Organizational Responses to Labor Turnover: Evidence from Hospitals*

Danyelle Branco Bladimir Carrillo Luiz Fontes Leticia Nunes

February 13, 2025

Abstract

Labor turnover is a common and potentially disruptive event for firms. When a key worker leaves, managers must decide not only about hiring a suitable replacement but also assess whether and how to reorganize the production process to minimize operational disruptions. Yet, empirical evidence on these responses and their role in shaping output dynamics remains limited. In this paper, we investigate how the performance and internal organization of hospitals respond to physician turnover after death. The hospital industry is well suited for studying these questions given that inputs, outputs, and task allocation can be accurately measured. Using rich administrative data from Brazil, we document that hospitals respond to the sudden exit of a physician by strategically reallocating affected tasks and roles toward physicians already familiar with the local hospital environment, broadening their clinical scope without worsening the skill-to-task matches. We demonstrate that this strategy effectively mitigates disruptions to care quality but a reduction in the scale of operations appears to be necessary to achieve this balance. This suggests that hospitals prioritize quality over quantity when faced with disruptive events.

JEL codes: *M11, M51, M54, L23, D90*

Keywords: *labor turnover, organization of production, health care, active personnel management*

*Contact information: Branco: Sao Paulo School of Economics-FGV, Sao Paulo, SP, Brazil, 01332-000 (e-mail: danyelle.branco@fgv.br). Carrillo: Sao Paulo School of Economics-FGV, Sao Paulo, SP, Brazil, 01332-000 (e-mail: bladimir.carrillo@fgv.br). Fontes: Nunes: Insper - Institute of Education and Research, Sao Paulo, SP, Brazil, 04546-042 (e-mail: luiz.fontes@insper.edu.br). Nunes: Insper - Institute of Education and Research, Sao Paulo, SP, Brazil, 04546-042 (e-mail: leticiafcn@insper.edu.br). Financial support for this research from the Brazilian National Council for Scientific and Technological Development (CNPq —# 04/2021) is gratefully acknowledged. We thank Emanuel Ornelas, Fabio Miessi, Lucas Finamor, and Vladimir Ponczek, as well as participants various seminar participants for helpful conversations and comments on this project. We are solely responsible for this paper's contents. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

1 Introduction

Every month, many employees leave their jobs voluntarily. The latest US turnover survey, for example, reveals that approximately 5.1 million Americans left their jobs, and three of every five did so voluntarily.[1] Firms are thus frequently exposed to the loss of valuable labor. When a worker in charge of key tasks leaves, managers must make decisions beyond hiring a suitable replacement. They may need to adjust routines, reallocate tasks, update job roles, or provide additional support to remaining staff. Since new workers could take time to become proficient both in the technical and cultural aspects of the firm, managers' ability to adopt productive organizational changes may be crucial for mitigating any operational disruptions during the transition (Doeringer and Piore, 1971; Holmstrom and Tirole, 1989).

How firms manage these transitions could have far-reaching implications. A well-established body of work has documented substantial and persistent differences in productivity across producers, leading to significant economic disparities between countries and over time (Syverson, 2011; Bartelsman et al., 2013). There is also extensive evidence suggesting that much of these differences are explained by differences in management across firms (Ichniowski and Shaw, 1997; Gibbons and Henderson, 2012; Bloom et al., 2013, 2019). Understanding how the internal organization of firms responds to highly frequent, disruptive events, such as labor turnover, may shed light on when and why different firms succeed and thus help explain why management practices account for a sizeable portion of the much-debated variation in productivity. Yet, empirical evidence on the response of firms to labor turnover remains limited. As such, many questions remain unanswered: When a worker leaves, are all of their tasks are allocated to new entrants? How does this allocation shape output dynamics? Do firms rather rely on incumbents by broadening the scope of their tasks? Does this expansion affect the worker-to-task match quality? If necessary, do firms prioritize quality or quantity of output?

Tackling these questions confronts some empirical challenges. One of the biggest challenges is that we rarely have data with detailed information on tasks, workers, and output for a number of different firms. Standard employer-employee datasets are generally limited to payroll and basic demographic information about workers. When richer data are available, they are often restricted to a single organization, making cross-firm comparisons and empirical analyses of turnover unfeasible. Another major challenge is establishing causality. Factors that affect a firm's behavior and performance are likely to simultaneously affect turnover. Ideally, we would like to examine the effects of employees leaving a firm for reasons unrelated to the firm's potential outcomes.

In this paper, we seek to make progress on these challenges and offer a credible answer to the aforementioned questions. We do so in the context of the hospital industry, where inputs, outputs, and the allocation of tasks can be accurately measured. To overcome the identification challenge, we focus on the exit of physicians because of death. Using rich administrative data from Brazil,

[1] These statistics are published monthly by the US Bureau of Labor Statistics (see https://www.bls.gov/news.release/pdf/jolts.pdf, last accessed on January 24, 2025). Labor turnover is also a highly prevalent phenomenon in other countries. For example, Beccaria and Maurizio (2018) analyze annual turnover statistics for Argentina, Brazil, Ecuador, Mexico, Paraguay, and Peru, and report prevalence rates between 24 and 44 percent.

we document that hospitals respond to the sudden exit of a physician by strategically reallocating affected tasks and roles toward physicians already familiar with the local hospital environment, broadening their clinical scope without worsening the skill-to-task matches. We demonstrate that this strategy effectively mitigates disruptions to care quality but a reduction in the scale of operations appears to be necessary to achieve this balance, suggesting that hospitals prioritize quality over quantity.

The healthcare sector is of substantive interest because of its size, heavy fiscal burden, and prominent role in patient welfare. We bring a rich compilation of administrative registers that allow us to track the universe of physicians over time, including their background characteristics, assigned roles, detailed inpatient services, and patient outcomes. Our main analysis draws on approximately 189 million physician-patient interactions and exploits turnover after death as a source of exogenous variation in a hospital's labor supply. To limit the scope for anticipation effects, we restrict the death events to physicians who were actively working at the time of their death. Because physicians often differ in their clinical focus within hospitals, much like distinct production lines in a factory, a physician's exit is likely to be felt differently across different areas of care.[2] We exploit this variation and trace the evolution of outcomes across hospitals in a given area that relied to varying degrees on the deceased physician, as measured by their patient shares, in a difference-in-differences setup. To avoid biases from "forbidden" comparisons that could arise due to heterogeneity in treatment timing, we use a matching procedure to identify appropriate comparison hospitals and implement a stacked difference-in-differences design (Wing et al., 2024; Butters et al., 2022; Cengiz et al., 2019).

We begin by showing that the death of a physician constitutes a major labor supply shock to hospitals that were more dependent on their expertise. We estimate a sharp decline in physician staffing levels immediately after the death event, followed by a quick recovery, with staffing returning to pre-event levels by the fifth quarter. Hospitals adjust primarily by hiring new physicians, but we also find some evidence of increased retention. Importantly, there are no pre-existing differential trends correlated with turnover after death, consistent with the unanticipated nature of the event and the validity of our statistical approach.

Having documented the quantitative importance of the event, we then investigate how hospitals organize internally in response to the shock and its consequences for output. The rest of the paper proceeds in three steps. In the first part of the paper, we explore changes in the allocation of tasks. We find positive effects on the patient load allocated to incumbents as well as on the number of distinct health conditions they treat, with some mean-reverting dynamics over time. We also find a clear and persistent shift in the allocation of nonclinical duties toward incumbents. Deceased physicians are typically older and more involved in non-clinical duties, so their exit requires reassigning these responsibilities and incumbents can be viewed as the most natural replacements. Notably, there are no changes in incumbents' total hours worked, suggesting that they accommodate the ad-

[2] As described in Section 4, we define areas of care or disease areas using information on the specific cause of admission of the patients, categorized according to the relevant 3-digit ICD codes. The classification is relatively narrow. For instance, within the ischemic heart disease category, there are separate codes for myocardial infarction, angina pectoris, and chronic ischemic heart disease.

ditional clinical and non-clinical workload through increased productivity. Over time, however, this dual adjustment appears to be unsustainable, given that the increase in the clinical load eventually fades.

We then examine how different areas of care are affected. Because the adjustment in staffing levels is not instantaneous, hospitals must find ways to ensure that all areas continue to operate in the immediate aftermath of the event. They could do so by relying on incumbent physicians who already operated in those areas or by broadening the clinical responsibilities of physicians from other areas. We find no evidence of a decline in the number of physicians available for care in more exposed areas, not even in the very short term. This suggests that the expansion of incumbents' tasks is the key mechanism of adjustment. Remarkably, we do not find evidence that this task expansion is associated with an increase in the mismatch between a physician's area of specialization and the specialty clinically relevant for a patient's condition.[3] Thus, hospitals manage to expand the scope of tasks for physicians without significantly worsening the skill-to-task matches.

We also directly assess how much of the caseload left by deceased physicians is reallocated to new entrants. Our estimates indicate that newcomers are far from absorbing the entire workload. While we find that patients in more affected areas are more likely to be attended by a new physician immediately after the event, the magnitude of this effect is modest. If new physicians were to take on the entire service gap left by the deceased physician, fully exposed areas would experience a 100 percentage-point change in patient shares. However, we find point estimates falling well within the range of 14 to 20 percentage points. We show that this limited patient reallocation to new physicians is not due to delays in recruitment, or a lack of necessary qualifications among new entrants to perform the medical procedures previously handled by the deceased physicians. Instead, hospitals appear to rely primarily on incumbent physicians to manage the labor supply shock.

The organizational responses of hospitals are unlikely to involve physicians solely. It is possible that hospitals also adjust on dimensions that are harder to observe, including the reorganization of the support staff or operational protocols. We provide evidence suggesting that hospitals indeed make behind-the-scenes adjustments. Our results suggest that when faced with the death of a physician, hospitals manage to retain existing nursing staff, who most likely have already established working relationships with physicians and other staff members. Increased retention, in turn, is associated with a decline in the hiring of new nurses, potentially alleviating the financial and operational burdens of recruitment and training. These findings are important given the central role that nurses play in the functioning of hospitals.

In the second part of the paper, we examine the implications of these organizational responses for output. We examine the quantity and quality dimensions of performance. On the quantity side, we find a persistent decline in the number of inpatient admissions, which implies a reduction in the

[3] We generate this measure of mismatch for each patient-physician pair using data from the Ministry of Health, which determines the list of specialties considered clinically relevant for each procedure. As an example, consider the clinical treatment of acute stroke. The specialties listed as relevant include neurology, pediatrics, internal medicine, intensive care, and geriatrics. We say that there is a mismatch between the physician's expertise and this procedure if the physician's specialty is not one of those listed as relevant for the treatment of acute stroke. In the data, approximately 11 percent of patients are treated by physicians without the specialization clinically relevant for the condition.

utilization of bed capacity. Our estimates indicate that areas with full exposure to the death event accommodate 10 percent fewer patients in the post-event period. Importantly, the data show that providers target these cuts where they are likely to have the least detrimental impact, prioritizing patients with higher mortality risk. At the same time, we do not see significant changes in the length of stay and total number of procedures administered to a patient. Thus, while bed utilization declines following the exit of a key physician, the intensity of care provided to individual patients remains largely unchanged.

On the quality side, we find no evidence of deterioration, or at least none quantitatively important to be detectable in widely used metrics of care quality. Our main measure of care quality is inpatient mortality among patients at the highest end of the mortality risk distribution, similar to Silver (2021). The focus on "at-risk" patients is compelling not only because they are particularly sensitive to changes in the quality of care delivered, but also because these cases often involve acute conditions that require immediate care, leaving little room for selection.[4] Our results indicate that the estimated effect of physician turnover on patient mortality can be bounded to a tight interval around zero. We find similarly small and statistically insignificant estimates when we examine 30-day mortality, or when we use less extreme measures of care quality, such as readmission and use of high-technology, highly complex procedures believed to be beneficial for patient outcomes.

Our reading of the evidence is that despite their organizational efforts, hospitals experience a decline in productivity. Hospitals achieve the same patient outcomes but reduce their overall output in scale. There are two major classes of mechanisms that may explain these findings. The first channel is doctor "quality." If the expertise of the average physician is easy to replace and deceased physicians are not very different from the average one, then it could be less challenging for hospitals to maintain the quality of care delivered. Another possible mechanism is that hospitals prioritize care quality over quantity in their responses. Suppose that hospitals admit patients up to the point where the expected quality of care they can deliver per patient approaches a minimum care standard threshold. The decline in operations we observe might reflect, for example, that hospitals are not willing to reduce their minimum standard of care by assuming risks with physicians unfamiliar with the hospital's procedures in high-exposed, strained areas.

In the final part of the paper, we provide evidence that both the individual-specific capital and quality prioritization channels are important. We use the inpatient mortality among at-risk patients and generate mortality predictions for each physician as a proxy for their individual-specific capital. Our calculations suggest that there is strong heterogeneity among deceased physicians, but in the end, deceased physicians are only slightly more effective than their peers on average. This suggests that replacing the skills of the deceased physician among the remaining staff is unlikely to impose extreme challenges.

We exploit the rich heterogeneity in mortality predictions to gain further insights into the mechanism at play. In particular, we document that the effect of physician turnover on mortality varies heterogeneously with the predicted mortality change. Our estimates indicate that the exit of "good"

[4] In line with this reasoning, we show that there is no evidence of compositional changes in the patient case mix correlated with the exposure to the death event in this subsample.

physicians causes an increase in mortality, but the exact opposite pattern holds when a "bad" physician leaves. The magnitude of these effects is remarkably asymmetric: good physicians have a much smaller impact on mortality than bad physicians. We interpret these asymmetric effects as evidence consistent with the quality prioritization channel. When a high-performing physician departs, hospitals may adopt stronger and more effective organizational responses. In contrast, the exit of a low-quality physician does not trigger a similar response, as their absence alone already serves as a corrective mechanism.

We also show that the pattern of patient allocation is effective in mitigating disruptions to care quality. Zooming into the pre- and post-event change in mortality, we find that more exposed areas witness an increase in mortality when considering those patients treated by new physicians joining in the post-event period. This effect decays rapidly over time, suggesting that physicians develop some degree of hospital-specific capital. The magnitude of the effect is large. If new entrants had an overall patient share of 30 rather than 13 percent, fully exposed areas would have experienced a 10 percent increase in mortality. Once again, these findings are consistent with the quality prioritization channel. If the focus of hospitals were solely on maximizing volume, new entrants would be required to assume the exact workload of the deceased physician, irrespective of whether their adaptation is more challenging in highly exposed, strained areas. In practice, however, new entrants take on only a limited fraction of the workload and this is why we do not detect meaningful changes in overall mortality rates.

Our final set of results assess whether hospitals are sensitive to the potential extent of disruption caused by physician turnover. The quality prioritization channel predicts that providers adjust their internal production process in proportion to the expected disruption to care quality. We find that the patient share allocated to new entrants in more exposed areas is significantly larger when a bad physician leaves than when a good physician departs. The differences are sizeable, on the order of a factor of two. Given that new entrants are associated with worse patient outcomes in more exposed areas, this pattern suggests that hospitals strategically adjust task allocation depending on the perceived severity of the disruption.

Related Literature. The findings of this paper speak to a broad literature on management and corporate performance. Numerous studies in this area have documented that managerial and organizational factors significantly influence firm and worker productivity (e.g. Bertrand and Schoar, 2003; Bloom et al., 2013; Gosnell et al., 2020; Hoffman and Tadelis, 2021; Fenizia, 2022; Otero and Munoz, 2024; Minni, 2024). However, we still know relatively little about how firms make practical organizational decisions and whether they react to different stimuli or adapt under varying environments (Hoffman and Stanton, 2024). One notable exception is Adhvaryu et al. (2022a), who use data from an Indian garment firm to investigate how worker-to-task assignments change in response to short-run productivity shocks caused by pollution. Our study contributes to the literature by documenting how firms adjust their internal organization when a key member of the workforce departs. Unlike temporary shocks, the departure of a key worker represents a permanent shock that may require structural rather than temporary solutions. Indeed, we find evidence of lasting

adjustments to task distribution and operational scale. More broadly, our findings provide novel insights into how firms prioritize certain aspects of performance—such as quality over scale—and tailor their organizational changes accordingly.

