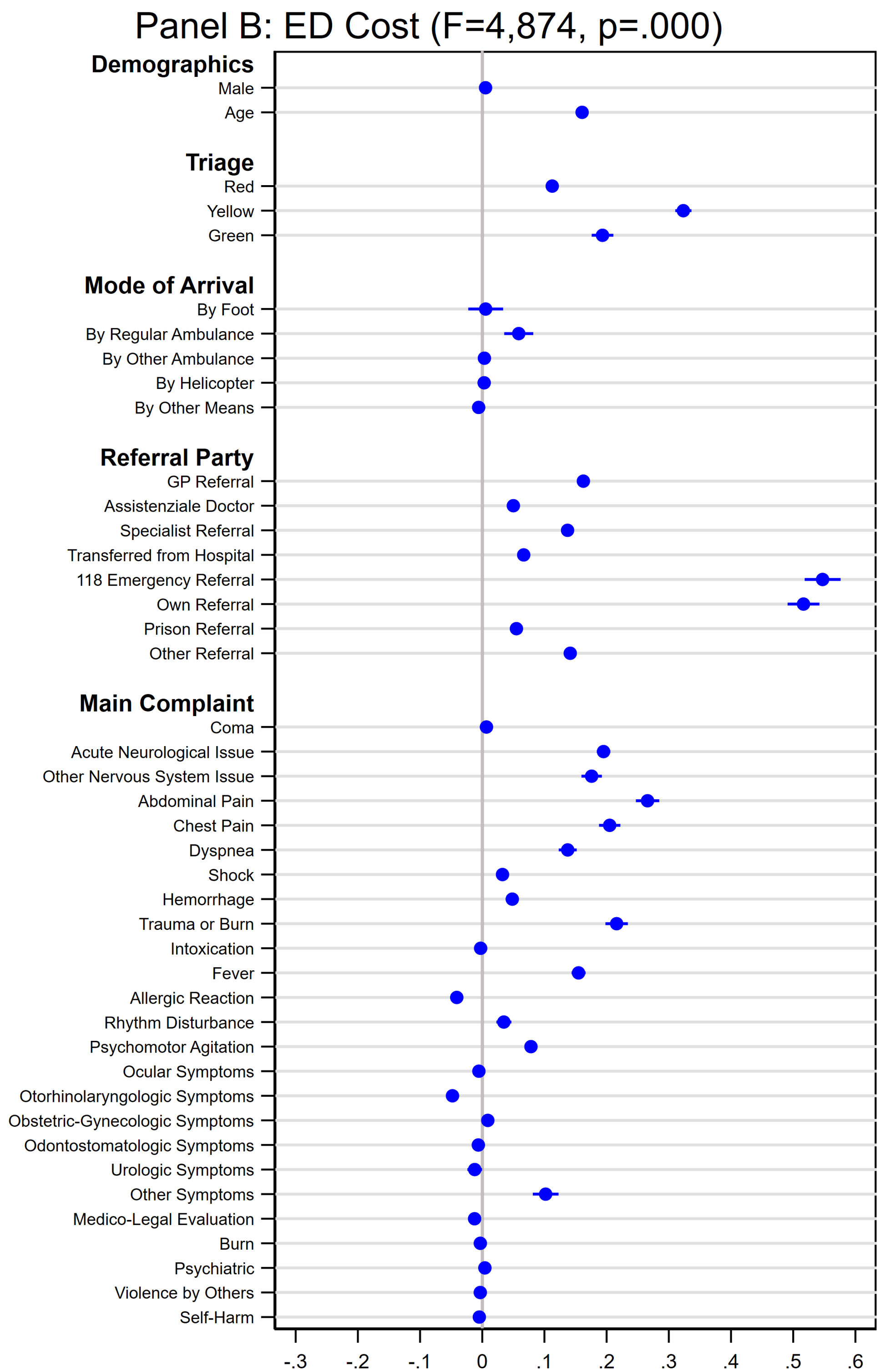
TABLE 7
MECHANISM: EFFECTS ON OVERSTAY VARIABLE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Employee		.019***	.014***	.013***	.006	.001	.002
		(.003)	(.005)	(.005)	(.006)	(.006)	(.006)
Employee × Incoming Employees					.015***	.015**	.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	.001	.004
Doctor Fixed Effects	Yes	No	Yes	Yes	No	Yes	Yes

Standard errors two-way clustered at the doctor and shift-by-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 and referral party). 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 305,169.

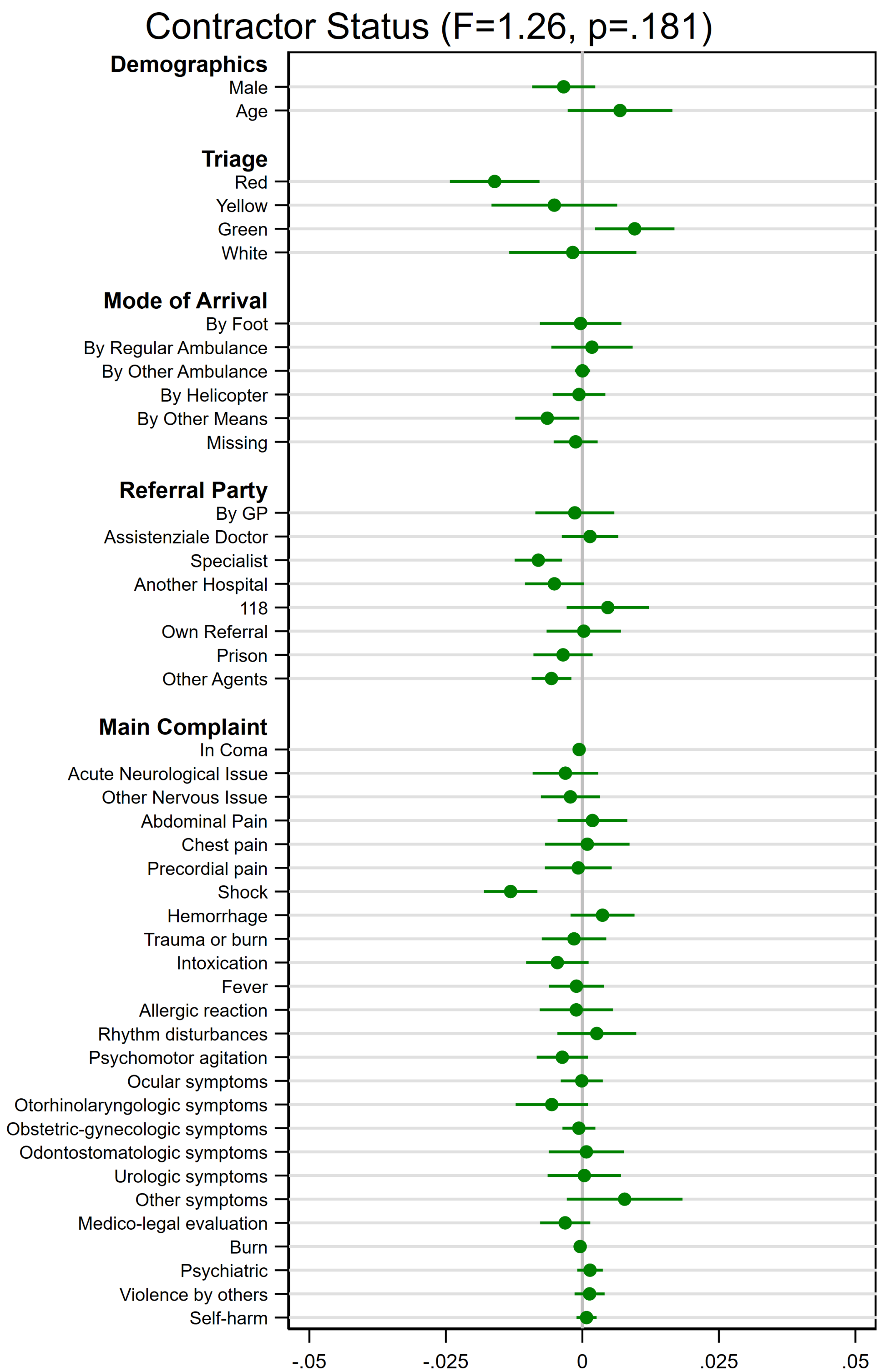
FIGURE 1: PREDICTIVE POWER OF PATIENT CHARACTERISTICS





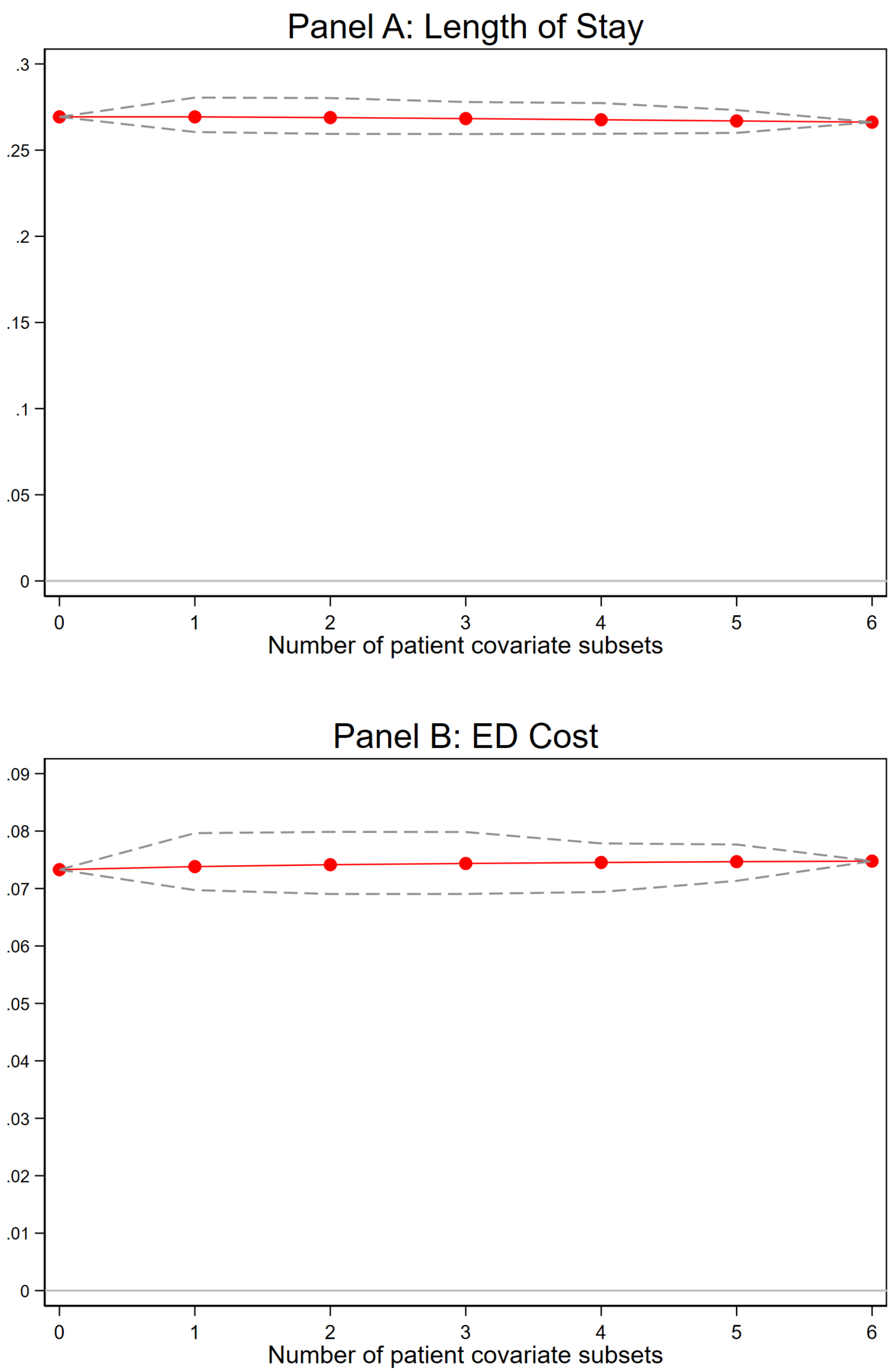
This figure displays coefficients and 95% confidence intervals from regressions of length of stay and ED cost on patient characteristics, controlling for the shift-by-pod indicators. All variables are standardised. Standard errors are clustered two-way at the doctor and shift-by-pod level. The F-statistics 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



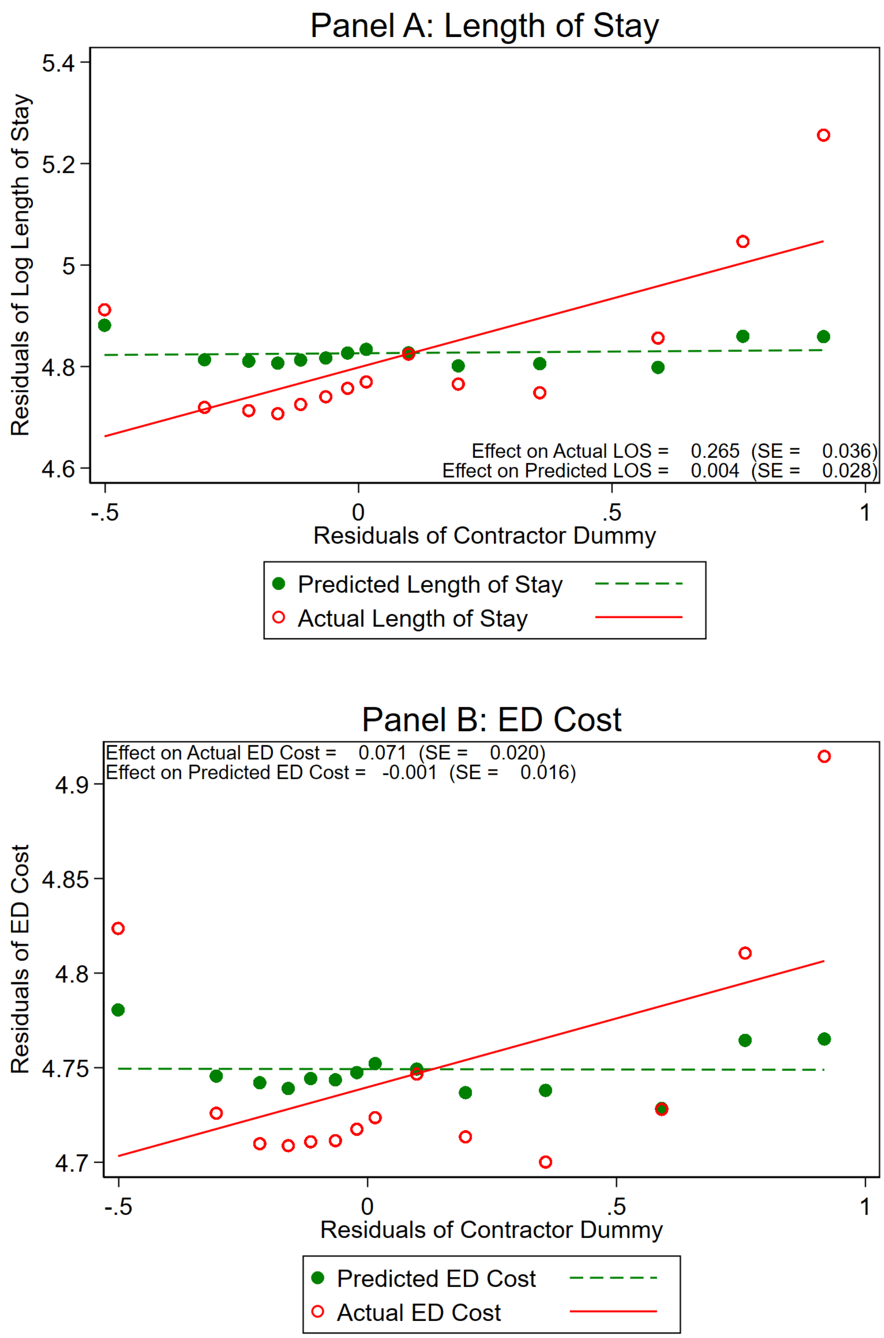
This figure displays coefficients and 95% confidence intervals from regressions of patient characteristics on contractor status, controlling for the shift-by-pod indicators. All variables are standardised. Standard errors are clustered two-way at the doctor and shift-by-pod level. The F-statistic at 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