Our paper also relates to a large body of empirical research on labor turnover. Many studies have examined the nature of turnover (Jolivet et al., 2006) and identified its determinants, including worker characteristics (Lazear and McCue, 2018), workplace policies (Hoffman, 2017; Adhvaryu et al., 2022b), labor market regulations (Card and Krueger, 1994; Saez et al., 2023), and macroeconomic conditions (Shimer, 2012). Our paper is more closely related to research investigating the consequences of turnover. While most of this literature focuses on the effects of job displacement on workers (Jacobson et al., 1993; Davis and Von Wachter, 2011; Flaaen et al., 2019),[5] much less is known about how worker exits impact firms. Recent studies have started to fill this gap. Jäger et al. (2022) study how worker departures caused by deaths affect coworkers' wages and retention. Sauvagnat and Schivardi (2024) conduct a similar analysis for executive exits, assessing their effect on firms' return on assets.[6] We extend this literature by examining how firms internally reorganize in response to worker exits, uncovering when and why turnover shocks affect firm performance. We also contribute by investigating the impact of these shocks on outcomes that directly affect consumers and, more broadly, the quality of an essential public good (healthcare). One key takeaway from our analysis is that a firm's ability to internally reallocate workers across tasks is central to mitigating disruptions in service quality resulting from the loss of key personnel.

Our paper also builds on a longstanding literature in health economics that studies variation in the quality and style of care across providers (Skinner, 2011; Chandra et al., 2011; Abaluck et al., 2016; Chandra and Staiger, 2020). Our findings indicate that variation in physician turnover and the ways hospitals respond to it can lead to differences in hospital productivity and care quality. At a more micro level, we provide novel evidence on how physician-to-task assignments can shape hospital performance. In doing so, our paper speaks more directly to the strand of this literature focusing on personnel organization as a key driver of medical outcomes, whereas prior work has largely focused on the optimal configuration of medical teams (Chan, 2016; Chen, 2021; Doyle and

[5] These papers have identified persistent earnings losses among job losers, often interpreted as evidence that human capital has a firm-specific component (Becker, 1962; Lazear, 2009). Additional evidence comes from studies documenting that worker productivity and wages rise with tenure (Topel, 1991; Huckman and Pisano, 2006; Buchinsky et al., 2010). This work also provides evidence that firm-specific human capital is important for individual performance. Our results reveal that newcomers struggle to maintain baseline service quality in areas most exposed to the exit of key personnel, whereas physicians familiar with the workplace do not. More importantly, we show firms appear to respond to these differences by limiting the allocation of tasks to newcomers and prioritizing incumbents in more exposed areas. This ultimately preserves the quality of firm service. Therefore, we establish a link between individual-level effects of firm-specific human capital, organizational responses, and firm performance.

[6] A related strand of recent research investigates the effects of employment interruptions driven by family leave and retirement policies on firms (e.g., Ginja et al., 2023; Schmutte and Skira, 2023; Bianchi et al., 2023; Huebener et al., 2024). While economically relevant, such turnover shocks are typically anticipated and, in some cases, temporary. Moreover, these studies focus on workforce size adjustments without examining internal reorganization or how such exits affect firm performance. In addition, there is a more descriptive literature that examines the costs firms incur when replacing departing workers, often relying on surveys about recruitment expenses or personnel records from specific firms (e.g., Hamermesh and Pfann, 1996; Abowd and Kramarz, 2003; Alan, 2011; Kuhn and Yu, 2021). Kuhn and Yu (2021) attempt to provide a more causal answer on these costs in an event-study analysis exploiting variation in voluntary worker departures across different stores of a Chinese menswear firm.

Staiger, 2022; Branco et al., 2024). By examining how the impact of physician turnover varies with the mortality prediction of each physician, we also complement the branch of work on doctor effectiveness (Doyle et al., 2010; Chan and Chen, 2022; Currie and Zhang, 2023; Ginja et al., 2024). Finally, our work contributes to the literature on how healthcare providers balance potential trade-offs between service quantity and quality. Recent studies show that increasing the speed or volume of care can undermine quality (Silver, 2021; Aghamolla et al., 2024). Our paper offers a novel perspective by exploring how hospitals prioritize quality over volume in the face of adverse shocks.

Outline. The remainder of the paper is organized as follows. Section 2 provides brief background information on our setting. Section 3 describes the data, sample, and key variables. Section 4 outlines our baseline research design and presents evidence on the quantitative importance of the shock to hospitals' labor supply. Section 5 investigates how hospitals organize internally in response to the shock. Section 6 examines its consequences for hospital performance, focusing on both the quantity and quality of care. Section 7 explores potential mechanisms behind these findings, emphasizing the role of doctor quality and the ways hospitals may choose to prioritize care quality over quantity. Finally, Section 8 concludes.

2 Setting

Brazilian Health System. Following the enactment of the 1988 Constitution, Brazil established universal access to healthcare as a fundamental right, giving rise to the Unified Health System (SUS). The SUS ensures access to preventive and curative care through both publicly managed institutions and a network of private facilities affiliated with the SUS. Hospitals in the SUS network deliver healthcare services to patients without charging them directly. Instead, hospitals are reimbursed by the government based on the quantity and type of services provided. Reimbursement rates are set nationally and managed by the Ministry of Health. Patients incur no co-payments, coinsurance, deductibles, or any other out-of-pocket expenses under this system.

People who opt for private care outside the SUS can receive ambulatory and hospital services either through out-of-pocket payments or private health insurance plans. The demand for health insurance is mostly from companies offering insurance coverage to their employees. In practice, the majority of the population relies on the SUS for healthcare services. Only a small percentage, approximately 26 percent, opt for private health plans (Paim et al., 2011).

The Physician's Role. Our analysis focuses on the main physician responsible for the care of a patient during their hospital stay. This physician is tasked with the critical responsibilities of diagnosing the patient's condition, determining the urgency of care, and identifying the most appropriate treatment path. For instance, in the case of pneumonia, the physician must decide whether the patient requires immediate treatment with strong antibiotics or if a less intensive approach, like rest and fluids, would be sufficient. This decision often involves assessing the patient's symptoms, reviewing test results, and considering how well the patient is responding to initial treatments to

determine the best next step. In many cases, these decisions are time-sensitive and delays or errors in treatment can significantly worsen the patient's condition (Brennan et al., 1991; Bates et al., 2023). After the primary intervention, they oversee the patient's recovery, monitor for complications, and adjust treatment plans as needed to ensure optimal outcomes.

The physician in charge also supervises other members of the healthcare team, which may include residents, fellows, medical students, nurses, and other specialists. This involves ensuring that the tasks assigned are carried out appropriately and within the required timeframes. They can correct any discrepancies when necessary and give instructions to realign tasks with the overall care goals for the patient.

Hospital Labor Market. The structure of the medical labor market in Brazil reflects the unique characteristics of the country's healthcare system, where physicians frequently work across public and private sectors, often holding multiple jobs simultaneously (Scheffer et al., 2020). In the public sector, employment is primarily organized under the Unified Health System (SUS), with hospitals hiring physicians through a range of contractual arrangements. The most common types include formal contracts and agreements with independent professionals, both of which offer hospitals significant flexibility in managing their workforce, including hiring and terminating employment as needed. While these arrangements differ primarily in terms of tax obligations and access to social security benefits, both enable hospitals to adapt effectively to changing workforce demands.

A key feature of the hospital labor market is the reliance on affiliated and on-call physicians. These are professionals who maintain a formal or informal connection with a hospital—through existing contracts, prior professional relationships, or inclusion in hospital rosters—but may not be actively involved in routine operations. During periods of increased demand or emergencies, they are often mobilized to fill critical staffing gaps. This practice enhances hospital flexibility, which may allow for quick adjustments to fluctuations in patient volume or workforce availability. Additionally, hospitals may extend physicians' shifts strategically to address surges in demand and ensure consistent care during peak periods.

3 Data Sources, Samples, and Variable Definitions

This section describes the data sources, defines key variables, and outlines the empirical strategy.

3.1 Data Sources

We assemble a new dataset combining information from several administrative registries managed by the Ministry of Health, Regional Medical Council, and National Registry Office. The data cover all hospital admissions within the entire public health system, which includes both public hospitals and private hospitals affiliated with the public health system. At the patient level, we observe a set of clinical characteristics such as the cause of admission (as classified by the International Classification of Diseases, Tenth Revision, or ICD- 10 code), details on each procedure received,

date of discharge, and date of death for those who passed away in the hospital. The data also include basic demographic characteristics such as date of birth, gender, race, and place of residence at the zip code level. Importantly, we observe the identity of the main physician for each patient, and unique tax identifiers allow us to track physicians over time across patients.

We have information on a number of background and human resources characteristics covering the universe of physicians and other professionals in the country. Key professional information includes a physician's occupations, hours worked, and the identifier codes of all the establishments for which they provide any inpatient or outpatient services. This administrative information is updated on a monthly basis, which allows us to examine sharp changes in a hospital's labor supply even within intervals shorter than one year. We also observe basic background characteristics, such as gender, date of birth, and experience since the start of medical practice. Key to the purposes of this study, we observe a physician's mortality status and the exact date of death. We have limited information on the cause of death, which is available only for some periods in our sample.

Appendix C provides a more detailed description of these data and the construction of key variables. Our analysis focuses on the 2008-2024 period, a choice determined on the front end by the availability of physician data that can be linked to patients. The data encompass approximately 189.2 million inpatient admissions between 2008 and 2024. To reduce noise in the data, we choose a quarter as the time unit of reference for the analysis.

3.2 Turnover After Death

There are 309,854 physicians throughout the study period. Of these physicians, approximately 6,806 passed away. Many of these deaths could not have come as a surprise to their colleagues. For example, in cases where the physician had frequently been absent from work due to serious or terminal health conditions, colleagues might have been aware of the potential risks to their health. Terminal illnesses, such as cancer or multiple sclerosis, typically require prolonged treatments and close monitoring, making it difficult for individuals to continue working actively. These individuals may even choose to leave the hospital or resign from their positions well before death, in order to focus on managing their condition and receiving necessary care. Therefore, to focus on deaths that are less likely to be anticipated, we restrict the sample to physicians who were actively working at the time of their death, defined as those who provided care to patients during the quarter in which they died. By definition, this excludes physicians who have experienced long sick leave spells associated with terminal, debilitating illnesses. In the end, we have 1,351 deaths, or approximately 20 percent of all deaths.

One might apply further restrictions to better identify unexpected deaths, such as limiting the data to deaths due only to accidents or heart attacks for the subset of physicians with available information on the cause of death. However, because the death of a physician is already a relatively rare event, such restrictions would be somewhat extreme and could significantly reduce the statistical power of our analysis. There is, thus, a trade-off between precision and bias. The extent to which such bias is important and hospital managers are able to accurately anticipate physician

deaths would likely result in noticeable changes in organizational outcomes before the physician's actual departure. In practice, we do not see any signs that this is the case for the large set of outcomes we examine, even for those where such anticipatory effects would be easily visible. In light of this evidence, we focus on our higher-powered and more parsimonious strategy.

Characteristics of Alive and Deceased Physicians. Table 1 shows that deceased physicians tend, on average, to be older, more experienced, more likely to be male, and more likely to have nonclinical duties than alive physicians. This remains true even after removing hospital×time effects, and thus only physicians within the same hospital and time frame are compared (odd-numbered columns of Appendix Table B.2). As such, these differences are a consequence of the law of large numbers. The majority of arguably unexpected causes of death, such as heart attacks, heart failures, and strokes, are more likely to occur at older ages and among males. Consistent with this fact, the differences become substantially smaller once we account for differences in age (even-numbered columns of Appendix Table B.2). The differences between deceased and alive physicians primarily reflect that they are at different stages of their life cycles. In Section 6.3, we explore the implications of these differences for our results.

3.3 Sample and Matched Controls

There are 7185 hospitals in the period of analysis. Of these hospitals, 1023 had at least one physician who passed away on duty. About 23 percent experienced exactly two events, and 17 percent had three or more deaths. Among those with at least two deaths, the average interval of time between consecutive deaths is 15 quarters. For our analysis, in cases of hospitals with multiple events, we include all such events by stacking the data and treat each hospital-death pair as a separate unit of analysis. For example, if a hospital witnesses two deaths, we create two separate copies of that hospital, a time series for each event. The baseline sample consists of 1347 hospital-event pairs.

Physician deaths are not randomly distributed across hospitals. They are more likely to occur in larger hospitals with more resources and a higher volume of patients (Columns 1 and 3, Appendix Table B.1). This is consistent with the fact that the death of a physician is a rare event that is expected to be observed more frequently in larger populations. Although such differences are not necessarily a problem for identification, we implement a two-step matching procedure to identify comparable hospitals. In the first step, we estimate the propensity to experience the death of a physician based on a set of predetermined characteristics. These characteristics include patient volume, teaching status, offering of specialized services, presence of emergency department, presence of operating rooms, and number of physicians, nurses, and beds. In the second step, we compute the Mahalanobis metric and identify control hospitals as those with the smallest distance from the treated hospitals in terms of the propensity score and set of pre-event characteristics from the first step. Appendix Table B.1 documents that this approach essentially eliminates differences between treated and untreated hospitals.

3.4 Risk-Adjustment and Aggregation

In our analysis of performance, we examine changes in hospital mortality and other patient outcomes. Part of the variation in patient outcomes reflects differences not only in performance but also in the types of patients admitted to the hospital. We implement a two-step procedure to reduce the variation that is due to patient characteristics. In the first step, we regress a given outcome variable, y_{it}, for patient i in quarter t on the vector of patient characteristics, $\mathbf{X}_{it}$:

$$y_{it} = \beta \mathbf{X}_{it} + \mu_{jht} + \xi_{it}$$

Where μ_{jht} represents a fully interacted set of cause-of-admission $\times$ hospital $\times$ quarter-year indicators. In the second step, we calculate the risk-adjusted outcome as follows:

$$y^*_{it} = y_{it} - \hat{\beta} \mathbf{X}_{it} \equiv \mu_{jht} + \xi_{it}$$

The vector of patient characteristics includes age, gender, race, zip code of residence, as well as a set of admission-level comorbidities. Our risk-adjustment procedure essentially uses within hospital-disease-quarter variation to estimate β, which reduces biases due to omitted factors correlated with patient outcomes and performance. To see this, suppose that high-mortality risk patients sort into high-quality hospitals. By estimating β without controlling for hospital fixed effects, we could eliminate all of the variation in outcomes across hospitals that is due not only to patient case mix but also to differences in hospital quality.

To ease the computational burden, we collapse the residualized outcomes at the hospital-by-cause-of-admission-by-quarter level and weight all regressions by cell size to account for differences in the precision with which the cell means are estimated. This is the level of aggregation used in our main analysis. Similar approaches have been applied in a variety of other settings (Shapiro, 2006; Albouy, 2009; Rossin-Slater and Wüst, 2020), and Donald and Lang (2007) demonstrate that this type of two-step procedure to estimate treatment effects is equivalent to estimating individual-level regressions.

4 Research Design

4.1 Areas of Care and Exposure to the Deceased

Not all separations are equally relevant for a hospital. The exit of a physician who managed a limited caseload may have a negligible impact on hospital operations, especially in high-volume hospitals with a diverse range of medical staff. Additionally, there could be important differences within hospitals, as physicians generally focus on treating specific types of patients. For example, some physicians may focus on conditions such as ischemic heart disease, arrhythmias, or heart attacks, while others may specialize in managing orthopedic injuries, or respiratory disorders. This structure of medical care in hospitals is analogous to the segmented nature of manufacturing processes in a factory. Different production lines handle specific stages of manufacturing. Therefore, the departure

of a physician is likely to have a more significant impact on the areas or "production lines" where they were most involved.

To capture these nuances and make the most of all available information, we measure the extent to which a given area of care within a hospital was concentrated in the hands of the deceased physician. Areas of care or disease areas are defined using information on the specific cause of admission of the patients, categorized according to the relevant 3-digit ICD codes.[7] The classification is relatively narrow. For instance, within the ischemic heart disease category, there are separate codes for myocardial infarction, angina pectoris, and chronic ischemic heart disease. We then compute the exposure of a disease area to a physician as the share of patients within that specific area who that physician treated during their hospital stay:

$$\text{Exposure}_{js} = \frac{N_{js}}{\sum_{k \in \mathcal{K}(s)} N_{ks}} \tag{1}$$

Where N_{js} is the number of patients managed by physician j within area of care s, and $\mathcal{K}(s)$ is the set of physicians providing care to those types of patients. The basic idea is that the more patients the physician treated, the more exposed the disease area is to their expertise, and therefore, the greater the potential impact of their exit. When the deceased's patient share approaches 1, it indicates that all those patients were treated exclusively by that physician. Appendix Figure A.2 shows that there is substantial variation in the exposure to the death event. Within the disease areas where deceased physicians practiced, the standard deviation of the exposure measure is about 20 percentage points.