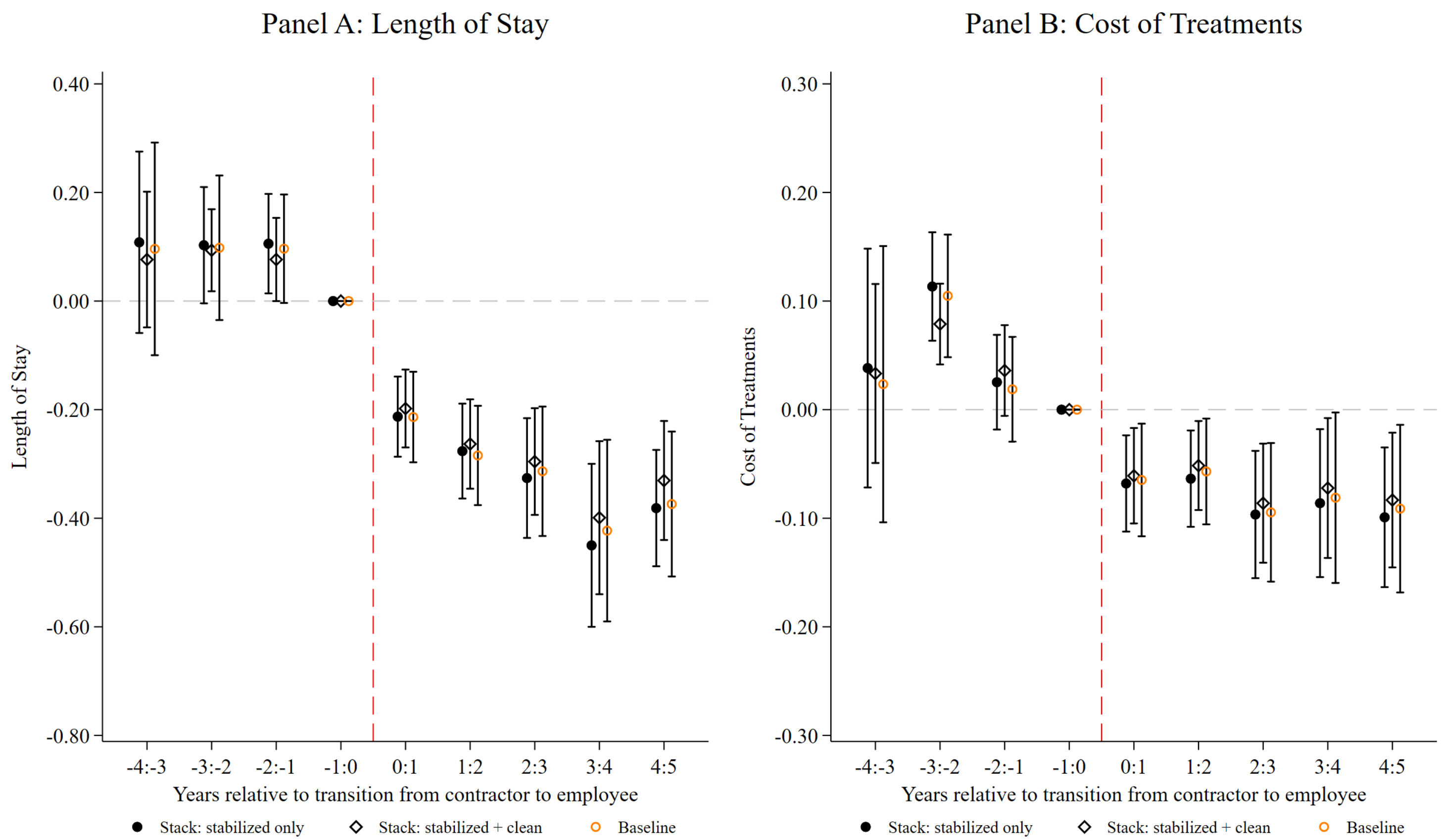
This figure displays the robustness of the coefficients to the inclusion or exclusion of different subsets of patient characteristics. Our subsets of patient characteristics are gender, age (in 50 quantiles), triage, main complaint, referral party and mode of arrival. We run separate regressions that control for each of the $2^6 = 64$ 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-by-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 and referral party). 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 shift-by-pod level), of the regressions on actual and predicted outcomes are reported inside the graphs.

FIGURE 5: EVENT STUDY AROUND THE TRANSITION



This figure reports event-study coefficients π_τ around contractor-to-employee transitions. The baseline coefficients are obtained by estimating the following specification on the baseline sample:

$$y_j = \sum_{\tau=-4}^{4} \pi_\tau \left(transition_{i(j)} \times year_\tau(i,j)\right) + \kappa_{i(j)} + \eta_{sp(j)} + \theta_{h(j)} + \mathbf{X}_j'\lambda + \varepsilon_j, \quad (1)$$

where $transition_{i(j)} = 1$ if doctor i is a transitioning doctor, and $year_\tau(i,j) = 1$ indicates that observation j falls in year τ relative to i's transition date. $\kappa_{i(j)}$, $\eta_{sp(j)}$ and $\theta_{h(j)}$ are doctor, shift-by-pod and exact hour fixed effects, respectively. $\mathbf{X}_j$ includes the baseline patient characteristics. The figure also reports two stacked variants estimated over the same event-time window. For each contractor-to-employee transition, we construct an episode centred on the transition date, retain the focal doctor's event-time observations, and pool them with control observations drawn from the same calendar-time window (four years before to five years after the transition). The episodes are then stacked. In these stacked specifications, we include episode fixed effects and weight observations by the inverse of the number of episode windows in which they appear. The control pool consists of either (i) stabilised employees—doctors always observed as employees in our sample, plus former contractors observed at least five years after their own transition—or (ii) stabilised employees plus clean controls (i.e., never-treated and not-yet-treated observed at least one year before their own transition). All specifications exclude students. Standard errors are two-way clustered at the doctor and shift-by-pod levels.

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	
	(1)	(2)	(3)	(4)
	Original	Reassigned	Original	Reassigned
Panel A: Without Doctor Fixed Effects				
Contractor	.105***	.100***	.323***	.323***
	(.018)	(.020)	(.045)	(.041)
Panel B: With Doctor Fixed Effects				
Contractor	.090***	.064***	.316***	.229***
	(.018)	(.017)	(.032)	(.032)

Columns 1 and 3 are equivalent to Columns 1 and 2 in Panels C and D Table 2. In the original sample, sometimes a doctor appears in more than one pod during a shift. In Columns 2 and 4 we repeat the baseline regressions after reassigning 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 302,202.

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	.105***	.105***	.096***	.323***	.323***	.304***
	(.018)	(.018)	(.018)	(.045)	(.045)	(.043)
Hours to End of Shift		.008	-.011*		.004	-.034***
		(.007)	(.006)		(.013)	(.012)
Cases Previously Assigned in Shift			-.011***			-.023***
			(.001)			(.003)
Panel B: With Doctor Fixed Effects						
Contractor	.090***	.085***	.085***	.316***	.304***	.304***
	(.018)	(.018)	(.018)	(.032)	(.032)	(.032)
Hours to End of Shift		-.005	-.005		-.023*	-.023*
		(.006)	(.006)		(.012)	(.012)
Cases Previously Assigned in Shift			-.007***			-.017***
			(.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 and referral party), and shift-by-pod indicators. Panel B further controls for doctor fixed effects and exact hour (year $\times$ month $\times$ day $\times$ hour). Columns 1 and 4 are equivalent to Columns 1 and 2 Panels C and D Table 2. Standard errors two-way clustered at the doctor and shift-by-pod level. The number of observations is 305,205.

TABLE A4
BASELINE CONTROLLING FOR RECENT ROTATION PATTERNS

	ED Cost		Length of Stay	
	(1)	(2)	(3)	(4)
Contractor	.105***	.104***	.323***	.315***
	(.018)	(.017)	(.045)	(.037)
Total Shifts		-.005		.001
		(.006)		(.015)
Morning Shifts		-.006		-.018**
		(.004)		(.009)
Afternoon Shifts		.003		.001
		(.004)		(.010)
M1 Shifts		.011***		.004
		(.004)		(.009)
M2 Shifts		.004		-.013
		(.004)		(.009)
M3 Shifts		.001		-.007
		(.005)		(.010)
M4 Shifts		.004		-.007
		(.005)		(.011)
Medica Shifts		.006		.008
		(.004)		(.009)
Monday Shifts		-.002		-.008
		(.005)		(.010)
Tuesday Shifts		.003		.014
		(.006)		(.012)
Wednesday Shifts		-.002		-.004
		(.006)		(.012)
Thursday Shifts		.006		.007
		(.006)		(.011)
Friday Shifts		.003		.023*
		(.006)		(.014)
Saturday Shifts		.010		.018
		(.007)		(.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 total number of shifts, numbers 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 shift-by-pod indicators, 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 and referral party). Standard errors two-way clustered at the doctor and shift-by-pod level. The number of observations is 286,330.

TABLE A5
BASELINE CONTROLLING FOR RECENT ROTATION PATTERNS

	ED Cost		Length of Stay	
	(1)	(2)	(3)	(4)
Contractor	.090***	.090***	.316***	.290***
	(.018)	(.018)	(.032)	(.032)
Total Shifts		-.005		-.013
		(.005)		(.011)
Morning Shifts		-.001		.006
		(.003)		(.005)
Afternoon Shifts		.002		.011*
		(.003)		(.006)
M1 Shifts		.003		-.006
		(.004)		(.008)
M2 Shifts		-.001		-.011
		(.004)		(.007)
M3 Shifts		.001		-.008
		(.004)		(.007)
M4 Shifts		-.001		-.007
		(.005)		(.008)
Medica Shifts		.005		.009
		(.003)		(.006)
Monday Shifts		.001		-.002
		(.004)		(.010)
Tuesday Shifts		.007		.016
		(.005)		(.011)
Wednesday Shifts		.002		.001
		(.005)		(.011)
Thursday Shifts		.002		.002
		(.005)		(.011)
Friday Shifts		.001		.008
		(.005)		(.011)
Saturday Shifts		.008		.013
		(.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 total number of shifts, numbers 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 shift-by-pod indicators, 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 and referral party). Doctor fixed effects are absorbed. Standard errors two-way clustered at the doctor and shift-by-pod level. The number of observations is 286,330.

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	.129	.2712	.026	.0009	.0123	.0161	.0248	.0275
S.D. Bayes Shrunk Fixed Effects	.1192	.2528	.0133	.0003	.0031	.0044	.0067	.006
Mean Variable	Logs	Logs	.1854	.0008	.0364	.065	.1328	.1741

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 as opposed to discharged home. Death in ED is a dummy variable for whether the patient dies in the ED. Return to ED 2 is a dummy variable for whether the patient returns to the ED 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 and referral party). The number of observations is 305,205.

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

Cost Item	Percentage	Item Type
Palliative care consultation (first visit)	7.38	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	4.56	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	3.08	Other
Chest X-ray	2.61	Other
Amylase	2.6	Laboratory
Electrocardiogram (ECG/EKG)	2.26	Other
Troponin T	1.69	Laboratory
CT scan of the brain (without contrast)	1.05	Other
Microbiological specimen collection	.85	Laboratory
Abdominal X-ray (without contrast)	.77	Other
Neurology consultation (first visit)	.53	Other
Viral nucleic acid detection in biological material by hybridization following PCR	.52	Laboratory
Blood culture (hemoculture)	.5	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)	17.07	Other
CT scan of the brain (without contrast)	10.69	Other
Chest X-ray	4.67	Other
CT scan of the abdomen, complete, with and without contrast	3.38	Other
Viral nucleic acid detection in biological material by hybridization following PCR	3.28	Laboratory
Troponin T	3.23	Laboratory
Injection or infusion of specific drugs	3.01	Other
Electrocardiogram (ECG/EKG)	2.7	Other
C-reactive protein: quantitative assay	2.69	Laboratory
Complete abdominal ultrasound	2.63	Other
CT scan of the chest with and without contrast	2.22	Other
CT scan of the brain with and without contrast	2.18	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	1.6	Other
CT scan of the lumbosacral spine (3 vertebrae and 2 interspaces) without contrast	1.48	Other
CT scan of the chest (without contrast)	1.45	Other
Blood culture (hemoculture)	1.42	Laboratory
Venous blood draw	1.36	Laboratory
Upper abdominal ultrasound	1.31	Other
Neurology consultation (first visit)	1.23	Other
CT scan of the facial bones (without contrast)	1.23	Other
Echocardiography	1.22	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	.89	Laboratory
Creatinine	.89	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	7.34	Laboratory
Urinary potassium	6.14	Laboratory
Blood culture	5.98	Laboratory
Creatine kinase (CPK or CK)	5.7	Laboratory
Complete blood count with differential	5.68	Laboratory
Transaminase GPT (ALT)	5.58	Laboratory
Creatinine	4.61	Laboratory
Urinary sodium	4.61	Laboratory
Urea (blood urea nitrogen)	4.51	Laboratory
Pre-prandial glucose	4.45	Laboratory
C-reactive protein (CRP): quantitative measurement	4.02	Laboratory
Total bilirubin reflex, incl. possible fractionated bilirubin determination	3.81	Laboratory
Total calcium	3.66	Laboratory
Venous blood draw	2.98	Laboratory
Prothrombin time (PT)	2.83	Laboratory
Partial thromboplastin time (PTT)	2.81	Laboratory
Electrocardiogram (ECG/EKG)	2.4	Other
Palliative care consultation (first visit)	2.11	Other
Amylase	2.06	Laboratory
Drugs of abuse: heroin metabolites in urine	1.59	Laboratory
Chest X-ray	1.44	Other
Left hip X-ray	1.24	Other
Neurology consultation (first visit)	1.16	Other
CT scan of the brain without contrast	1.01	Other
Left wrist X-ray	.96	Other
Orthopedic-traumatology consultation (first visit)	.72	Other
Right arm X-ray	.63	Other
Right tibiotarsal X-ray	.63	Other
Left clavicle X-ray	.56	Other
Left leg X-ray	.56	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	.441***			.298***	
	(.007)			(.009)	
Total Number of Doctors		.277***			.223***
		(.004)			(.004)
Length of Stay			.133***	.080***	.071***
			(.004)	(.004)	(.003)
Panel B: Dep. Var. = Share of Duplicated Cost Items					
Handover to Next Shift	.050***			.031***	
	(.001)			(.001)	
Total Number of Doctors		.033***			.027***
		(.001)			(.001)
Length of Stay			.016***	.010***	.009***
			(.000)	(.000)	(.000)

Standard errors two-way clustered at the doctor and shift-by-pod level. The average number of duplicated cost items is .71. 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 and referral party). The number of observations is 305,205.

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	-.017
	(.007)	(.017)
Panel B: Adding Patient Characteristics		
Number of Contractors in Pod	-.001	-.019
	(.007)	(.016)
Panel C: Adding Doctor Fixed Effects		
Number of Contractors in Pod	-.003	-.020
	(.006)	(.014)

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 and referral party. Standard errors two-way clustered at the doctor and shift-by-pod interaction level. The number of observations is 261,225.

TABLE A12
EFFECTS ON OVERSTAY VARIABLE
ONLY OBSERVATIONS THAT NOT
DISCHARGED BY END OF SHIFT

	(1)	(2)	(3)	(4)
Employee		.0655***	.0475***	.0423***
		(.0094)	(.0107)	(.0105)
Difference in Gender	-.0061			-.0060
	(.0059)			(.0058)
Difference in Age	-.0015***			-.0011**
	(.0005)			(.0005)
Doctor Fixed Effects	Yes	No	Yes	Yes

Standard errors two-way clustered at the doctor and shift-by-pod level. Overstay takes value one when a patient is discharged by the doctor originally assigned to the case 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 and referral party). 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 119,614.

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

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Employee × Incoming Employees	.015**	.015***	.015***	.015***	.017***	.016***	.019***	.017***
	(.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-by-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 and referral party), and doctor fixed effects. The mean of the dependent variable Overstay is .037. The number of observations is 304,447.

TABLE A14
EFFECTS ON OVERSTAY VARIABLE - ROBUSTNESS

	(1)	(2)	(3)
Panel A: Unweighted Average Among Next-Shift Doctors			
Employee	.009	.005	.005
	(.006)	(.007)	(.006)
Employee × Incoming Employees	.011	.011	.008
	(.007)	(.007)	(.007)
Panel B: Using the Next-Shift Doctor with Most Cases			
Employee	.010**	.006	.006
	(.005)	(.006)	(.006)
Employee × Incoming Employees	.010**	.010**	.008*
	(.004)	(.005)	(.004)
Panel C: Overstay Calculated Over 2-Hour Window			
Employee	.004	-.001	.000
	(.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	-.002	-.002
	(.006)	(.007)	(.007)
Employee × Incoming Employees	.014**	.016**	.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	-.323***	-.316***	-.201***	-.249***	-.251***
	(.045)	(.032)	(.064)	(.048)	(.048)
Employee × Incoming Employees			-.138***	-.077	-.063
			(.050)	(.047)	(.047)
Difference in Gender					.023*
					(.013)
Difference in Age / 10					-.014
					(.015)
P-value 1+2			.000	.000	.000
Doctor Fixed Effects	No	Yes	No	Yes	Yes
Panel A: Dependent Variable = ED Cost					
Employee	-.105***	-.090***	-.059**	-.067***	-.068***
	(.018)	(.018)	(.026)	(.027)	(.028)
Employee × Incoming Employee			-.052**	-.026	-.023
			(.023)	(.021)	(.022)
Difference in Gender					.006
					(.006)
Difference in Age / 10					-.006
					(.008)
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-by-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 and referral party). The number of observations is 302,166.

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 and referral party). 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-by-pod interaction level. The number of observations is 302,166.

TABLE A17
OVERLAP VARIABLES AND OVERSTAY

	(1)	(2)
Horizontal Overlap (Past)	.030	.022
	(.020)	(.018)
Backward Handover (Past)	-.090***	-.058***
	(.026)	(.024)
Forward Handover (Past)	-.032	-.012
	(.033)	(.031)
Horizontal Overlap (Future)	.030	.027*
	(.021)	(.016)
Backward Handover (Future)	-.083*	-.030
	(.043)	(.034)
Forward Handover (Future)	.028	.049
	(.036)	(.033)
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 and referral party). 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-by-pod interaction level. The number of observations is 286,330.