4.2 Specification and Identification

Our baseline design compares areas of care that differ in their exposure to death-induced exits across hospitals over time in a difference-in-differences setup. The matched comparison controls represent a specific case where all areas of care had no exposure to the death of any physician. To illustrate the idea behind this approach, consider three hospitals treating cases of myocardial infarction, two of which experienced the death of a physician in the same quarter. In the first hospital, nearly all admissions related to cardiovascular conditions were managed by the deceased physician, while in the second hospital, these cases were more evenly distributed among the staff. The labor supply shock was therefore felt to a greater extent in the former than in the latter hospital. The third hospital did not face any labor supply shock, thus serving as a comparison group for the other hospitals. Under a common trends assumption, one could isolate the causal effects of the shocks to the labor supply of hospitals by comparing the outcomes related to cardiovascular admissions of the three hospitals before and after the event.

We can generalize from the example above by pooling all types of hospital admissions together and estimating regressions controlling for hospital-by-area-of-care and time effects. A complication is that there is variation in the timing of deaths across different hospitals. A recent literature has established that, because the conventional two-way fixed effects estimator is a weighted average of all

[7] We use the terms "disease area" and "area of care" interchangeably.

possible two-group/two-period before-after comparisons, it is biased in the presence of variation in treatment timing and heterogeneous treatment effects (De Chaisemartin and d'Haultfoeuille, 2020; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; Goodman-Bacon, 2021). Our approach to avoid biases from "forbidden" comparisons is a stacked difference-in-differences design (Wing et al., 2024; Butters et al., 2022; Cengiz et al., 2019). To implement this approach, we first construct a separate dataset for each event q sub-experiment. Each dataset s is composed of hospitals that experienced the death of a physician in event quarter q and of their respective matched hospital controls. We then append all these sub-samples into a single file to form the final stacked sample. We estimate the parameters of interest using this stacked sample and control for a full set of sub-experiment fixed effects interacted with unit and time fixed effects. Because the matched control hospitals are chosen with replacement from the pool of possible candidates, some control hospitals may be over-represented in the sample as a whole. We use corrective weights to offset this over-representation, as suggested by Wing et al. (2024).

We use the following event-study specification to estimate the effect of death-induced exits:

$$\begin{aligned} y_{ihts} = \alpha + \sum_{\tau} \beta_{\tau} \text{Exposure}_{ihs} \times \mathbb{1}[\text{quarter since death} = \tau] \\ + \ \mu_{ihs} + \ \lambda_{its} \ + \eta_{ihts} \end{aligned} \tag{2}$$

where y_{ihts} is the outcome of interest for area of care i within hospital h observed in quarter t and event sub-sample s. *Exposure* is the time-invariant, continuous measure that captures the extent to which the area of care relied on the deceased physician, based on equation (1). It is zero for matched comparison hospitals by definition. The coefficients $\{\beta_{\tau}\}$ provide information on the path of outcomes around the event. The time-to-treatment of comparison is $\tau = -1$, a quarter before the death. The term μ_{ihs} controls for time-invariant characteristics of areas of care specific to each hospital. With the inclusion of area-of-care-by-quarter-year-by-subsample fixed effects, λ_{its}, the parameter of interest is identified primarily by comparing the same types of admissions across different hospitals —not between different areas of care within the same hospital. The residual term, η_{ihts}, is clustered at the hospital level to allow for serial correlation.[8]

To summarize magnitudes in table format and perform specification checks parsimoniously, we also estimate a reduced version of (2):

$$\begin{aligned} y_{ihts} = \alpha + \beta_0 \text{Exposure}_{ihs} \times \text{Short-Run}_{ts} \ + \ \beta_1 \text{Exposure}_{ihs} \times \text{Long-Run}_{ts} \\ + \ \mu_{ihs} + \ \lambda_{its} \ + \eta_{ihts} \end{aligned} \tag{3}$$

The interaction between *Exposure* and *Short-Run* captures the effects in the two quarters immediately following the event. In the second interaction term, the variable *Long-Run* is an indicator that takes the value of 1 from the third period onwards after the event, and 0 otherwise. It summarizes the effects of the labor supply shock over a longer horizon. The remaining variables and coefficients are

[8] Appendix Table B.10 shows our results are barely affected when we implement alternative inference procedures: (i) clustering by matching pair; and (ii) clustering by hospital (as in our main specification) with the degree-of-freedom corrections suggested by Young (2016).

identical to those outlined in equation (2). Because changes in the quarter of the event are difficult to interpret in some cases, since some deceased physicians handled cases, we exclude that period to avoid any ambiguity when interpreting the short-run coefficient.

Identification. A causal interpretation of our estimates requires the assumption that areas of care more and less exposed to the event would have followed the same trends in the absence of the event. This assumption could be violated if the timing of deaths coincides with other shocks affecting hospital performance. There is no reason to believe that factors key to hospital performance, such as hospital funding, staffing shortages, demand for medical care, or shifts in healthcare policies, are systematically correlated with physician deaths. Of greater concern might be natural disasters and other large-scale adverse events that increase mortality in the general population, including physicians. In practice, such events are very infrequent and thus unlikely to introduce an important bias. Consistent with this logic, Panel B of Appendix Table B.11 documents that our key results hardly change when excluding deaths that occurred during one of the most severe communicable disease outbreaks in the last 100 years: COVID-19 pandemics. We also evaluate the robustness of the estimates to excluding hospitals that experienced multiple deaths in the same quarter, which excludes those due to larger disasters, without any material impact on our estimates (Table B.11, Panel C).

Another possible concern is anticipation effects. If the hospital administration can anticipate the physician's death, they may take proactive steps to minimize disruptions to workflow and maintain operational continuity. For example, the hospital could hire new physicians in advance, or reallocate patients and other duties to other physicians. Because we focus on physicians who were actively working at the time of their death, which excludes those whose deaths were preceded by prolonged sick leaves —arguably the most likely to be foreseen—, the risk of anticipation effects should be reduced. Prior studies have shown that firms typically adjust their staffing levels in advance when they anticipate a worker's exit (e.g., Huebener et al., 2024; Schmutte and Skira, 2023). We can thus evaluate the presence of anticipation effects and, more generally, the plausibility of the identifying assumption using the event-study specification above. If the identifying assumption is valid, the trajectory of the quarter-specific coefficients should be flat prior to the event and begin to diverge only after the physician's death. We will show throughout the paper that the dynamics of pre-event trends are broadly consistent with the validity of our statistical approach. In Appendix Table B.12, we summarize the results of pre-trend tests for key outcomes and show that in no case is possible to reject the null hypothesis that all pre-treatment coefficients are jointly equal to zero.

We estimate the models above on a fully balanced sample of hospitals both before and after the event. This implies excluding observations with events in the first and last quarters of our dataset. The longer the event study, the more observations we tend to lose. To reduce the loss of observations and explore longer-run dynamics, our window of analysis starts 4 quarters before the event and extends to 10 quarters after the event. Since the average interval between events is 15 quarters among hospitals experiencing multiple events, this choice minimizes temporal overlaps between events. In Appendix Table B.12, we show that the tests for pretrends are robust to alternative

window choices that include longer pre-event periods.

4.3 Changes in a Hospital's Labor Supply

Before turning to our main findings, we first examine changes in a hospital's labor supply. We look at the log number of affiliated physicians and newly affiliated physicians, converted into full-time equivalent (FTE) physicians assuming a full-time workweek of 40 hours. Since a physician's affiliation varies essentially at the hospital level, we estimate the hospital-level version of the event-study specification given by equation (2):

$$\begin{aligned} y_{hts} = \alpha + \sum_{\tau} \beta_{\tau} \text{Exposure}_{hs} \times \mathbb{1}[\text{quarter since death} = \tau] \\ + \ \mu_{hs} + \ \lambda_{ts} \ + \eta_{hts} \end{aligned} \tag{4}$$

Figure 1 display the results. There are no differential trends between more and less exposed hospitals before the event, consistent with no relevant anticipation effects and supportive of the parallel trends assumption. Immediately following the death, there is a sharp decline in the number of FTE physicians in the first quarter, with a gradual recovery over time and the estimated effects becoming insignificant by the fifth period. In line with this result, we observe an immediate increase in the number of newly FTE physicians, followed by a subsequent decline in the magnitude of treatment effects.[9] We also investigate effects on retention, which could have been another important margin of adjustment. To do so, we explore changes in the log number of FTE separations, excluding that of the deceased to avoid mechanical effects (Appendix Figure B.2). There is a decline in separations in the quarter of the event, but the estimated coefficients for subsequent periods are notably smaller in magnitude and statistically insignificant.

Quantitatively, our estimates imply that moving from zero to full exposure would lead to a decline of 22 percent in the number of FTE physicians in the quarter immediately after the event. At the hospital level, this is an extreme comparison in view of the low share of cases handled by the typical physician. In terms of a one-standard-deviation increase in the exposure measure, the point estimate implies a decline of 4 percent in the number of FTE physicians in the quarter immediately after the event. The magnitude of this effect is unsurprisingly small, as the relative importance of a single physician for the hospital as a whole is minimal. However, this can represent a large decline in the area where the deceased physician practiced within the hospital. A simple back-of-the-envelope calculation that rescales this effect by the share of the deceased's main specialty among the hospital's medical staff suggests a decline of approximately 18 percent in the supply of physicians in the relevant area of practice.

To approximate the path of staffing levels without any hiring or retention response, the dashed

[9] To account for the fact that many hospital-quarters have no hiring activity, we follow the approach recently recommended by Chen and Roth (2024) and calibrate the extensive margin. Specifically, we assign a value of 100 log points to an extensive margin change between 0 and the minimum non-zero value of the dependent variable. Appendix Figure B.1 shows that these results are robust to alternative calibrations of the extensive margin. For example, the results are statistically identical to the baseline if we assign values of 25, 50, or 75 log points changes to an extensive margin change between 0 and the minimum non-zero value of the dependent variable.

line in Panel A of Figure 1 shows the evolution in the log number of FTE physicians assuming that the staffing level in the post-event period would have remained the same as the one observed in the quarter before the event but without the deceased physician. This calculation also adjusts for the probability of exit if the physician had survived, which reduces the quantitative importance of the shock over time.[10] The calculations suggest that, in the absence of a hiring or retention response, the decline in staffing levels would have been 3.6 times larger in the quarter of the event. This pronounced divergence indicates that hospitals mitigate the shock quickly.

Characteristics of Physicians. Although hospitals replace deceased physicians in headcount relatively quickly, replicating their attributes may be far more challenging. We explore this question in Appendix Table B.3. The data reveal that the death event is associated with a significant decline in age, experience, and share of male physicians, consistent with the fact that deceased physicians are predominantly seniors and that older cohorts of physicians in Brazil tend to be male. In contrast, we find no major changes in the number of specialties. This could reflect that, since medical specialization typically occurs early in a physician's career,[11] specialization is weakly correlated with seniority.

Overall, the results suggest that the death of a physician significantly disrupts hospital staffing, but hospitals respond quickly through increased hiring and retention. The adjustment was accompanied by a shift toward younger and less experienced staff.

5 Organizational Responses

This section investigates how hospitals organize internally in response to the labor supply shock.

5.1 Demand for Incumbents

We start by exploring how the shock to hospital's labor supply affected the demand for incumbents. Although hospitals adjust relatively quickly in terms of staffing levels, new physicians could require time to familiarize themselves with the local hospital practices, and incumbents could absorb the additional workload during the transition to minimize any drop in hospital performance. To explore this question, we estimate the physician-level analogous of model (2):

$$\begin{aligned} vy_{jhts} = \alpha + \sum_{\tau} \beta_{\tau} \text{Exposure}_{hs} \times \mathbb{1}[\text{quarter since death} = \tau] \\ + \ \mu_{jhs} + \ \lambda_{ts} \ + \eta_{jhts} \end{aligned} \tag{5}$$

The outcome of interest is y, measured for physician j affiliated to hospital h in quarter t and sub-experiment group s. Note that the exposure variable is defined at the hospital level. We exploit

[10] It is important to note that even if the physician had survived, there was still a chance that she or he might have left the hospital during the post-event period for other exogenous factors. We use data from all physicians to compute the cumulative probability of exit over time and use it to adjust our counterfactual calculations.

[11] In Brazil, 60% of specialist physicians start their specialization within one year of graduation from medical school, and 90% do so within three years (Scheffer et al., 2023).

within physician-hospital variation and control for a rich set of physician-by-hospital-subsample and quarter-year-by-subsample fixed effects. We also present the average short- and long-run estimates to summarize magnitudes in table format. The coefficients represent the effect of moving from a hospital without any death to another where all cases were entirely concentrated in the hands of the deceased physician. Because this is an extreme comparison at the hospital level, we interpret the results in terms of the effect of a one-standard-deviation increase in the exposure measure.

5.1.A Case Load

Figure 2 presents the results from estimating the dynamic model (5) for the log number of patients, whereas Table 2 shows the average treatment effects for the short and long run. We observe a sharp rise in the number of patients seen by incumbents in the immediate aftermath of the colleague's death. The short-run coefficient of 0.43 implies that physicians practicing in a hospital one standard deviation higher in the exposure distribution experienced an increase of 8.6 percent in their patient volume (0.43×0.20). The magnitude of this effect decays gradually over time but remains statistically distinguishable from zero over the long term. Our estimates in Section 4 indicate that hospitals have adjusted in terms of staffing levels by the fifth quarter after the event. The persistent effect on incumbents' patient load suggests that the newly affiliated physicians do not fully compensate for the loss of the deceased physician, leading to a sustained redistribution of the workload among existing staff.

Note that the increase in the incumbents' patient load does not automatically imply that hospitals were able to completely maintain their overall pre-event operation levels. There is certainly a limit to the number of patients a physician can manage, and if this limit is binding, the hospital's overall capacity may still have been constrained by the loss of the deceased physician. We will return to this point in Section 6.1.

5.1.B Broadening Clinical Scope

The increase in the patient load of incumbents coincides with a significant broadening of the types of conditions they treat. Indeed, Panel B of Figure 2 and column 2 of Table 2 show a significant increase in the number of distinct primary diagnosis codes incumbents manage after the colleague's death. Each additional standard deviation increase in the exposure to the deceased leads to 0.74 new distinct admission causes managed in the short run and 0.44 in the long run. This represents an increase of 5 to 9 percent relative to the pre-event mean of 8.7. The diagnosis categories are based on 3-digit ICD-10 codes, which roughly correspond to a disease. In column 1 of Appendix Table B.4, we instead consider the number of distinct sub-chapters of the ICD classification, which correspond to a broader group of related diseases within a disease category. The results remain quantitatively and qualitatively similar.[12] Incumbents have not only increased the intensity of the

[12]For example, within the "Diseases of the respiratory system" chapter (ICD-10 codes J00-J99), one sub-chapter includes "Acute upper respiratory infections" (codes J00-J06). Another example is the "Diseases of the circulatory system" chapter (ICD-10 codes I00-I99), where a sub-chapter might include "Ischemic heart diseases" (codes I20-I25). Each sub-chapter

tasks they already perform but have also taken on a broader variety of clinical duties.

5.1.C Non-Clinical Duties

The negative shock to the hospital's labor supply may also have affected the distribution of non-clinical duties, such as research, education, and administrative roles within the hospital, which are crucial for the functioning of the institution. Deceased physicians tend to be more experienced and older on average, and senior physicians are generally more engaged in such institutional roles due to their expertise and familiarity with the hospital's operations, policies, and culture. Thus, the exit of such physicians implies that such responsibilities need to be reassigned, and incumbents may be viewed as the most natural candidates to take on these roles. Panel C of Figure 2 examines changes in an indicator for holding nonclinical duties and shows that this is indeed the case. Incumbents are significantly more likely to hold nonclinical tasks following the colleague's death. These effects are persistent and economically large, with a one standard deviation increase in the exposure to the deceased implying an increase of 100 to 145 percent relative to the pre-event mean. The persistence of this effect may indicate a preference for physicians with greater tenure for these tasks.[13]

In sum, hospitals respond to a physician's exit by significantly increasing incumbents' workloads, broadening their responsibilities, and assigning additional non-clinical tasks.

5.1.D How Do Physicians Absorb Extra Work?

The results so far suggest that incumbents have taken on greater clinical and nonclinical responsibilities. But how do they manage the increased workload? Incumbents may have accommodated the additional patient load by working longer hours or increasing productivity (i.e., seeing more patients per hour). In column 2 of Appendix Table B.4, we examine how physician turnover after death affects incumbents' total of hours worked (columns 3 and 4, Table B.4). To analyze possible changes across the distribution, we also consider indicators for whether the physician is in the bottom or top quartile of the working hours distribution. There is no evidence that incumbent physicians work significantly longer hours following the exit of a colleague. While the point estimates are positive, it is not possible to reject the null hypothesis that they are statistically identical to zero. Not surprisingly, scaling the clinical production of incumbents by the hours worked has no material impact on the conclusions (column 5, Table B.4).

These results suggest that incumbents absorb the additional workload primarily through increased productivity. They produce more spending the same amount of time.

contains diseases that are closely related to one another within the broader disease category.

[13]Hospitals may have limited resources, and it may be more efficient for the remaining incumbents to continue managing nonclinical tasks. Training new staff to handle these responsibilities can be time-consuming and costly, especially if the nonclinical duties are complex and require a deep understanding of the hospital's operations. Therefore, the incumbents may retain these roles because they are already performing them effectively and can manage both clinical and nonclinical responsibilities.

5.2 Redistribution Across Disease Areas

This section investigates hospital responses at the more granular level of disease area. We investigate how labor is redistributed internally across areas of care. Our unit of analysis is at the hospital-by-area-of-care, and the estimating equations follow models (2) and (3).

5.2.A Physicians Available for Care

Although hospitals respond to the negative labor supply shock by bringing in new affiliated physicians, the adjustment process is not immediate. In the short run, hospitals must then find ways to ensure that all areas of care continue to operate. In doing so, hospitals may increase the patient load for physicians already assigned to affected areas of care, rather than relying on physicians from other areas. In this case, the death of a physician would result in a one-for-one decline in the number of different physicians available in a given area of care. Alternatively, incumbents from other areas of care may temporarily expand their scope of practice to manage patients in affected areas. This may be especially true for physicians with generalist training or those with overlapping skills relevant to the vacant position. For instance, an internist might take over managing some less complex conditions typically handled by a specialist. The results from the previous subsection provide evidence suggesting that incumbents indeed take on a broader range of conditions than they typically manage. If this expansion in the scope of practice is sufficiently important, we should observe limited changes in the supply of physicians available for care in an area.

Panel A of Figure 3 shows the dynamic treatment effect on the log number of physicians in an area of care. As in the hospital-level analysis, we convert these numbers into FTE physicians. There is no reduction in labor supply across different areas of care in any meaningful way, either in the short run or long run. While the point estimates are slightly negative across event times, they are economically small and statistically insignificant. Table 3 displays the average treatment effects for the short- and long-run periods. Column 1 uses the number of FTE physicians in logs. The rest of the columns explore the robustness of the results to alternative specifications. Column 2 uses simple headcounts of physicians (in logs) as the dependent variable, while columns 3-4 employ count data models (Poisson and Negative Binomial) instead of OLS. These results confirm the muted effects of physician death on labor supply across disease areas. The results remain small and statistically insignificant, irrespective of the estimation strategy. These results suggest that the expansion of physicians' scope of practice, treating a broader range of conditions, has played a major role in mitigating disruptions to the supply of physicians available for care in more affected areas.

5.2.B Skill-to-Task Mismatch

Given that physicians are managing a wider range of medical conditions than they normally would, a natural question is whether this has led to a mismatch between their expertise and the medical needs of their patients. For example, an anesthesiologist providing treatment for a heart attack

because a cardiologist or internist is no longer available would constitute a mismatch.[14] We generate a measure of mismatch for each patient-physician pair using data from the Ministry of Health, which determines the list of specialties considered clinically relevant for each procedure.[15] In our sample, approximately 11 percent of physician-patient pairs are mismatched, and mismatch is a strong predictor of inpatient mortality, suggesting that patients may indeed receive suboptimal care when treated by physicians without the relevant specialization.[16]Appendix Figure B.3 shows no evidence that patients are more likely to be mismatched following the physician's death. The estimated coefficients are small and statistically indistinguishable from zero, both before and after the event.

We view this null result as evidence consistent with a strong organizational capacity of hospitals. Although hospitals need to expand the scope of medical practice for physicians when one of their colleagues leaves, they manage to do so without significantly worsening the physician-patient match quality.

5.2.C Allocation of New Entrants

We now look at the caseload of new entrants in the quarter they start operations in the hospital. New entrants may need time to familiarize with the local hospital environment, and this transition may be particularly challenging in more affected areas. In such areas, there is a greater gap in service provision following the physician's death. Hence, new physicians entering these areas may face a heavier workload and increased pressure, which may increase the risk of fatigue-related errors such as misdiagnoses, complications, and treatment delays (Weissman et al., 2007). Moreover, new physicians might struggle to integrate into teams that are already strained by the increased workload, further complicating their ability to learn and adapt efficiently. Given these challenges, hospitals might opt to reassign only a portion of the deceased physician's caseload to new physicians in more exposed areas.

Panel B of Figure 3 and column 5 of Table 3 examine the probability that a patient is treated by a physician that started their clinical duties in that quarter. If new physicians take over the entire service gap left by the deceased physician, the coefficients immediately after the event should be close to 1. As can be seen from the figure, we observe a jump in the share of patients attended by new physicians in the short run, but the estimated coefficients are very far from 1. The largest

[14]While anesthesiologists are skilled in managing critical care and physiological monitoring, their expertise is primarily focused on perioperative care and anesthesia management, rather than on the acute medical management of cardiac emergencies.

[15]To be more precise, we use data from the Brazilian health system's procedure management tool, known as *SIGTAP*, which includes descriptions of all inpatient procedures along with a complete list of the specialties considered relevant for each procedure for billing purposes. As an example, consider the clinical treatment of acute stroke. The specialties listed as relevant include neurology, pediatrics, internal medicine, intensive care, and geriatrics. We say that there is a mismatch between the physician's expertise and this procedure if the physician's specialty is not one of those listed as relevant for the treatment of acute stroke.

[16]Controlling for hospital and time fixed effects as well as patient characteristics, we find that patients who are mismatched for the procedure are 2.2 percentage points more likely to die during their hospital stay. The magnitude of this point estimate is large relative to the baseline mortality rate of 5.2 percent.

estimate implies that new entrants assume about 14 percent of the deceased's caseload. Over time, this effect goes to zero, as hospitals do not disproportionately allocate new cohorts of entrants to areas with greater exposure to the deceased.

A complication in interpreting these results is that they also capture the extensive margin of hiring. The small effects on the allocation of patients to new entrants may simply indicate that some hospitals were unable to recruit promptly, rather than a strategic allocation response of hospital managers. To get a better sense of magnitude, we can rescale the estimated effects by the death-induced change in the hospital's share of new entrants.[17] The rescaled effect would then be interpreted as the patient share of new entrants among hospitals with new physicians. The dashed line in Panel B of Figure 3 depicts the rescaled effects for quarters 0 and 1 since the event and documents that they are still relatively small. The largest estimated effect moves from 0.14 to around 0.23. This indicates that while new physicians have a greater portion of the patient share in more exposed areas, which most likely reflects an increased demand for physicians in these areas, they are very far from replicating the deceased's patient distribution.

A related concern is that hospitals may encounter difficulties in finding a suitable replacement with the necessary skills to perform the medical procedures previously handled by the deceased physician. This may explain the relatively low share of patients allocated to new entrants. For example, suppose that the hospital lost an interventional cardiologist and is able to find only internists in the market. In this case, all patients undergoing invasive coronary interventions may be assigned to incumbent cardiologists due to skill constraints rather than deliberate efforts by managers to facilitate the transition of new entrants. Appendix Figure B.4 shows that this is not the case. It displays the dynamic treatment effects on the ratio of new entrants to the total number of physicians, distinguishing between new hires with high and low overlap with the deceased in the medical procedures that they are qualified to perform. As the figure illustrates, nearly all of the increase in new hires comes from those whose qualifications align closely —over 90 percent —with the medical procedures previously performed by the deceased physician.

Overall, hospitals rely primarily on incumbents to manage the labor supply shock in more exposed, strained areas.

5.3 Nurses

Nurses have a central position in the hospital and thus may play a critical role in managing staffing shortages. As such, nurses may face an increased intensity of work when a key physician suddenly leaves. They may be required to spend more time monitoring patients, take on greater roles in care coordination across departments, handle a higher volume of discharge planning tasks, and update patient records more frequently. Detailed data on interactions between patients and nurses are not systematically collected in Brazil. Nevertheless, we can examine turnover. During times of uncertainty, hospitals may prioritize retaining nurses who are already familiar with the local clinical

[17] In practice, we use the hospital-level version of equation (2) to estimate the dynamic treatment effect on the probability of observing a new entrant. We use the estimated effects for quarters 0 and 1 as rescaling factors.

environment and thus better equipped to handle the increased workload. Moreover, hiring new nurses may require time and effort for recruitment and training, which can further strain hospital resources and delay the adjustment process.

We explore this question in Appendix Table B.5 using the hospital-level version of model (3). There is a significant decline both in separations and hiring of nurses, without meaningful changes in the hospital's overall nursing staff levels. This evidence is consistent with an increased demand for incumbent nurses to deal with the additional workload. Increased retention reduces the need for new hires and this most likely explains the decline in hiring activity.

6 Effects on Performance

In this section, we examine whether hospitals have been effective in offsetting any adverse consequences of the labor supply shock, or whether it has implied a trade-off between the quantity and quality of care delivered.

6.1 Effects on Quantity of Care

6.1.A Admission Decisions

Hospitals have considerable discretion over admission decisions and may adjust their admission practices as a last resort if other organizational changes are not sufficient to preserve care quality. Hospitals might achieve this by redirecting patients to other facilities, implementing stricter admission thresholds, postponing non-urgent surgeries, or reducing transfers and referrals for non-urgent cases. Figure 4 presents event study estimates of the treatment effects on the log number of patients in a disease area, while column 1 of Table 4 provides the corresponding results in regression format. The trend in the pre-event period is flat. Then, we observe a significant and persistent decline in patient volume, with an initial reduction of 9.3 percent during the first two quarters after the event in fully exposed areas. This effect deepens slightly over time, averaging 11.7 percent in the long run and peaking at 15 percent in later event times. Columns 2 and 3 of Table 4 show that the results are robust to using count data models. For example, the long-run coefficient goes from -0.1036 to -0.1006 when the model is estimated via Poisson Pseudo-Maximum Likelihood technique.

Table 4 also explores alternative measures of volume productivity. Column 4 examines the number of patients per bed, while column 5 looks at the patients per physician ratio. These measures capture the efficiency with which hospitals utilize their physical and human resources. Both metrics decline significantly following the exit of a physician, mirroring the magnitude of the reduction in overall patient volume. These results indicate that the volume declines are not driven by a proportional contraction in hospital inputs but instead reflect a deterioration in operational performance.

Because the unit of analysis is a disease area, a possible concern is spillover effects. The decline in patient volume in one disease area could potentially be offset by an increase in another, and, as a result, the aggregate impact may be negligible. Appendix Figure B.5 replicates the event study

analysis at the hospital level and shows that the treatment effects are very similar to the baseline. These findings reduce concerns about spillovers and suggest the observed reduction in hospital-wide output is primarily driven by disruptions in production within more affected areas.

Where Do Hospitals Reduce Care? Hospitals admit fewer patients in response to the labor supply shock, but providers that value the health of their patients may target such cuts where they will have the least detrimental impact. Providers could prioritize severe conditions and defer or manage non-urgent cases through outpatient care. Column 1 of Table 5 reports the treatment effects on the predicted inpatient mortality rate. We predict mortality by regressing it on patient characteristics, including demographics and indicators for a comprehensive set of admission-level comorbidities. The estimated coefficients are positive, but only statistically significant for the long-run coefficient. These results may hide heterogeneity in treatment effects across the mortality risk distribution. Columns 2 and 3 look at the share of patients above the 75 and 90th percentiles of the mortality risk distribution, respectively. There is clear shift in the distribution of cases toward in the top of the mortality risk distribution. Our estimates point to a short- and long-run increase in the share of cases above the percentile 90th of 1.5 and 1.7 percentage points in areas fully exposed to the event. These findings suggest that hospitals indeed prioritize critical care, scaling back services where the impact is likely to be least harmful.

6.1.B Intensity of Care

If admitting fewer patients is not enough to manage the strain caused by the sudden exit of a physician, hospitals may also scale back services during a patient's stay. This reduction in the intensive margin of care quantity can take several forms. Hospitals may cut down on diagnostic tests, particularly routine ones for patients already showing signs of improvement. They could also reduce the frequency of specialist assessments. Instead of daily rounds by specialists, patients may see them only every few days, especially if the patient's condition is stable. If patients receive fewer services, the length of stay might also be affected. The fewer services provided, the less time a patient may need to remain hospitalized. Moreover, patients with milder conditions could be quickly discharged and sent home with instructions for follow-up in outpatient settings.[18]

In columns 4 and 5 of Table 5, we examine changes in the intensity of care. We focus on the total number of procedures and length of stay. A challenge in carrying out this analysis is the observed change in the pattern of admissions. If patients with higher mortality risk receive fewer procedures, this could introduce a bias toward finding a spurious decline in the intensity of care.[19] To reduce biases due to changes in the composition of admitted patients, we implement the risk adjustment

[18] For example, a patient presenting with a mild case of pneumonia might be discharged after a short hospital stay once their symptoms are under control, with instructions to continue oral antibiotics at home and schedule a follow-up visit with their primary care physician to ensure full recovery.

[19] If a patient dies during an intervention, they would not receive rehabilitation services, and their length of stay would mechanically be shorter. Relative to otherwise identical patients, those with higher mortality risk may be associated with fewer services received. In this case, changes in the composition of patients toward those with higher mortality risk may lead us to find a decline in the number of services offered, even if providers do not actually alter their care practices.

and aggregation described in Section 3.4. We then estimate the baseline specification (3). We find no evidence of a statistically meaningful change in the amount of care delivered. While the coefficients are negative, they are small and insignificant. For example, disease areas with full exposure to the deceased experience a statistically insignificant decline of only 0.03 days in the length of stay in the short run. Thus, while patient volume decreases following the exit of a physician, the intensity of care provided to individual patients remains largely unchanged.

6.2 Effects on Quality of Care

Although the death event does not affect the quantity of care patients receive, it could still affect the quality of care delivered. As such, the number of procedures and length of stay are unlikely to be the only determinants of care quality. Factors such as the timeliness of care, quality of decision-making, and team dynamics could have been disrupted even if the same procedures were performed. Moreover, staff might be overworked in the immediate aftermath of the event due to the need to cover for the absent physician. This increased pressure could lead to fatigue-related errors, such as misreading patient charts, administering incorrect medication dosages, overlooking subtle signs of patient deterioration, or delaying critical interventions. It thus remains important to empirically examine potential changes in care quality.

Inpatient Mortality. Our key metric of care quality is inpatient mortality, which has its own strengths and limitations. It is well-defined, welfare-relevant, and exhibits considerable variation across hospitals, arguably driven by differences in the quality of care delivered. However, the vast majority of cases do not involve life-threatening conditions. Inpatient mortality is unlikely to provide relevant information on care quality, for example, for young individuals admitted because of fractures, sprains, superficial wounds, minor burns, or concussions. While common, these cases rarely involve significant mortality risk, and detecting treatment effects would thus require a design with extremely high statistical power. We therefore focus on the subsample of patients who are at risk of mortality, for whom small differences in care quality could translate into large differences in survival chances. As previously mentioned, we define these at-risk cases as patients who are in the top quartile of the predicted mortality risk distribution.[20] The mortality rate in this subsample is approximately 13 percent. This approach of focusing on patients with significant mortality risk is similar in spirit to recent work on the determinants of care quality (Batty and Ippolito, 2017; Silver, 2021).

The focus on patients at risk of mortality is also compelling because these cases often involve acute conditions that require immediate care, leaving little room for selection.[21] Without prompt medical intervention, such conditions can result in irreversible damage or death. Consistent with this reasoning, column 1 of Table 6 shows that treatment effects on the predicted inpatient mortality

[20] We find similar results when we instead define at-risk patients as those above percentile 80, 85, or 90th of the predicted mortality distribution (Appendix Figure B.6).