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Employee		.021***	.015***	.014***	.006	.001	.002
		(.003)	(.004)	(.004)	(.006)	(.006)	(.006)
Employee × Incoming Employees					.017***	.016***	.014**
					(.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

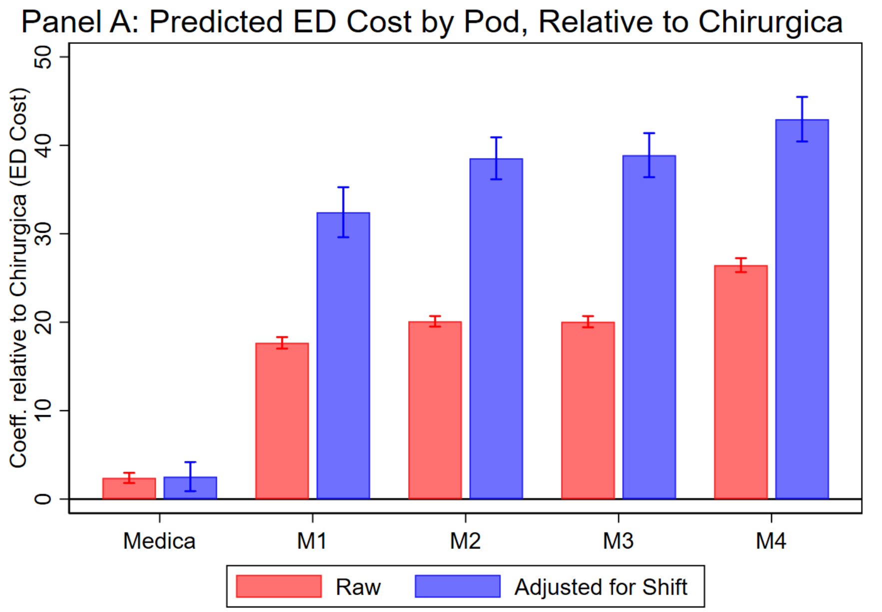
Standard errors two-way clustered at the doctor and shift-by-pod level. Overstay takes value one when a patient is discharged by the doctor originally assigned to the case, but in the sixty minutes after that doctor's shift has ended. Difference in gender is the absolute value of the difference between 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 and referral party). 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 286,330.

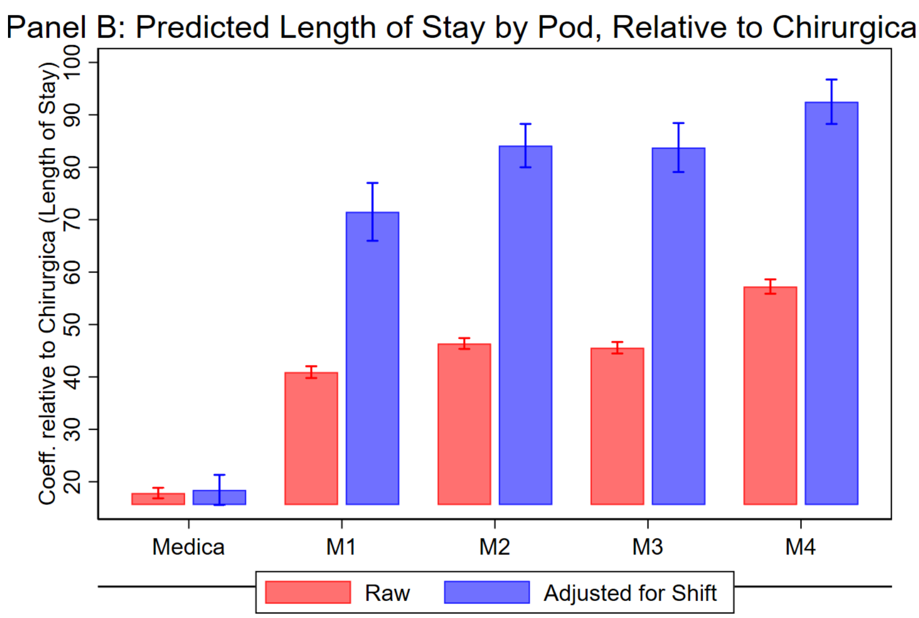
TABLE A19
POST-ADJUSTMENT TREND TEST

	Length of Stay		Cost of Treatments	
	$\tau = 2$	$\tau = 3$	$\tau = 2$	$\tau = 3$
Post	-0.3315***	-0.3394***	-0.0991***	-0.0988***
	(0.0324)	(0.0332)	(0.0216)	(0.0210)
γ (post-τ linear trend)	-0.0099	0.0036	0.0086	0.0151
	(0.0167)	(0.0212)	(0.0131)	(0.0154)

This table reports estimates from a parametric version of the event study that allows the post-transition effect to evolve linearly after an adjustment horizon $\tau \in \{2, 3\}$. *Post* denotes the common post-transition effect over the first τ years after transition, while γ is the coefficient on the post-τ linear trend, defined as treated $\times$ (years since transition $-\tau$) and set to zero before year τ. All specifications include doctor, shift $\times$ pod, and exact-hour fixed effects, together with patient characteristics indicators. Standard errors are two-way clustered by doctor and shift-date. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The number of observations is 266,529

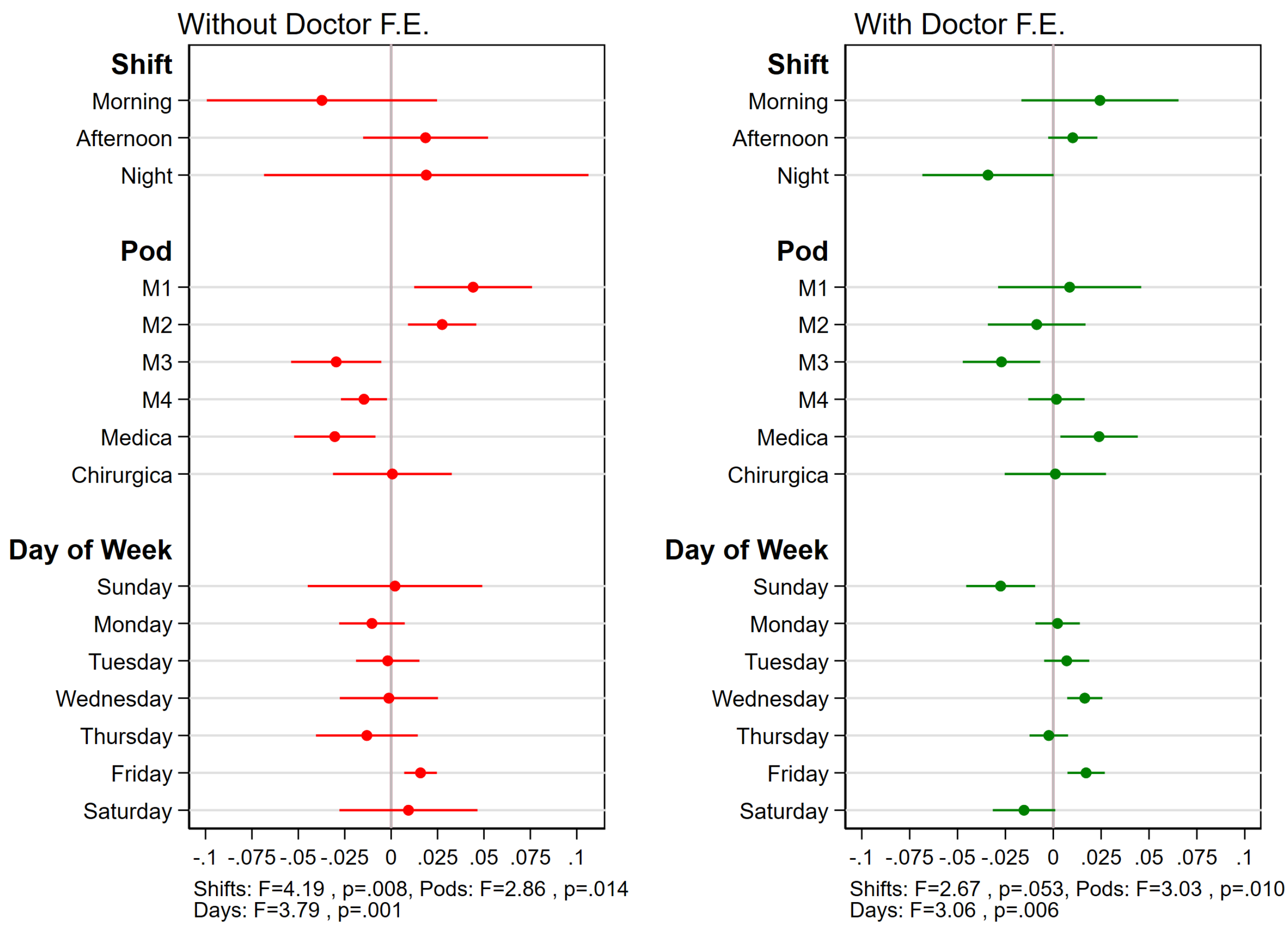
FIGURE A1: AVERAGE PREDICTED COST AND PREDICTED LENGTH OF STAY, BY POD





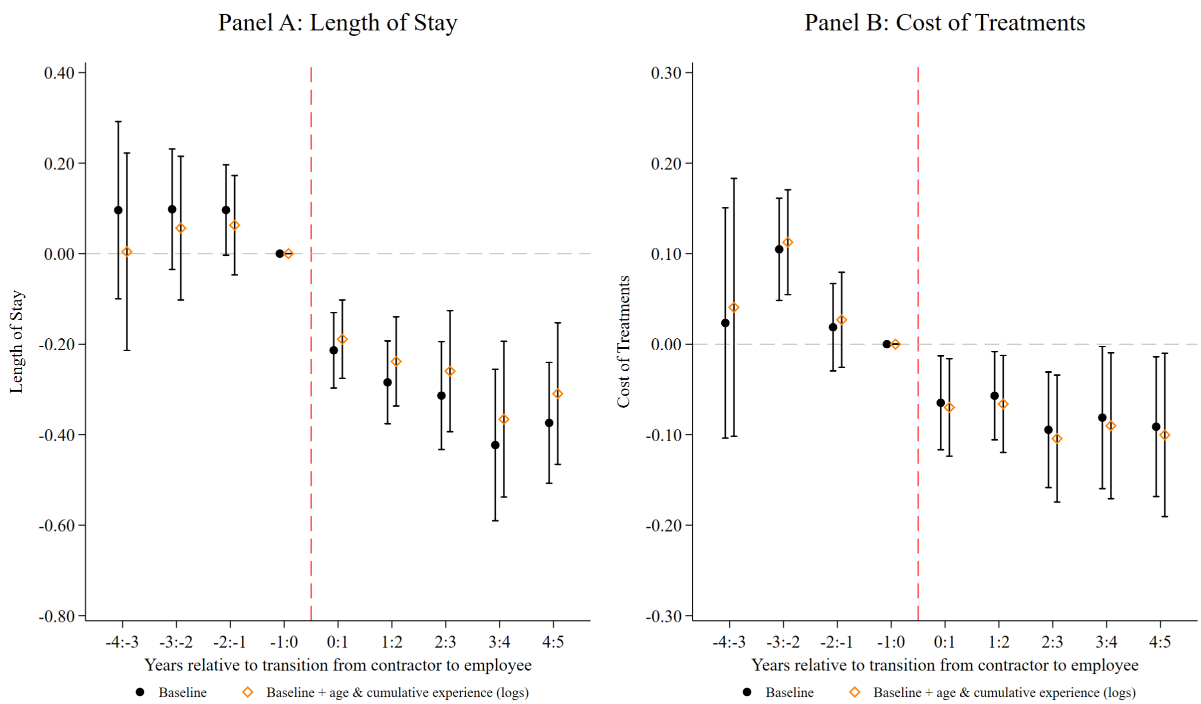
This figure displays regression coefficients of predicted cost and predicted length of stay on the pod to which the incident is assigned. The predictions are done based on patient characteristics (age, gender, triage, main complaint, mode of arrival and referral party). Cost and length of stay are in logs. The regressions further 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-by-pod interaction level. The sample is the baseline sample. The number of observations is 305,221.

FIGURE A2: ASSIGNMENT OF SHIFTS AND PODS BY CONTRACTOR STATUS



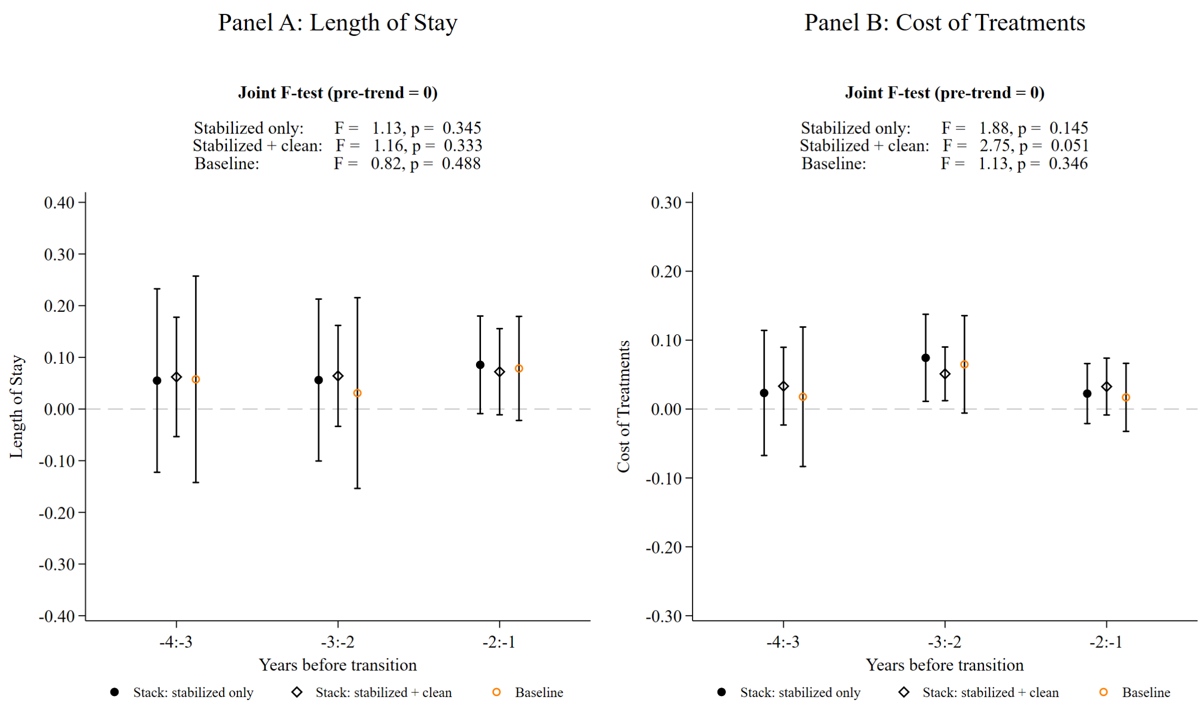
This figure displays regression coefficients of shift and pod characteristics on contractor status. The sample is a sample of shift-by-pod and doctors. Each row displays the coefficient of a separate regression with the row characteristic on the left hand side and contractor status on the right hand 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 levels.

FIGURE A3: EVENT-STUDY, CONTROLLING FOR AGE AND EXPERIENCE



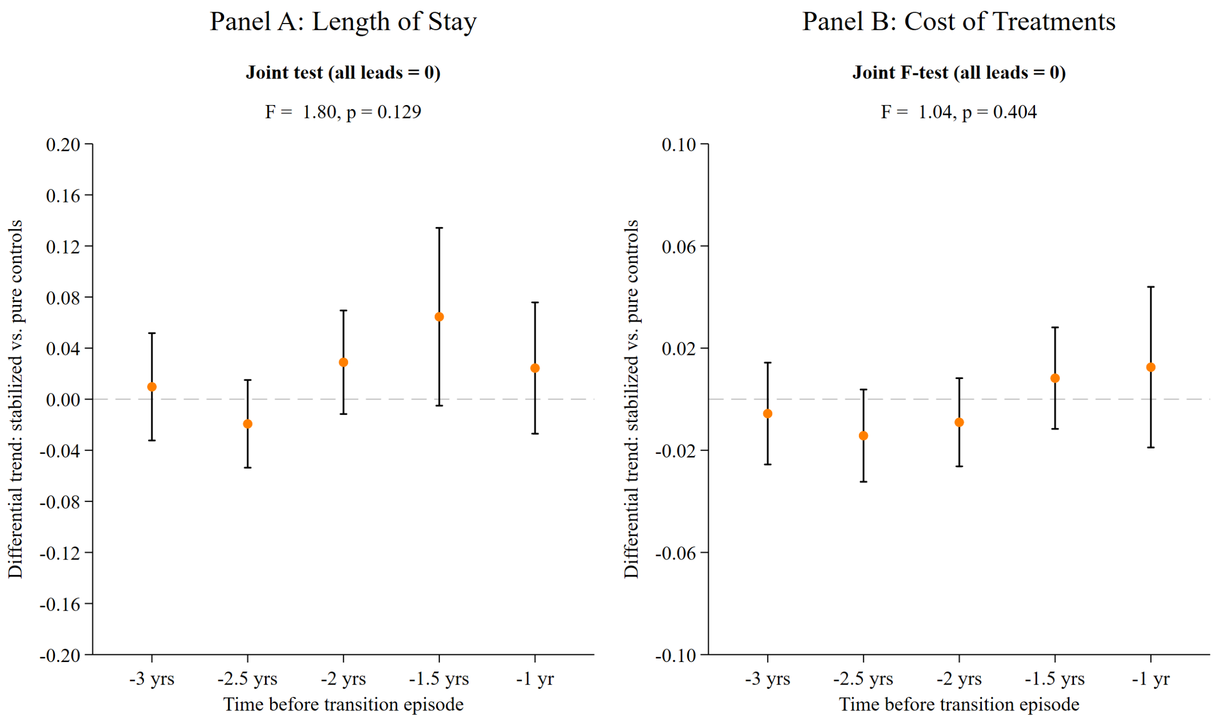
This figure compares the baseline event-study estimates to those obtained after augmenting the baseline specification with controls for the log of doctor age and log cumulative cases (as a proxy for experience). Fixed effects, clustering, and sample restrictions are as in Figure 5.