[21] Appendix Table B.6 displays the top 10 admission causes in this subsample. Conditions such as heart failure, myocardial infarction, strokes, sepsis, and fracture of the femur are particularly overrepresented in this group.

for this subsample are small and statistically indistinguishable from zero. We are thus less worried about changes in the composition of patients in this subsample.

Figure 5 presents the results from estimating the event study model (2) for the actual mortality rate. We find no statistically meaningful relationship between the death event and patient mortality. The point estimates both before and after the death event are flat, close to zero, and statistically insignificant. While a few point estimates are positive, the overall pattern of dynamic effects does not indicate a systematic decline in care quality. Column 2 of Table 6 confirms the visual evidence. We estimate short- and long-run coefficients of 0.009, both of which are statistically indistinguishable from zero.

A possible concern with these results is spillovers. Reallocation of staff across disease areas in response to the event could negatively affect the quality of care in less affected areas. If such spillovers are present, they alone could explain the lack of mortality effects. While the inclusion of disease-area-by-time fixed effects already minimizes the relevance of cross-disease-areas comparisons, we address this concern further by reruning the basic specification using our hospital-level measure of exposure to the deceased. Appendix Table B.7 shows that the effects remain small and statistically insignificant.

Another possible explanation for the null effect on mortality is spillover effects from treated to untreated hospitals. The death event may represent a shock to the labor supply of not only the hospital hit by the event but also of other hospitals within the same market. Treated hospitals could poach physicians from nearby control hospitals, which could affect their own operations and patient outcomes. As a robustness check, we repeat the matching algorithm but with the restriction that the control hospitals must be located *outside* the state of the treated hospital. We then rerun the baseline specification using this alternative comparison group. Appendix Figure B.7 documents that the results remain extremely similar to the baseline. Cross-hospital spillovers are unlikely to account for the muted effects on mortality.

Other Measures of Care Quality. The consequences of suboptimal care might not lead to immediate fatalities but can result in complications or deterioration that emerge only after discharge. To examine this possibility, Panel A of Figure 6 displays the results from estimating the dynamic treatment effects on 30-day mortality in the at-risk subsample. There is no evidence of a statistically significant increase in 30-day mortality, either in the immediate post-event periods or over longer horizons.

The data suggest no significant changes in mortality, but the death event could still affect less severe outcomes. We examine two complementary outcomes. The first is 30-day readmission. While some readmissions are necessary regardless of the care quality delivered, the medical literature has established that they are particularly responsive to the adoption of best clinical practices, such as adherence to clinical guidelines (Pennsylvania Health Care Cost Containment Council, 1996; Hannan et al., 2003), and pre-discharge counseling (Naylor et al., 1999).[22] The second measure

[22]Unlike the United States, there are no explicit financial incentives to reduce readmissions in Brazil. This alleviates concerns about selection and coding manipulation, which have been shown to be important in the context of Medicare

we examine is an indicator variable for the use of high-technology, highly complex procedures, which are believed to be beneficial and lead to better outcomes (Farquhar, 2008). Examples include coronary artery bypass graft, abdominal aortic aneurysm repair, and carotid endarterectomy. In our data, the use of these procedures is associated with a 37 percent decline in mortality after controlling for patient characteristics as well as hospital and time fixed effects.

Panels B-D of Figure 6 present the dynamic of treatment effects for these outcomes. While these outcomes apply to a wide range of conditions beyond life-threatening situations, we focus on the subsample of patients at risk of mortality to reduce concerns about compositional changes in the patient case mix. When looking at the indicator for beneficial procedures, we also present results from a subsample that further restricts the data to patients who are candidates for such procedures.[23] In no case do we find statistically significant changes that suggest a deterioration in care quality. The trends in these outcomes are essentially flat both before and after the event.

6.3 External Validity

Because physicians who exit due to death often differ in terms of seniority and other characteristics from their peers, an important question is how generalizable the effects on hospital performance are to the average physician. To address such external validity concerns, we implement the approach developed by Andrews and Oster (2019). They propose a procedure that adjusts the point estimates by accounting for heterogeneity in treatment effects across observables influencing participation in the design and provides bounds under different assumptions on the relative importance of unobservables. Appendix Table B.8 performs this procedure considering the characteristics of deceased physicians displayed in Table 1.[24] The conclusions remain essentially the same even under relatively extreme scenarios where unobservables are substantially more important than observables in predicting the treatment effects.[25]

7 Understanding the Quality-Quantity Tradeoff

Our reading of the evidence is that despite their organizational efforts, hospitals experience a decline in productivity. Hospitals reduce their overall output in scale while maintaining the same patient outcomes. There are two major classes of mechanisms that may explain the nature of this finding. The first is a quality prioritization channel. The organizational responses of hospitals could directly

patients (Gupta, 2021).

[23] Many conditions do not require the use of these procedures. For example, uncomplicated pneumonia can often be treated with antibiotics, dehydration is typically managed with intravenous fluids, and minor fractures are commonly addressed with simple casting techniques. We therefore restrict the data to patients whose conditions involve more frequent use of high-technology, high-complexity procedures. Specifically, we define such conditions as those with a utilization rate above the 75 percentile of the utilization distribution.

[24] Because a physician's experience and age are almost perfectly correlated, we exclude the former for this exercise.

[25] For example, selection into the sample on unobservables would need to be 8 times as large as that on observables to render to yield a long-run treatment effect of zero. Because the observables include a physician's age, a key characteristic that explains almost all of the differences between deceased and alive physicians, this approximation suggests that our findings are fairly robust.

target prioritizing care quality over quantity. To be more precise, suppose that providers admit patients up to the point where the expected quality of care they can deliver per patient approaches a minimum care standard threshold. The decline in admissions we observe might reflect, for example, that hospitals are not willing to reduce their minimum standard of care by assuming risks with new physicians in more exposed, strained areas. The second channel is doctor quality. Recent work suggests that physicians differ widely in their individual-specific productivity, with significant overlap between physicians with different levels of experience and training (Gowrisankaran et al., 2023; Chan and Chen, 2022). While deceased physicians are generally more experienced, they may still differ widely in their individual-specific capital, with both low- and high-performing types, such that the average physician is not different from their alive peers. If this is the case, it could be less challenging for hospitals to replace the skills of deceased physicians and mitigate any consequences on care quality.

In this section, we provide evidence that both dimensions matter and are likely to interact with each other. Before turning to this empirical evidence, we first introduce a simple conceptual to fix ideas and guide our analysis. While the evidence suggests that hospitals undertake a variety of organizational changes, all of which may be equally important, this section focuses on the allocation of incumbent and new physicians to tasks performed previously by the deceased.

7.1 Conceptual Framework

Let us introduce a simple conceptual framework to better understand the potential mechanisms discussed above. There are N physicians affiliated with hospital h in period t, each with an individual output y_{iht} and a patient share α_{it}. Physicians are indexed by the subscript i, with the deceased represented by $i = d$ and new entrants by $i = e$. New physicians enter the hospital only when an incumbent exits the hospital. The output of a physician depends on their specific skills, β_i, and a hospital-specific capital component, λ_{iht}. For simplicity, we assume that the level of hospital-specific capital increases with tenure and is identical for all physicians in the subsequent period after their entry, such that $\lambda_{iht} = \lambda_h^I$ for physicians with more than one period of tenure. We also assume that the level of the hospital-specific capital in the period of entry is the same for all new entrants, but varies inversely with the service gap left by the deceased physician as measured by α_{dt-1}. That is, $\partial \lambda_h^e(\alpha_{dt-1})/\alpha_{dt-1} < 0$. This captures the idea that new entrants face greater challenges adapting in more affected areas.

The aggregate output Y_{ht} in an area of care is the weighted average of individual physicians' output, where the weights correspond to the share of patients treated by each physician. The change in aggregate output after the death of a physician in period $t-1$ can be written as follows:

$$\Delta Y_{ht} = \underbrace{\left[\sum \alpha_{jt}\beta_j + \alpha_{et}\beta_e\right] - \left[\alpha_{dt-1}\beta_d + \sum \alpha_{jt-1}\beta_j\right]}_{\text{heterogeneity in physician quality}} + \underbrace{\alpha_e\left[\lambda_h^e(\alpha_{dt-1}) - \lambda_h^I\right]}_{\text{hospital-specific capital}} \tag{6}$$

The first term shows how the exit of a physician affects the average physician quality. The higher the deceased physician's quality, the larger the decline in aggregate output, holding the rest constant.

The second term captures the loss of firm-specific capital when the physician with longer tenure leaves and is replaced by new entrants. The decline in aggregate output will be larger when the caseload left by the deceased is greater. Notably, this effect is proportional to the patient share allocated to new entrants. Because the firm-specific capital component is identical for all physicians in the subsequent period, what matters for output is how much of the tasks performed by the deceased are taken over by new physicians immediately following their entry.

Equation (6) formalizes the intuition potentially behind the limited effects on care quality we observe: the deceased physicians are not very different from the average physician, and hospitals minimize their reliance on new entrants in more affected areas. We next evaluate the empirical relevance of these channels.

7.2 Physician Quality

We proxy for the quality of a physician in a disease area with the inpatient mortality rate in that area, as in Chan and Chen (2022). This is an admittedly imperfect proxy. Differences in mortality rates across physicians may reflect not only differences in a physician's skills but also differences in patient case mix. To diminish this concern, we use the risk-adjusted mortality in the subsample of patients at risk of mortality. We refer to this measure simply as mortality prediction or related terms. To understand how far the deceased is from the average physician, we normalize this variable by subtracting its area-specific mean. On average, patients treated by deceased physicians have a mortality rate that is 2.4 percentage points lower relative to the area-specific mean, suggesting that deceased physicians are only slightly more effective than their peers. The fact that this average gap is not sizeable is consistent with the lack of mortality effects documented in the previous subsection.

The relatively small gap in mortality predictions hides considerable heterogeneity across deceased physicians (see Appendix Figure A.3). The standard deviation in the normalized mortality predictions is about 17 percentage points. While some physicians on the right side of the distribution perform substantially worse than their peers, some others on the left do exceptionally well. We can exploit this rich heterogeneity to gain further insights into the physician-specific capital channel. If physicians have causal effects, then patient outcomes should shift in the opposite direction of the deceased physician's mortality prediction. When a physician with exceptional skills departs, it could be particularly challenging to find a comparable replacement, and, as a result, there could be an increase in mortality. One would expect the exact opposite when a relatively low-performing physician departs.

To test this prediction, we split the treated sample into two groups based on whether the deceased physician's normalized mortality expectation is below or above the median. We then estimate the dynamic treatment effects separately for each group, using the full set of control hospitals in both subsamples to maximize precision. The mean normalized mortality rate among deceased physicians below and above the median is -12.3 and 7.7 percentage points, respectively. Figure 7 shows that there is strong heterogeneity in the treatment effects. The point estimates are highly significant and move in opposite directions depending on the mortality prediction of the deceased.

When the deceased physician's mortality expectation is relatively high, their departure leads to an immediate and persistent decline in patient mortality. Conversely, if the deceased physician had a relatively low mortality expectation, we observe a highly significant increase in mortality rates following their death. Importantly, these results are not driven by changes in the sorting of patients into physicians. Appendix Figure B.8 documents that the observed heterogeneity completely disappears when we use predicted mortality as the dependent variable.

The pattern of mortality effects in Figure 7 appears somewhat asymmetric, especially toward the end of the post-event period. To account for differences in expected mortality changes and make the estimates as comparable as possible, we adjust the low-mortality coefficients by scaling them with the ratio of the average mortality prediction in the high-mortality subsample to that in the low-mortality subsample (lighter blue line in Figure 7). The differences become even larger, and we can reject the hypothesis of equality at the 5 percent level of significance.[26] The treatment effects in the high-mortality prediction subsample are 2 to 5 times larger in magnitude than those in the low-mortality prediction subsample. "Good" physicians have a much smaller impact on mortality than "bad" physicians.

This asymmetry aligns with the narrative that hospital managers implement organizational changes that preserve the quality of care. When a good physician departs, hospitals may respond more aggressively by reallocating resources, implementing contingency plans, or intensifying collaborative practices to maintain care quality. In contrast, the exit of a bad physician might not trigger a similar response, as their absence alone already serves as a corrective mechanism. Below, we will provide further evidence in line with such a strategic response.

7.3 Managing the Loss of Hospital-Specific Capital

Our prior results in Section 5.2 reveal that hospitals rely heavily on incumbents to maintain continuity of care in more affected areas when facing an unexpected physician exit. If new entrants perform worse in highly exposed areas because adapting to the local hospital environment is more challenging, this allocation strategy may help explain the lack of mortality effects. In this section, we investigate whether new physicians are indeed associated with a worse performance in more exposed areas.

To explore this question, one would ideally like to estimate a direct empirical analog of the hospital-specific component of equation (6) but it is unobservable to the econometrician. However, if we assume that the physician-specific capital does not vary systematically with the exposure to the deceased, we could obtain approximate estimates of the parameter of interest using the following first-difference model:

$$y^{e}_{ihs,k} - y_{ihs,-1} = \beta_k \text{Exposure}_{ihs} + \lambda_{is} + \eta_{ihs} \tag{7}$$

In this specification, $y^{e}_{ihs,k}$ represents the mortality rate in the post-event period k only from patients

[26] To test for the differences, we pool the subsamples together into a single file and estimate a model where all variables are fully interacted with subsample fixed effects. We then rescale the coefficients in the low-mortality prediction subsample, and test the joint hypothesis that $\mid \beta^{low}_{\tau} \mid = \mid \beta^{high}_{\tau} \mid$ for all $\tau = 0, 1, 2, ..., 10$. We find a p-value of 0.0108.

attended by physicians starting their operations at the hospital after the event, whereas $y_{ihs,-1}$ is the overall mortality rate from all patients in the pre-event period k. Put simply, the difference $y^e_{ihs,k} - y_{ihs,-1}$ captures how much the new entrants' mortality outcomes deviate from the pre-event overall mortality rate. Positive estimates of β_k would imply that new entrants have a harder adaptation in more exposed areas under the aforementioned assumption of no correlation between the event and physician-specific capital. This is a strong assumption, one that cannot be directly tested in the data. However, the hospital-specific and physician-specific capital channels generate specific predictions on the dynamic of treatment effects. The former implies treatment effects that decline over time, as new physicians familiarize with the hospital, while the latter implies persistence in the treatment effects given that it is more likely to capture permanent features of physicians. Thus, we can tease out the relative importance of both mechanisms by inspecting the dynamics of the treatment effects.

The specification in (7) is useful for developing intuition, and suggests estimating a first pre-post difference regression. In practice, we can simply replace the overall mortality rate in the post-event period for the new entrants' mortality rate and then estimate the specification (3). We do this in Figure 8. Because many deceased physicians attended patients in the quarter of the event, we calculate the new entrants' patient mortality rate since the first quarter after the event. As one can infer from the figure, new entrants are significantly associated with an increase in mortality only in the short run, consistent with the idea that physicians develop some degree of firm-specific capital. The magnitude of this short-run effect is large. Full exposure to the event would imply an increase of about 4 percentage points in the mortality rate, or about 33 percent relative to the pre-event mean.[27]

In stark contrast to new entrants, the estimated coefficient for incumbents are very close to zero and statistically insignificant. The magnitude of the short-run treatment effect is 10 times as small as the corresponding one for new entrants. From a causal analysis perspective, we view this null effect as evidence that the association between new entrants and mortality is unlikely to be driven by omitted factors that affect both groups similarly. In particular, if our estimates were driven by deceased physicians performing exceptionally well in areas where they managed a larger share of cases, we would have observed a jump in mortality immediately following the event both for new entrants and incumbents.

A possible concern with these estimates is that they could be biased in the presence of changes in the of composition of patients. If new physicians attend to patients with higher mortality risk in more exposed areas, we could observe an increase in mortality even if they perform as well as those in less exposed areas. However, note that for this to explain our results, the selection pattern would need to vary over time in a way that is consistent with the decline in the treatment effects we observe. It is difficult to think of scenarios where this is plausible. To examine this possibility empirically, we repeat the analysis but use the predicted mortality as the dependent variable. Appendix Figure B.9 shows that the effects of the death event on predicted mortality for new entrants can be bounded

[27] Appendix Table B.9 shows that the results remain almost unchanged when we rerun the model focusing solely on those who began attending cases in the quarter of the death event or the subsequent quarter. This is consistent with the fact that the bulk of new entries occur in the periods immediately following the death event.

to a tight interval around zero. This suggests that our estimates do not reflect differences in patient composition between new entrants and baseline physicians.