FIGURE A4: TESTS OF PRE-TRENDS USING ONLY THE PRE-TRANSITION SAMPLE



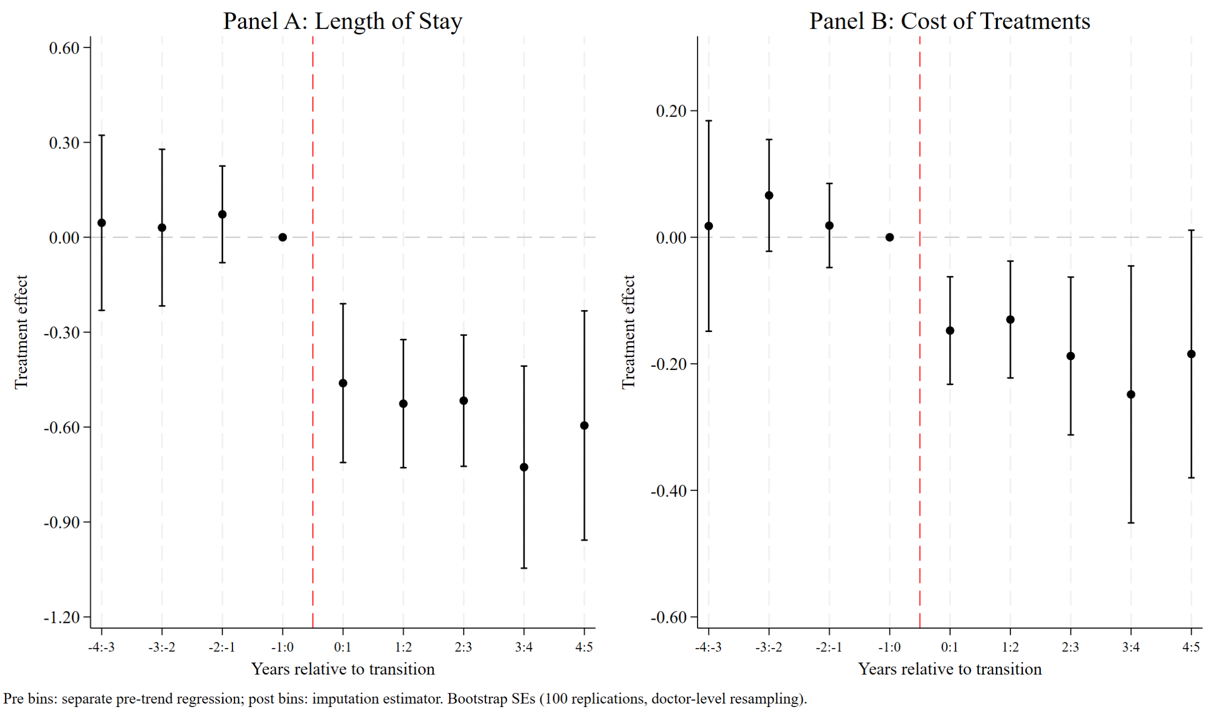
This figure reports pre-period lead coefficients from re-estimating each of the three specifications in Figure 5 on the pre-transition sample only, excluding all post-transition observations of switchers (Borusyak et al., 2024). For each specification, the reported joint F-test evaluates the null that all pre-period lead coefficients equal zero. Fixed effects, clustering, and sample restrictions are as in Figure 5.

FIGURE A5: COMPARING TRENDS BETWEEN STABILISED EMPLOYEES VS. CLEAN CONTROLS



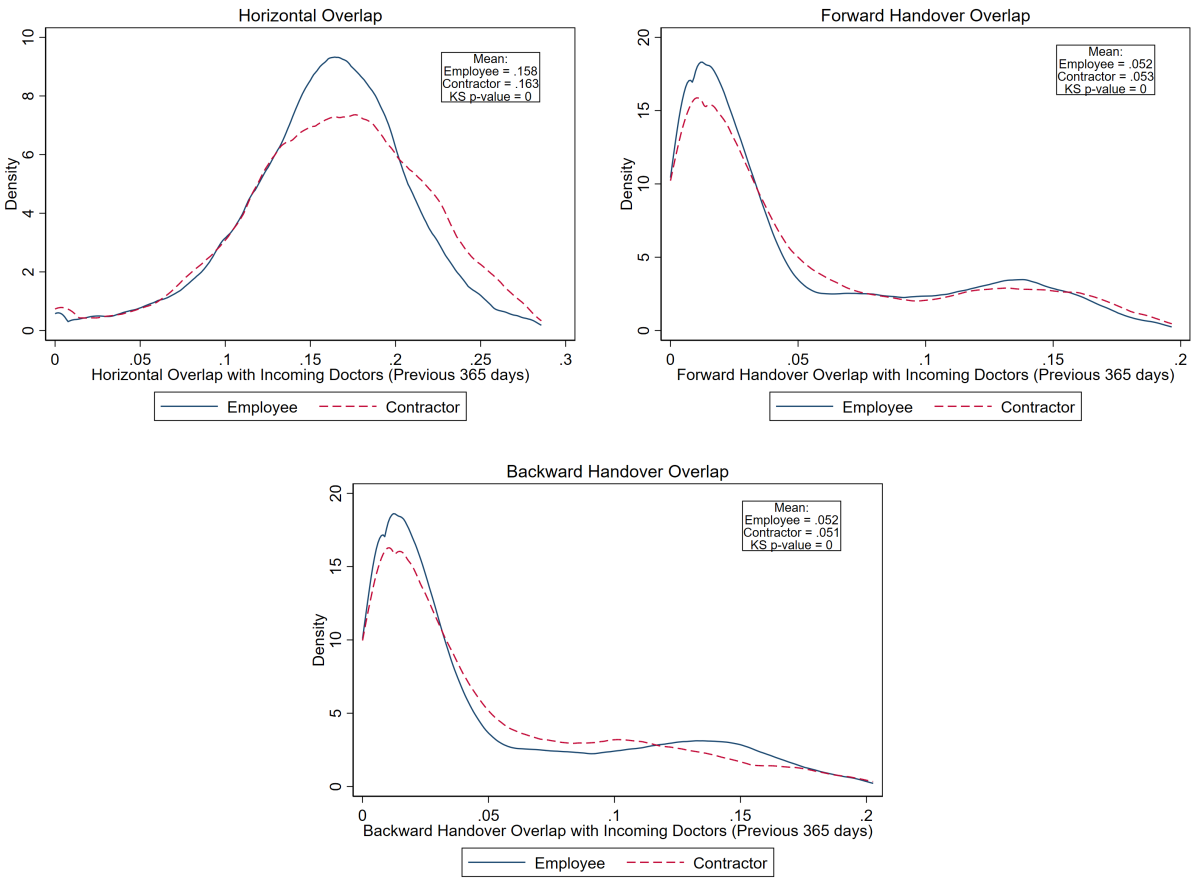
This figure displays the interaction coefficients δ_s from estimating equation (A1) on the stacked sample of pre-transition windows. For each of the fourteen transition episodes we retain the three years prior to the episode transition date. S_{je} indexes semester bins before the transition date (the semester immediately before transition is the omitted category), and $\text{Stabilised}_{i(j)} = 1$ for doctors who are always observed as employees in our sample or for switchers observed at least five years after their own transition. The plotted coefficients δ_s capture differential pre-period movements of stabilised employees relative to the clean control group (never-employees and not-yet-employees, defined using each doctor's own transition date) within episode windows. Fixed effects, clustering, and sample restrictions are as in Figure 5. The joint F-test reported in each panel tests $H_0 : \delta_s = 0$ for all displayed leads.

FIGURE A6: EVENT STUDY COEFFICIENTS UNDER THE IMPUTATION ESTIMATOR OF BORUSYAK ET AL. (2024)



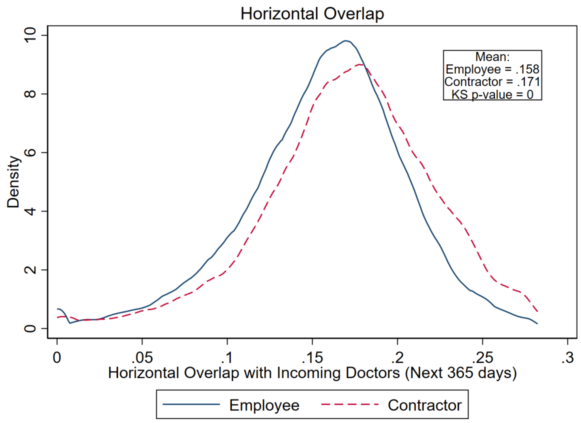
This figure reports dynamic treatment effects from a modified imputation estimator (see Borusyak et al. 2024 for details). The untreated-outcome model in equation (A2) is estimated on the donor sample of always-employees and pre-transition observations of switchers ($D_{it} = 0$), using the same fixed-effect structure and patient controls as the baseline specification. The fitted model is used to impute untreated potential outcomes for switchers after transition; post-transition event-study coefficients are the averages of the resulting imputed effects within event-time bins, normalized to the omitted $[-1, 0]$ bin. Pre-trend coefficients come from a separate regression run on pre-transition observations only. Standard errors are computed via a doctor-level block bootstrap with 400 replications.