Counterfactual Estimates. An important question is what the overall mortality effect would have been if new entrants attended a larger share of cases. To carry out this counterfactual exercise, we rely on the observation that the overall treatment effect is approximately equal to the weighted average of the treatment effects for the subsamples of new entrants and incumbents, with the weights corresponding to the proportion of patients in each subsample. Specifically, we combine the estimation subsamples into a single file and rerun the baseline specification but reweighting the importance of each subsample based on alternative patient shares attributed to new entrants. Figure 9 shows that the overall short-run effect would have been more than twice as large if new entrants had attended 60 percent of all cases rather than 13 percent. Fully exposed areas would have experienced a short-run increase in patient mortality of about 2 percentage points, or 17 percent of the pre-event mean. A less extreme scenario, where new entrants would have a patient share of 30 percent, implies a short-run increase in mortality of about 1.3 percentage points or 10 percent.

Overall, the evidence is consistent with the interpretation that physicians develop a degree of hospital-specific capital, which is effectively lost upon their departure and difficult to rebuild in areas where the deceased was central. These findings are consistent with the quality prioritization channel. If the focus of hospitals were solely on maximizing volume, new entrants would be required to assume the exact workload of the deceased physician, irrespective of whether their adaption is more challenging in highly exposed, strained areas. In practice, however, new entrants take on only a limited fraction of the workload and this is why we do not detect meaningful changes in overall mortality rates.

7.4 Expected Mortality Change and Allocation of New Physicians

The evidence suggests that the way how hospitals organize in response to an unexpected physician exit matters for patient outcomes. In this section, we further explore whether hospital managers make different allocation decisions depending on the anticipated impact of the event on mortality. If hospitals organize to preserve care quality, they could adopt more strategic responses when a shock is expected to be particularly disruptive. Given that new entrants are associated with increased mortality, managers may be more cautious in their allocation decisions when a high-performing physician suddenly leaves.

To explore this hypothesis, we use the deceased' normalized mortality prediction discussed in Section 7.2 and characterize high and low mortality predictions based on a median split. We then estimate the dynamic effects of the death event on the patient share of new entrants separately for each group. Figure 10 presents the results alongside the baseline estimates for ease of comparison. We find strong evidence of heterogeneity. In the immediate aftermath of the event, the caseload assigned to new entrants increases by around 8 percentage points when the deceased physician has a low mortality prediction or, equivalently, when the patient mortality is expected to increase

significantly. In contrast, this effect increases to more than 20 percentage points in the high mortality subsample, or equivalently, when a low-performing physician exits. This gap narrows in subsequent quarters as the overall effects converge toward zero, but the estimated effects remain consistently lower in the low mortality subsample.

Overall, these findings suggest that managers are sensitive to the potential extent of disruption caused by physician turnover and adjust their allocation decisions strategically to preserve care quality.

8 Concluding Remarks

Firms regularly face the challenge of losing valuable labor, prompting managers not only to find a suitable replacement but also to reorganize production processes to maintain continuity. However, we still know little about the specific ways organizations carry out these adaptations and how they ultimately affect output. This paper addresses this gap by examining how the performance and internal organization of hospitals respond to physician turnover caused by unexpected deaths. Using a novel administrative dataset from the Brazilian healthcare sector and a difference-in-differences design, we compare areas of care with varying reliance on the departing physician across hospitals over time.

Our findings reveal that hospitals respond to the sudden physician exit by reallocating tasks and roles to existing staff familiar with hospital operations, broadening their clinical and administrative duties. This strategy effectively avoids shortages in physician availability in the most affected areas without compromising the match between physician skills and patient needs. At the same time, hospitals deliberately limit the caseloads of newcomers in these strained areas—particularly when the exit poses a greater threat to care quality. This allocation pattern seems strategic, as our data indicates that newcomers struggle to maintain care quality standards in more exposed areas. In terms of output, hospital admissions show a persistent decline, reducing bed utilization. Crucially, hospitals scale back on lower-risk cases while maintaining admissions for higher-risk patients, and the intensity of care for those admitted remains unchanged. Ultimately, we find no evidence of deterioration across widely used metrics of care quality. These findings suggest that hospitals prioritize maintaining care quality following physician exits and effectively achieve this through organizational changes, even if such measures entail reducing the overall scale of operations.

Our results contribute to a broader understanding of how organizations manage sudden disruptions in specialized labor. We demonstrate that internal reorganization helps firms mitigate the adverse effects of labor turnover. This finding is particularly relevant given the high frequency of turnover shocks and their global prevalence, spanning developed and emerging economies. More generally, our results also offer lessons on how firms adopt practical organizational strategies. For instance, we learn about how hospitals reallocate tasks and duties following physician turnover and the relevance of such physician-to-task assignments for hospital quality. Hence, our findings underscore the importance of strategic task allocation as a managerial tool to sustain performance

amid workforce disruptions. Moreover, our results provide insights into how firms prioritize certain aspects of performance, such as quality over scale, and tailor their organizational strategies accordingly. Whether these strategies extend beyond the healthcare sector remains an open question. Health providers often balance patient outcomes with financial objectives, which may lead hospitals in our study to prioritize care quality and adopt targeted organizational responses. This dynamic may not apply in purely profit-driven firms. Future research examining other sectors could clarify whether non-healthcare firms adopt similar strategies when key personnel depart, shedding light on the broader applicability of the organizational responses we identify in this paper.

Bibliography

Abaluck, Jason, Leila Agha, Chris Kabrhel, Ali Raja, and Arjun Venkatesh, "The determinants of productivity in medical testing: Intensity and allocation of care," *American Economic Review*, 2016, *106* (12), 3730–3764.

Abowd, John M and Francis Kramarz, "The costs of hiring and separations," *Labour Economics*, 2003, *10* (5), 499–530.

Adhvaryu, Achyuta, Namrata Kala, and Anant Nyshadham, "Management and shocks to worker productivity," *Journal of Political Economy*, 2022, *130* (1), 1–47.

_ , Teresa Molina, and Anant Nyshadham, "Expectations, Wage Hikes and Worker Voice," *The Economic Journal*, 2022, *132* (645), 1978–1993.

Aghamolla, Cyrus, Pinar Karaca-Mandic, Xuelin Li, and Richard T Thakor, "Merchants of death: The effect of credit supply shocks on hospital outcomes," *American Economic Review*, 2024, *114* (11), 3623–3668.

Alan, Manning, "Imperfect competition in the labor market," in "Handbook of labor economics," Vol. 4, Elsevier, 2011, pp. 973–1041.

Albouy, David, "The unequal geographic burden of federal taxation," *Journal of Political Economy*, 2009, *117* (4), 635–667.

Andrews, Isaiah and Emily Oster, "A simple approximation for evaluating external validity bias," *Economics Letters*, 2019, *178*, 58–62.

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta, "Cross-country differences in productivity: The role of allocation and selection," *American Economic Review*, 2013, *103* (1), 305–334.

Bates, David W, David M Levine, Hojjat Salmasian, Ania Syrowatka, David M Shahian, Stuart Lipsitz, Jonathan P Zebrowski, Laura C Myers, Merranda S Logan, Christopher G Roy et al., "The safety of inpatient health care," *New England Journal of Medicine*, 2023, *388* (2), 142–153.

Batty, Michael and Benedic Ippolito, "Financial incentives, hospital care, and health outcomes: evidence from fair pricing laws," *American Economic Journal: Economic Policy*, 2017, *9* (2), 28–56.

Beccaria, L and R Maurizio, "Labor Turnover in Latin America how Intensive is it and How Differently Does it Behave Across Countries?," *International Labour Review*, 2018.

Becker, Gary S, "Investment in human capital: A theoretical analysis," *Journal of political economy*, 1962, *70* (5, Part 2), 9–49.

Bertrand, Marianne and Antoinette Schoar, "Managing with style: The effect of managers on firm policies," *The Quarterly journal of economics*, 2003, *118* (4), 1169–1208.

Bianchi, Nicola, Giulia Bovini, Jin Li, Matteo Paradisi, and Michael Powell, "Career spillovers in internal labour markets," *The Review of Economic Studies*, 2023, *90* (4), 1800–1831.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts, "Does management matter? Evidence from India," *The Quarterly journal of economics*, 2013, *128* (1), 1–51.

_ , Erik Brynjolfsson, Lucia Foster, Ron Jarmin, Megha Patnaik, Itay Saporta-Eksten, and John Van Reenen, "What drives differences in management practices?," *American Economic Review*, 2019, *109* (5), 1648–1683.

Branco, Danyelle, Bladimir Carrillo, Di Fang, and Wilman Iglesias, "Expertise Overlap and Team Productivity: Evidence from the Hospital Industry," Technical Report 2024.

Brennan, Troyen A, Lucian L Leape, Nan M Laird, Liesi Hebert, A Russell Localio, Ann G Lawthers, Joseph P Newhouse, Paul C Weiler, and Howard H Hiatt, "Incidence of adverse events and negligence in hospitalized patients: results of the Harvard Medical Practice Study I," *New England journal of medicine*, 1991, *324* (6), 370–376.

Buchinsky, Moshe, Denis Fougere, Francis Kramarz, and Rusty Tchernis, "Interfirm mobility, wages and the returns to seniority and experience in the United States," *The Review of economic*

studies, 2010, *77* (3), 972–1001.

Butters, R Andrew, Daniel W Sacks, and Boyoung Seo, "How do national firms respond to local cost shocks?," *American Economic Review*, 2022, *112* (5), 1737–1772.

Callaway, Brantly and Pedro HC Sant'Anna, "Difference-in-differences with multiple time periods," *Journal of econometrics*, 2021, *225* (2), 200–230.

Card, David and Alan Krueger, "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania," *American Economic Review*, 1994, *84* (4), 772–93.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer, "The effect of minimum wages on low-wage jobs," *The Quarterly Journal of Economics*, 2019, *134* (3), 1405–1454.

Chaisemartin, Clément De and Xavier d'Haultfoeuille, "Two-way fixed effects estimators with heterogeneous treatment effects," *American economic review*, 2020, *110* (9), 2964–2996.

Chan, David C, "Teamwork and moral hazard: evidence from the emergency department," *Journal of Political Economy*, 2016, *124* (3), 734–770.

_ **and Yiqun Chen**, "The productivity of professions: evidence from the emergency department," Technical Report, National bureau of economic research 2022.

Chandra, Amitabh and Douglas O Staiger, "Identifying sources of inefficiency in healthcare," *The quarterly journal of economics*, 2020, *135* (2), 785–843.

_ **, David Cutler, and Zirui Song**, "Who ordered that? The economics of treatment choices in medical care," *Handbook of health economics*, 2011, *2*, 397–432.

Chen, Jiafeng and Jonathan Roth, "Logs with zeros? Some problems and solutions," *The Quarterly Journal of Economics*, 2024, *139* (2), 891–936.

Chen, Yiqun, "Team-specific human capital and team performance: Evidence from doctors," *American Economic Review*, 2021, *111* (12), 3923–62.

Currie, Janet and Jonathan Zhang, "Doing more with less: Predicting primary care provider effectiveness," *Review of Economics and Statistics*, 2023, pp. 1–45.

Davis, Steven J and Till M Von Wachter, "Recessions and the cost of job loss," Technical Report, National Bureau of Economic Research 2011.

Doeringer, Peter B. and Michael J. Piore, *Internal labor markets and manpower analysis*, Routledge, 1971.

Donald, Stephen G and Kevin Lang, "Inference with difference-in-differences and other panel data," *The review of Economics and Statistics*, 2007, *89* (2), 221–233.

Doyle, Joseph J and Becky Staiger, "Physician Group Influences on Treatment Intensity and Health: Evidence from Physician Switchers," Technical Report, National Bureau of Economic Research 2022.

_ **, Steven M Ewer, and Todd H Wagner**, "Returns to physician human capital: Evidence from patients randomized to physician teams," *Journal of health economics*, 2010, *29* (6), 866–882.

Farquhar, Marybeth, "AHRQ quality indicators," *Patient safety and quality: an evidence-based handbook for nurses*, 2008.

Fenizia, Alessandra, "Managers and productivity in the public sector," *Econometrica*, 2022, *90* (3), 1063–1084.

Flaaen, Aaron, Matthew D Shapiro, and Isaac Sorkin, "Reconsidering the consequences of worker displacements: Firm versus worker perspective," *American Economic Journal: Macroeconomics*, 2019, *11* (2), 193–227.

Gibbons, Robert and Rebecca Henderson, *What do managers do?: Exploring persistent performance differences among seemingly similar enterprises*, Citeseer, 2012.

Ginja, Rita, Arizo Karimi, and Pengpeng Xiao, "Employer responses to family leave programs," *American Economic Journal: Applied Economics*, 2023, *15* (1), 107–135.

_ **, Julie Riise, Barton Willage, and Alexander Willen**, "Does Your Doctor Matter?," *Journal of*

Political Economy Microeconomics, 2024, *Forthcoming.*

Goodman-Bacon, Andrew, "Difference-in-differences with variation in treatment timing," *Journal of econometrics*, 2021, *225* (2), 254–277.

Gosnell, Greer K, John A List, and Robert D Metcalfe, "The impact of management practices on employee productivity: A field experiment with airline captains," *Journal of Political Economy*, 2020, *128* (4), 1195–1233.

Gowrisankaran, Gautam, Keith Joiner, and Pierre Thomas Léger, "Physician practice style and healthcare costs: Evidence from emergency departments," *Management Science*, 2023, *69* (6), 3202–3219.

Gupta, Atul, "Impacts of performance pay for hospitals: The readmissions reduction program," *American Economic Review*, 2021, *111* (4), 1241–1283.

Hamermesh, Daniel S and Gerard A Pfann, "Adjustment costs in factor demand," *Journal of Economic literature*, 1996, *34* (3), 1264–1292.

Hannan, Edward L, Michael J Racz, Gary Walford, Thomas J Ryan, O Wayne Isom, Edward Bennett, and Robert H Jones, "Predictors of readmission for complications of coronary artery bypass graft surgery," *Jama*, 2003, *290* (6), 773–780.

Hoffman, Mitchell, "The value of hiring through employee referrals in developed countries," *IZA World of Labor*, 2017.

_ **and Christopher T Stanton**, "People, Practices, and Productivity: A Review of New Advances in Personnel Economics," 2024.

_ **and Steven Tadelis**, "People management skills, employee attrition, and manager rewards: An empirical analysis," *Journal of political economy*, 2021, *129* (1), 243–285.

Holmstrom, Bengt R. and Jean Tirole, "The Theory of the Firm," in "Handbook of Industrial Organization," Vol. 1, Elsevier, 1989, pp. 61–133.

Huckman, Robert S and Gary P Pisano, "The firm specificity of individual performance: Evidence from cardiac surgery," *Management science*, 2006, *52* (4), 473–488.

Huebener, Mathias, Jonas Jessen, Daniel Kuehnle, and Michael Oberfichtner, "Parental Leave, Worker Substitutability, and Firms' Employment," *The Economic Journal*, 2024, *Forthcoming.*

Ichniowski, Casey and Kathryn Shaw, "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines.," *American Economic Review*, 1997, *87* (3).

Jacobson, Louis S, Robert J LaLonde, and Daniel G Sullivan, "Earnings losses of displaced workers," *The American economic review*, 1993, pp. 685–709.

Jäger, Simon, Jörg Heining, and Nathan Lazarus, "How substitutable are workers? evidence from worker deaths," Technical Report, National Bureau of Economic Research 2022.

Jolivet, Grégory, Fabien Postel-Vinay, and Jean-Marc Robin, "The empirical content of the job search model: Labor mobility and wage distributions in Europe and the US," *European Economic Review*, 2006, *50* (4), 877–907.

Kuhn, Peter and Lizi Yu, "How costly is turnover? Evidence from retail," *Journal of Labor Economics*, 2021, *39* (2), 461–496.

Lazear, Edward P, "Firm-specific human capital: A skill-weights approach," *Journal of political economy*, 2009, *117* (5), 914–940.

_ **and Kristin McCue**, "What causes labor turnover to vary?," Technical Report, National Bureau of Economic Research 2018.

Minni, Virginia Magda Luisa, "Making the invisible hand visible: Managers and the allocation of workers to jobs," 2024.

Naylor, Mary D, Dorothy Brooten, Roberta Campbell, Barbara S Jacobsen, Mathy D Mezey, Mark V Pauly, and J Sanford Schwartz, "Comprehensive discharge planning and home follow-up of hospitalized elders: a randomized clinical trial," *Jama*, 1999, *281* (7), 613–620.

Otero, Cristobal and Pablo Munoz, "Managers and public hospital performance," *University of California, Berkeley*, 2024.

Paim, Jairnilson, Claudia Travassos, Celia Almeida, Ligia Bahia, and James Macinko, "The Brazilian health system: history, advances, and challenges," *The Lancet*, 2011, *377* (9779), 1778–1797.

Pennsylvania Health Care Cost Containment Council, *Pennsylvania's Guide to Coronary Artery Bypass Graft Surgery, 1994-1995*, Pennsylvania: Pennsylvania Health Care Cost Containment Council, 1996.

Rossin-Slater, Maya and Miriam Wüst, "What is the added value of preschool for poor children? Long-term and intergenerational impacts and interactions with an infant health intervention," *American Economic Journal: Applied Economics*, 2020, *12* (3), 255–286.

Saez, Emmanuel, Benjamin Schoefer, and David G Seim, "Deadwood Labor: The Effects of Eliminating Employment Protection," Technical Report, National Bureau of Economic Research 2023.

Sauvagnat, Julien and Fabiano Schivardi, "Are executives in short supply? Evidence from death events," *Review of Economic Studies*, 2024, *91* (1), 519–559.

Scheffer, M. et al., *Demografia Médica no Brasil 2020*, São Paulo, SP: FMUSP, CFM, 2020.

Scheffer, Mário, Alexandre Guerra, Alex Cassenote, Alicia Matijasevich, Ana Pérola Drulla Brandão, Euclides Ayres de Castilho, Giuliano Russo, Jessica Pronestino de Lima Moreira, Ligia Bahia, Lucas Salvador Andrietta, Mario Roberto Dal Poz, Paulo Roberto de Castro Villela, Renata Alonso Miotto, Virginia Costa Duarte, and Walesca Araújo, *Demografia Médica no Brasil 2023*, São Paulo, SP: FMUSP, AMB, 2023.

Schmutte, Ian M and Meghan M Skira, "The response of firms to maternity leave and sickness absence," *Journal of Human Resources*, 2023.

Shapiro, Jesse M, "Smart cities: quality of life, productivity, and the growth effects of human capital," *The review of economics and statistics*, 2006, *88* (2), 324–335.

Shimer, Robert, "Reassessing the ins and outs of unemployment," *Review of Economic Dynamics*, 2012, *15* (2), 127–148.

Silver, David, "Haste or Waste? Peer pressure and productivity in the emergency department," *The Review of Economic Studies*, 2021, *88* (3), 1385–1417.

Skinner, Jonathan, "Causes and consequences of regional variations in health care," in "Handbook of health economics," Vol. 2, Elsevier, 2011, pp. 45–93.

Sun, Liyang and Sarah Abraham, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of econometrics*, 2021, *225* (2), 175–199.

Syverson, Chad, "What determines productivity?," *Journal of Economic Literature*, 2011, *49* (2), 326–365.

Topel, Robert, "Specific capital, mobility, and wages: Wages rise with job seniority," *Journal of political Economy*, 1991, *99* (1), 145–176.

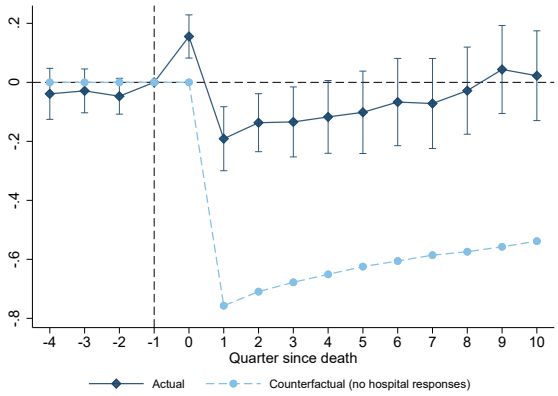
Weissman, Joel S, Jeffrey M Rothschild, Eran Bendavid, Peter Sprivulis, E Francis Cook, R Scott Evans, Yevgenia Kaganova, Melissa Bender, JoAnn David-Kasdan, Peter Haug et al., "Hospital workload and adverse events," *Medical care*, 2007, *45* (5), 448–455.

Wing, Coady, Seth M Freedman, and Alex Hollingsworth, "Stacked difference-in-differences," Technical Report, National Bureau of Economic Research 2024.

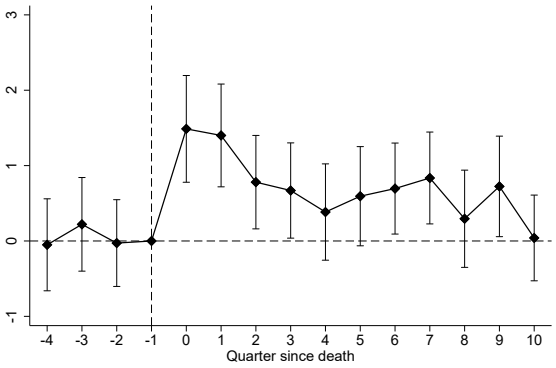
Young, Alwyn, "Improved, nearly exact, statistical inference with robust and clustered covariance matrices using effective degrees of freedom corrections," *Manuscript, London School of Economics*, 2016.

Figure 1: Effects on Employment and Hiring

Panel A. Log (number of affiliated physicians)



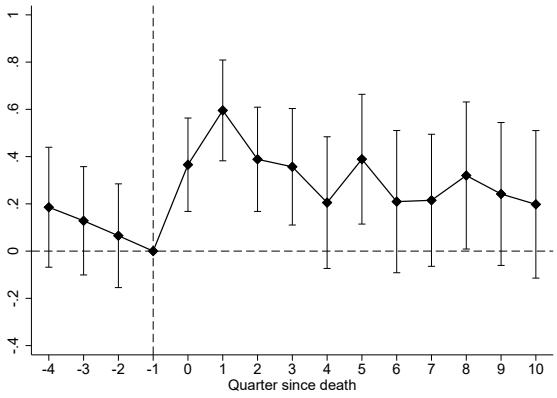
Panel B. Log (number of new affiliated physicians)



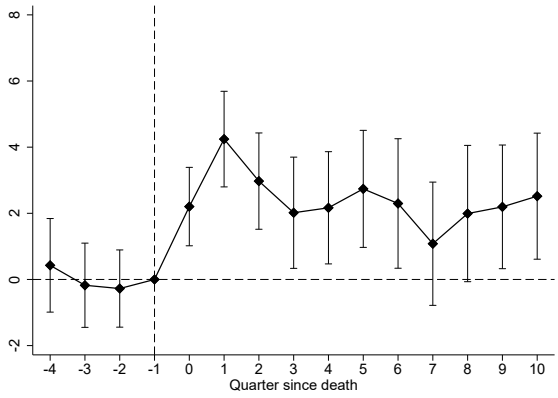
Notes. This figure presents estimates based on the model (4) with respective 95 percent confidence intervals. Count of total and new physicians are expressed in full-time equivalent (FTE) physicians assuming a full-time workweek of 40 hours. Panel A considers log total number of affiliated physicians in a hospital. The lighter blue line is counterfactual evolution of staffing levels without any hiring or retention response. This counterfactual is computed by assuming that the staffing level in the post-event period would have the same as the one observed in the quarter before the event but without the deceased physician. This calculation also adjusts for the probability of exit if the physician had survived, which reduces the quantitative importance of the shock over time. Dependent variable in panel B is the log number of new affiliated physicians. To account for the fact that many hospital-quarters have no hiring activity, the extensive margin is calibrated following Chen and Roth (2024): we assign a value of 100 log points to an extensive margin change between 0 and the minimum non-zero value of the dependent variable. Standard errors are clustered at the hospital level.

Figure 2: Effects on Incumbent Workload

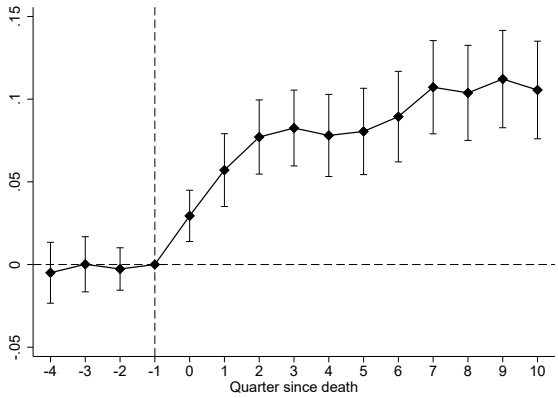
Panel A. Log (number of patients)



Panel B. Number of distinct diagnoses treated

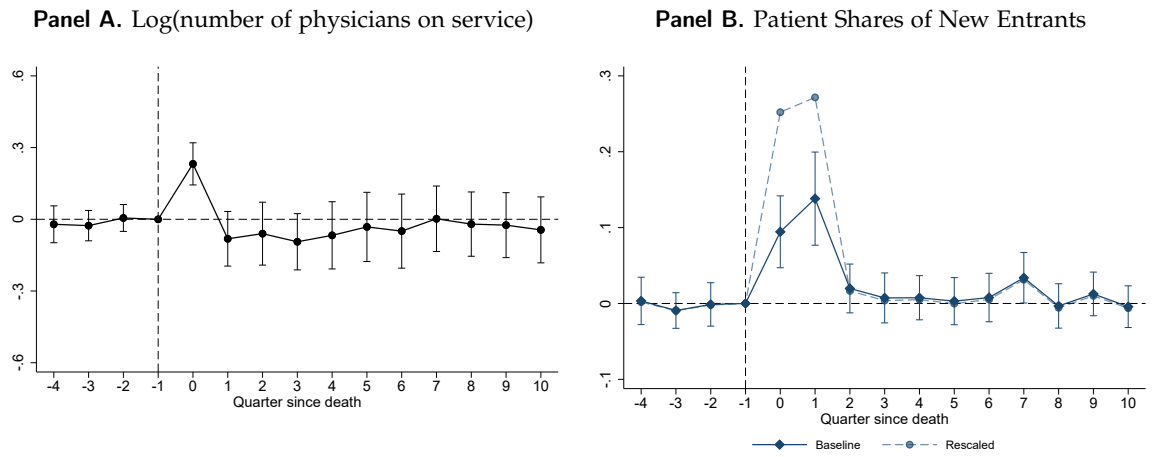


Panel C. Indicator for nonclinical activities



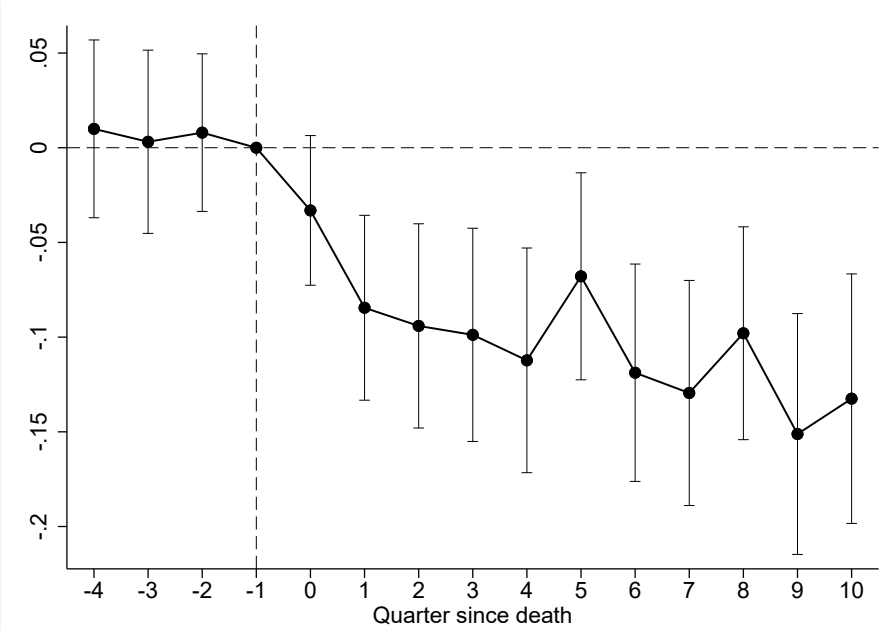
Notes. This figure presents estimates based on the model (5) with respective 95 percent confidence intervals. Incumbent physicians correspond to those who were affiliated with the hospital before the death event. Panel A considers changes in the number of patients managed, panel B looks at the number of distinct diagnoses, and panel C examines an indicator for whether the physician has nonclinical responsibilities attributed. Diagnoses are defined according to 3-digit ICD-10 codes, which roughly corresponds to a disease. Standard errors are clustered at the hospital level.

Figure 3: Effects on Allocation of Physicians across Areas of Care



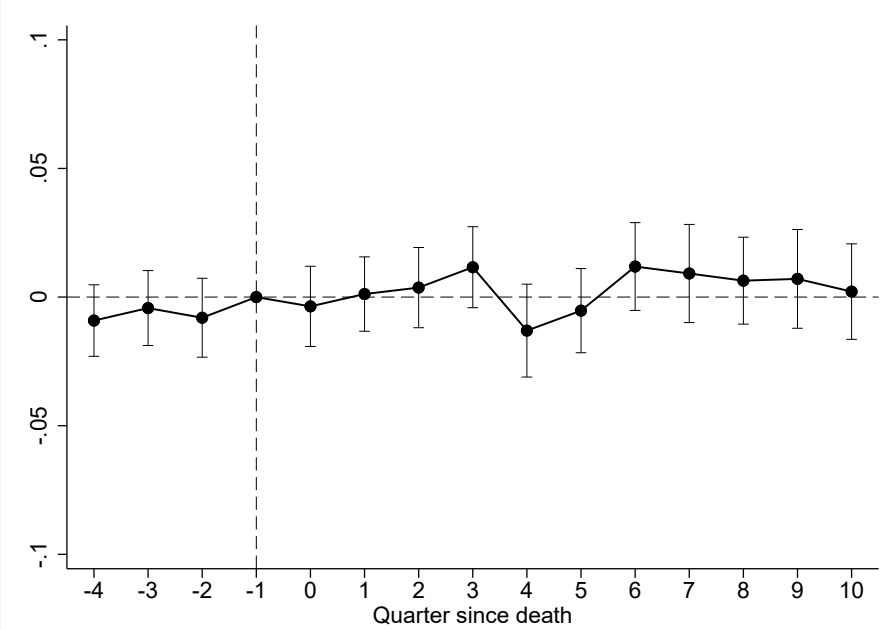
Notes. This figure presents estimates based on the model (2) with respective 95 percent confidence intervals. Panel A considers the log number of distinct physicians providing care in a given disease area. Panel B considers the patient share of new entrants in a disease area in the quarter they start operations in the hospital. The dashed line rescales the estimated effects by the death-induced change in the hospital's share of new entrants. To calculate this change in the share of new entrants, we use the event-study model (4). We use the estimated effects for event times 0 and 1 as rescaling factors. Standard errors are clustered at the hospital level.