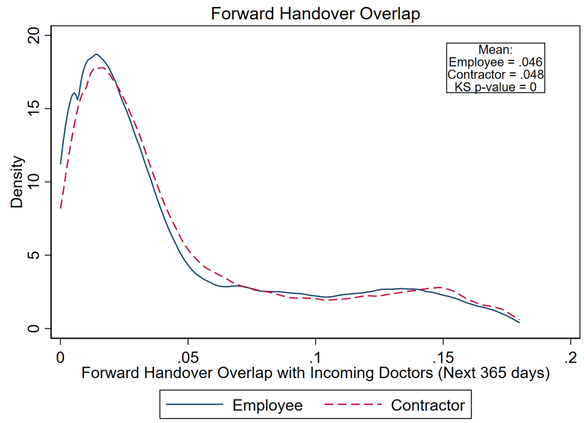
FIGURE A7: DISTRIBUTION OF OVERLAP BETWEEN FOCAL DOCTOR AND INCOMING DOCTOR(S) (PREVIOUS 365 DAYS)

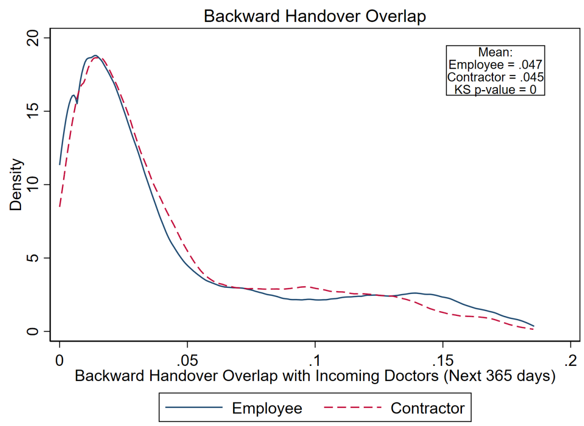


These figures display measures of overlap between the focal doctor and the doctors that are incoming in the same pod and next shift. The horizontal overlap measure is the number of shifts in which the focal doctor and the incoming doctors have coincided over the previous year, divided by the product 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 A8: 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: Details and Robustness of the Event Study Regressions

B.1. Construction of the Stacked Specifications with Stabilised Controls This subsection includes further details on the construction of the stacked DiD specifications. For each transition we construct an episode centred on the transition date, retain the focal doctor's event-time observations, and pool them with control observations drawn from the same calendar window (four years before to five years after the transition[1]). Episodes are then stacked, the regression absorbs episode fixed effects, and observations that appear in multiple episodes are weighted by the inverse of the number of episodes in which they appear (e.g. for details, Cengiz et al. 2019, Sun and Abraham 2021). In the first stacked specification the control pool is restricted to stabilised employees only. In the second we also include the small set of clean controls observed in each episode window. Figure 5 shows that both stacked specifications trace out a dynamic profile essentially identical to the baseline event study, confirming that the baseline estimates are predominantly identified from the stabilised group.[2]

A separate concern is that the timing of the transition may correlate with age or cumulative experience, so that the estimated effects could capture life-cycle or human capital accumulation rather than the effect of the transition. Figure A3 addresses this directly: augmenting the baseline specification with controls for log age and log cumulative cases leaves the estimated dynamic path essentially unchanged for both outcomes.

B.2. Testing Pre-Trends A visual inspection of Figure 5 reveals that some lead coefficients are statistically different from zero, particularly for cost of treatments. However, as discussed in Borusyak et al. (2024), the leads in a joint dynamic specification are not a clean diagnostic of pre-trends when estimated jointly with the post-transition coefficients, because heterogeneous treatment effects across units or event-time horizons contaminate the fitted lead coefficients. Borusyak et al. (2024) instead recommend assessing pre-trends using a regression estimated on the pre-period only, which isolates pre-transition variation and yields a cleaner test of the parallel trends assumption.

Figure A4 implements this diagnostic for each of the three specifications in Figure 5. Once post-transition observations are excluded from the sample, the lead coefficients become visibly smaller in all three cases. The joint F-test of the null that all pre-period leads are equal to zero does not reject in any specification (although one stacked specification yields a marginally non-significant result for the treatment cost outcome). The pre-period evidence is therefore broadly consistent with the parallel trends assumption, and the non-zero leads coefficients in Figure 5 appear to reflect the contamination from normalisation

[1] The pre-transition window is truncated at four years. The number of observations lying more than four years before the transition date is tiny (i.e., four observations).

[2] The five-year cutoff is justified if treatment effects stabilise within that horizon. Figure 5 suggests this: the coefficients are approximately flat in the last three post-transition bins, and a joint F-test of equality across those three bins does not reject for either outcome or specification. As this test may miss gradual drift, we also estimate a parametric specification that allows the treatment effect to evolve linearly beyond an adjustment horizon τ. For $\tau \in \{2, 3\}$ and both outcomes, the post-τ slope is small and statistically indistinguishable from zero (Table A19).

and composition effects inherent in the joint estimation of leads and lags.

B.3. Direct Comparison of Stabilised Employees and Clean Controls The most direct evidence on the validity of the control group in the stacked regressions comes from comparing stabilised employees to clean controls in the pre-transition windows where both are observed. While comparing transitioning doctors directly to never-treated and not-yet-treated observations is infeasible given the granularity of the fixed effects required for identification, the overlap between stabilised employees and clean controls is considerably wider, because stabilised employees appear in a much larger number of shift-by-pod cells. This makes it possible to test whether the two groups move in parallel in the pre-period, which in turn indicates whether stabilised employees represent a valid counterfactual path in the cells where clean controls are too sparse on their own. For each of the fourteen transition episodes we retain the three years preceding the transition date, stack these windows, and estimate:

$$y_{je} = \sum_{s=2}^{6} \gamma_s \, \mathbf{1}\{S_{je} = s\} + \sum_{s=2}^{6} \delta_s \big(\mathbf{1}\{S_{je} = s\} \times \text{Stabilised}_{i(j)}\big) + \eta_{i(j)} + \psi_e + \kappa_{sp(j)} + \phi_{h(j)} + \mathbf{X}_j' \lambda + \varepsilon_{je}, \tag{A1}$$

where S_{je} indexes semesters before the episode transition (the semester immediately before the transition is the omitted category) and $\text{Stabilised}_{i(j)} = 1$ for stabilised employees and 0 for clean controls. $\eta_{i(j)}$, ψ_e, $\kappa_{sp(j)}$, and $\phi_{h(j)}$ are doctor, episode, shift-by-pod, and exact-hour fixed effects, respectively, and $\mathbf{X}_j$ includes the baseline patient-level controls. The parallel trends assumption across the two groups can be tested by examining whether $\delta_s = 0$ for all s. Figure A5 shows that the δ_s estimates are small in all pre-transition periods for both outcomes, and the F-test does not reject the null that they are jointly zero. To the extent that clean controls identify the correct counterfactual time path where available, Figure A5 supports using stabilised employees to identify that same path in the shift-by-pod cells where clean controls are too sparse to contribute to identification.

B.4. Modified Imputation Estimator Figure A6 reports a complementary robustness exercise based on a modified version of the imputation estimator proposed by Borusyak et al. (2024). Because clean controls are too sparse to serve as the donor pool within our shift-by-pod cells, we instead use as donors the set of observations with $D_{it} = 0$, where $D_{it} = \mathbf{1}\{i \text{ is a switcher and } t > T_i\}$ and T_i denotes the transition date of doctor i. The donor sample therefore consists of always-employee observations and pre-transition observations of switchers. We estimate the untreated-outcome model:

$$y_{it} = \eta_{i(j)} + \kappa_{sp(j)} + \phi_{h(j)} + \mathbf{X}_{it}' \lambda + u_{it} \qquad \text{on the donor sample } D_{it} = 0, \tag{A2}$$

using the same fixed-effect structure as the baseline. The fitted model is then used to impute untreated potential outcomes $\widehat{y}_{it}(0)$ for transitioning doctors, and the post-transition event-study coefficients are computed as averages of the imputed treatment effects $\widehat{\tau}_{it} = y_{it} - \widehat{y}_{it}(0)$ within event-time bins, normalised relative to the omitted $[-1, 0]$ bin. Pre-trend coefficients are obtained from a separate regression on pre-transition observations only, and inference is based on a doctor-level block bootstrap with 400 replications.

The resulting dynamic profile is qualitatively consistent with the baseline and stacked specifications: pre-period coefficients are small and jointly insignificant, and post-transition coefficients are negative, emerge immediately, and persist. The imputation estimates are somewhat larger than in the stacked event study, and two factors plausibly account for this gap. First, the lead coefficients in the baseline and stacked specifications tend to be positive relative to the omitted bin, which is consistent with some degree of anticipation in the final months before the contract change. If so, this would attenuate the post-transition coefficients under the standard normalisation; the imputation estimator is less sensitive to this because it anchors the untreated counterfactual path on the full pre-transition history. Second, the two estimators weigh observations differently and use different samples to estimate the fixed effects, so they need not coincide in finite samples even under the same identification assumptions. The substantive conclusion is however the same across all three specifications: resource use falls sharply and persistently after doctors transition from contractor to employee status.

Appendix References

Borusyak, K., Jaravel, X., and Spiess, J. (2024), "Revisiting Event-Study Designs: Robust and Efficient Estimation," *Review of Economic Studies*, 91(6): 3253–3285.

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

Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019), "The Effect of Minimum Wages on Low-Wage Jobs," *Quarterly Journal of Economics*, 134(3): 1405–1454.

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