Figure 4: Effect on Log Number of Patients



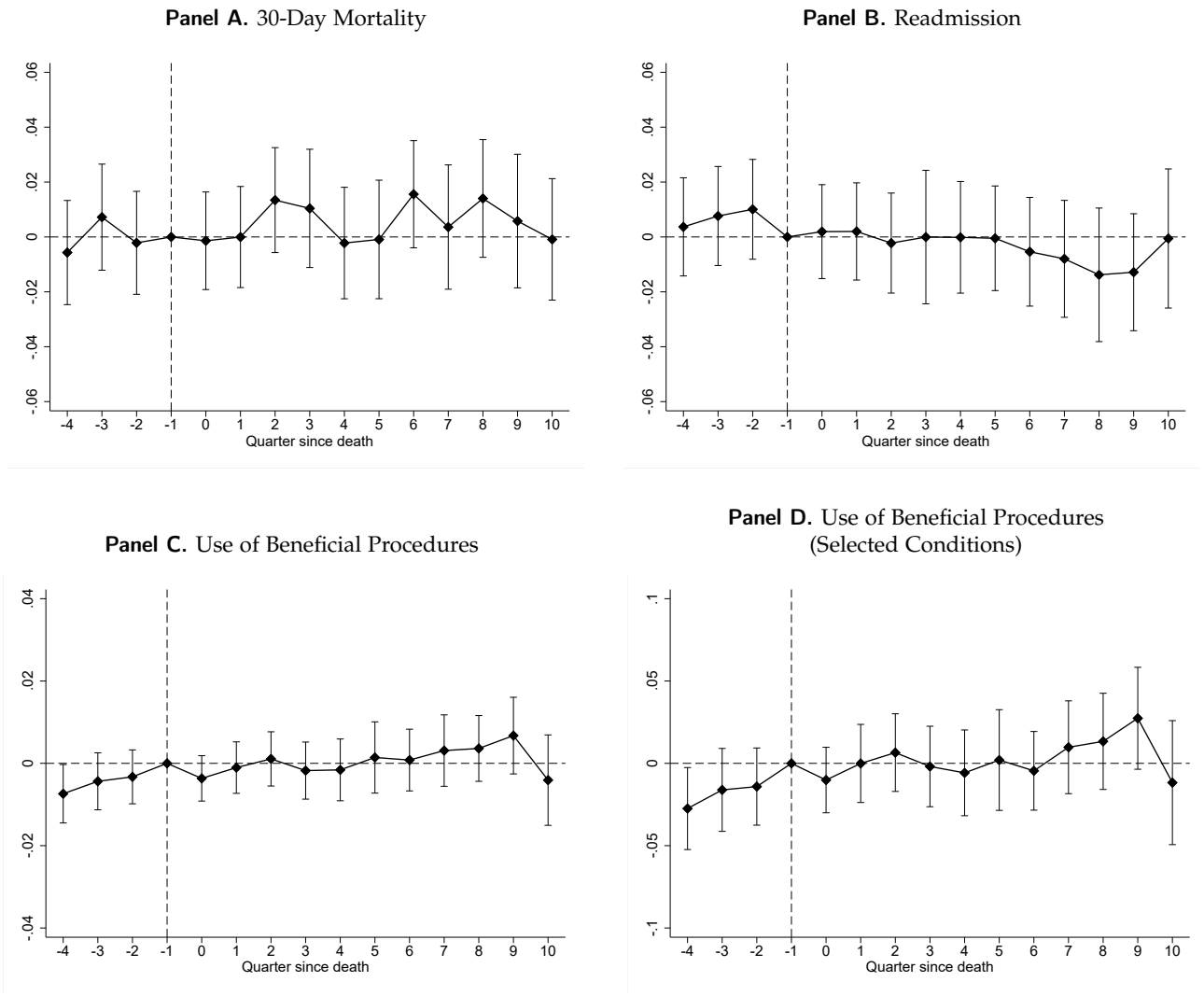
Notes. This figure presents estimates based on the model (2) with respective 95 percent confidence intervals. The dependent variable is the log number of patients in a disease area. Standard errors are clustered at the hospital level.

Figure 5: Effect on Inpatient Mortality:
At-Risk Patients



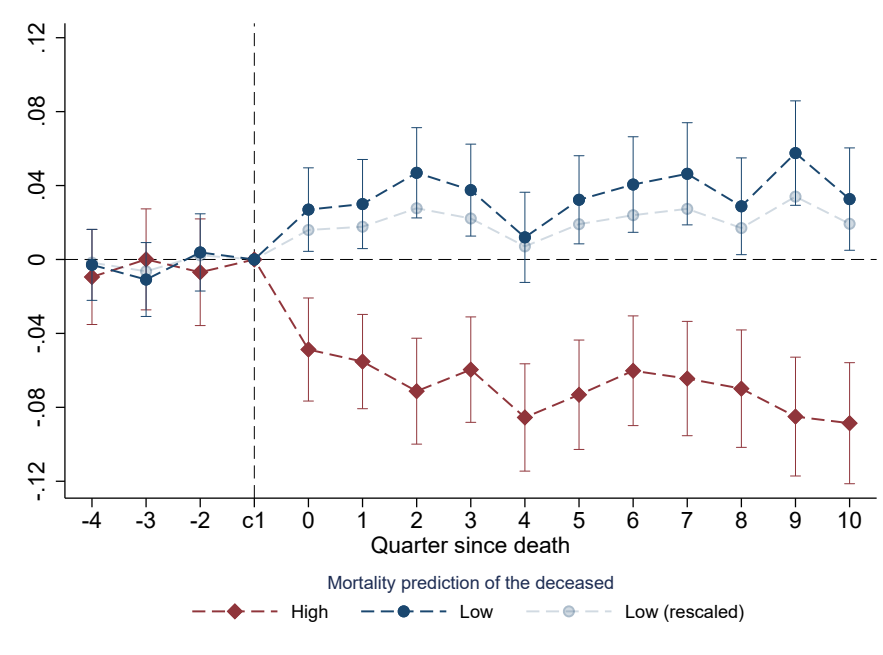
Notes. This figure presents estimates based on the model (2) with respective 95 percent confidence intervals. The dependent variable is the risk-adjusted inpatient mortality rate in a disease area among patients in the at-risk subsample. We compute risk-adjusted mortality rates using the two-step procedure described in Section 3.4. At-risk patients are defined as those in the top quartile of the predicted mortality distribution. The observations are weighted by cell size to account for differences in the precision with which the cell means are estimated. Standard errors are clustered at the hospital level.

Figure 6: Effects on Other Measures of Care Quality



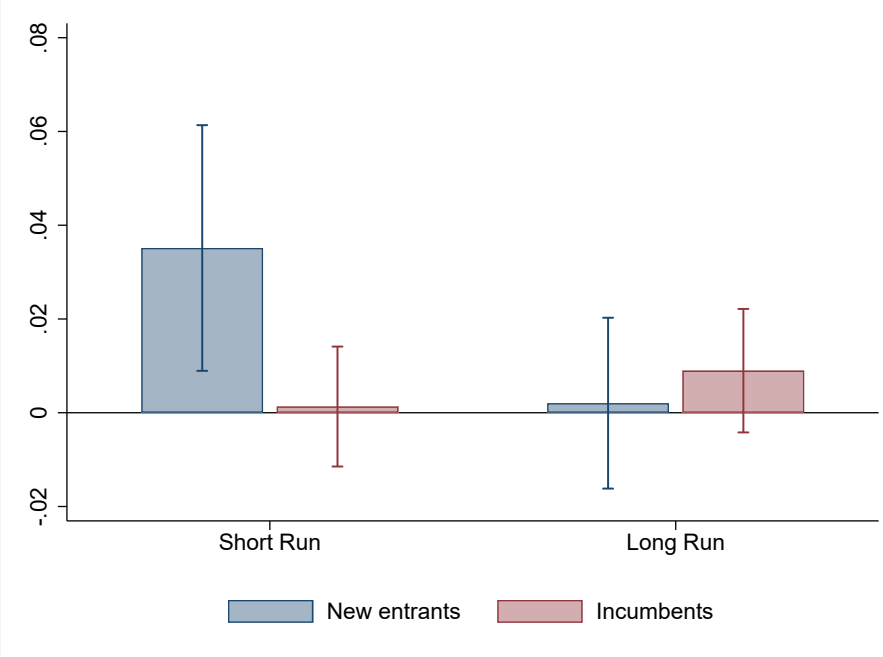
Notes. This figure presents estimates based on the model (2) with respective 95 percent confidence intervals. The sample is restricted to the at-risk sample, defined as those patients in the top quartile of the predicted mortality distribution. All outcomes are risk-adjusted using the two-step procedure described in Section 3.4. Readmissions corresponds to the share of patients readmitted in some hospital within 30 days since discharge. Beneficial procedures correspond to the use of high-technology, high-complexity procedures, as defined by the Brazilian Ministry of Health. Panel D further restricts the data to conditions whose the utilization rate of these procedures are above the 75 percentile of the utilization distribution. All regressions weight the observations by cell size to account for differences in the precision with which the cell means are estimated. Standard errors are clustered at the hospital level.

Figure 7: Effect on Inpatient Mortality:
Heterogeneity by Mortality Prediction of the Deceased Physician
(At-risk Patients)



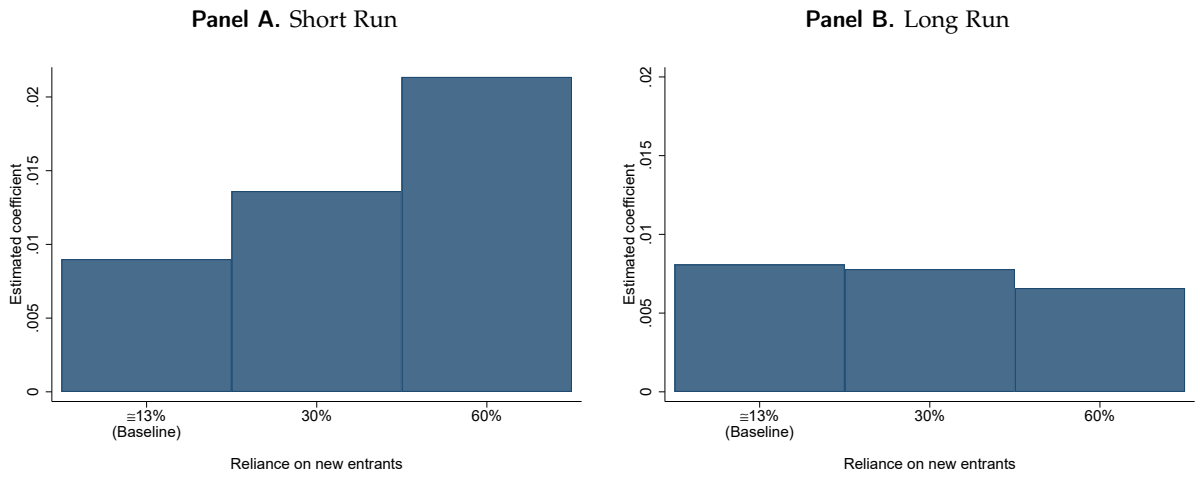
Notes. This figure presents estimates based on the model (2) with respective 95 percent confidence intervals. The dependent variable is the risk-adjusted inpatient mortality rate in a disease area among patients in the at-risk subsample. We compute risk-adjusted mortality rates using the two-step procedure described in Section 3.4. At-risk patients are defined as those in the top quartile of the predicted mortality distribution. The mortality prediction of a deceased physician is measured by their risk-adjusted inpatient mortality rate in a disease area, using the subsample of patients at risk of mortality. The mortality prediction is normalized by subtracting its disease-area-specific mean. The high- and low-mortality groups correspond to treated observations where the deceased physician's normalized mortality prediction is above and below the median, respectively. The dynamic treatment effects are estimated separately for each group, using the full set of control hospitals in both subsamples to maximize precision. To account for differences in expected mortality changes and make the estimates as comparable as possible, the low-mortality coefficients are adjusted by scaling them with the ratio of the average mortality prediction in the high-mortality subsample to that in the low-mortality subsample. The observations are weighted by cell size to account for differences in the precision with which the cell means are estimated. Standard errors are clustered at the hospital level.

Figure 8: Effect on Inpatient Mortality:
Post-period Decomposition
(At-risk Patients)



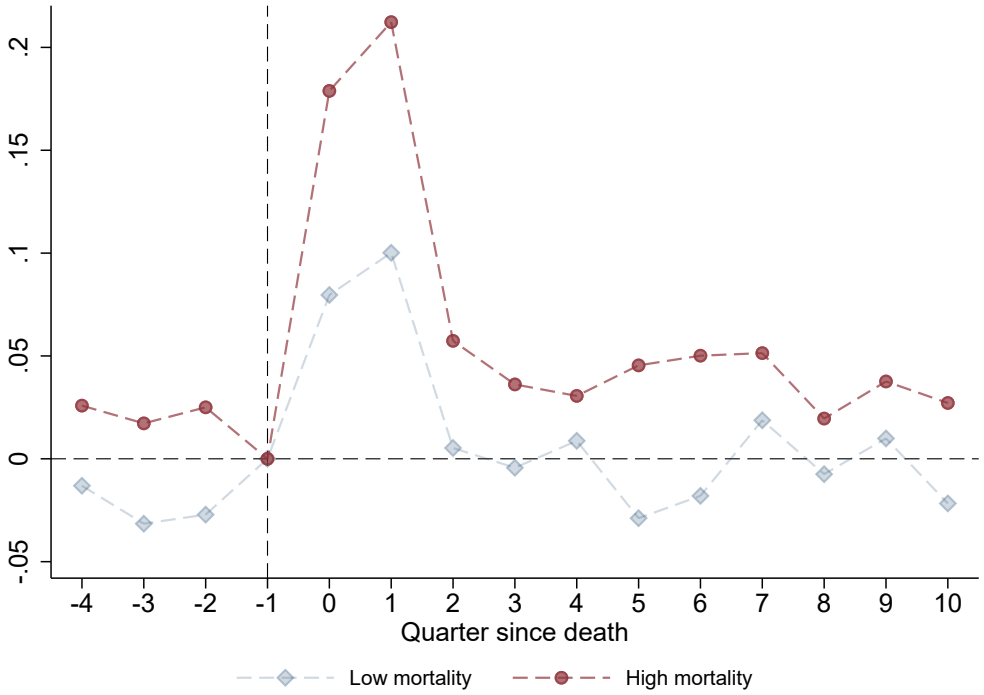
Notes. This figure presents estimates based on the specifications (3), with respective 95 percent confidence intervals. The dependent variable is the risk-adjusted inpatient mortality rate in a disease area among patients in the at-risk subsample. We compute risk-adjusted mortality rates using the two-step procedure described in Section 3.4. At-risk patients are defined as those in the top quartile of the predicted mortality distribution. The dependent variable in the post-event observations is computed from at-risk patients attended only by either new entrants or incumbents, as indicated in the figures. The dependent variable for the quarter of the event or the preceding quarters is computed from all patients in the at-risk subsample. All the observations are weighted by cell size to account for differences in the precision with which the cell means are estimated. Standard errors are clustered at the hospital level.

Figure 9: Counterfactual Effects on Inpatient Mortality



Notes. This figure presents counterfactual estimates of the treatment effect on inpatient mortality under scenarios where new entrants handle a larger share of cases. This exercise relies on the fact that the overall treatment effect is approximately equal to the weighted average of the treatment effects for the subsamples of new entrants and incumbents, with the weights corresponding to the proportion of patients in each subsample. To compute the counterfactual estimates, we combine the estimation subsamples into a single file and rerun the baseline specification but reweighting the importance of each subsample based on alternative patient shares attributed to new entrants. The dependent variable is the inpatient mortality rate in a disease area among patients in the at-risk subsample. We compute risk-adjusted mortality rates using the two-step procedure described in Section 3.4. At-risk patients are defined as those in the top quartile of the predicted mortality distribution.

Figure 10: Effects on the Patient Shares of New Entrants:
Heterogeneity by Mortality Prediction of the Deceased Physician



Notes. This figure presents estimates based on the model (2) with respective 95 percent confidence intervals. The dependent variable is the share of patients attended by new entrants in a disease area in the quarter they start operations in the hospital. The observations are weighted by cell size to account for differences in the precision with which the cell means are estimated. The mortality prediction of a deceased physician is measured by their risk-adjusted inpatient mortality rate in a disease (using the subsample of patients at risk of mortality) and normalized by subtracting its disease-area-specific mean. The high- and low-mortality groups correspond to treated observations where the deceased physician's normalized mortality prediction is above and below the median, respectively. The dynamic treatment effects are estimated separately for each group, using the full set of control hospitals in both subsamples to maximize precision. Standard errors are clustered at the hospital level.

Table 1: Summary Statistics:
Deceased versus Alive Physicians

	Deceased		Alive	
	Estimation sample (1)	All (2)	Estimation sample (3)	All (4)
Male	0.854	0.829	0.615	0.589
	[0.353]	[0.376]	[0.487]	[0.492]
Age	57.91	61.76	45.49	47.65
	[32.30]	[19.85]	[13.35]	[14.94]
Age at death	61.49	68.03	—	—
	[27.56]	[20.10]		
Experience	28.89	33.30	17.95	17.19
	[14.84]	[14.94]	[13.17]	[16.81]
Number of specialties	3.202	2.710	2.677	2.463
	[1.602]	[1.523]	[1.583]	[1.525]
Non-clinical duties	0.0431	0.0354	0.0165	0.0426
	[0.203]	[0.185]	[0.127]	[0.202]
Number of physicians	1351	6806	205716	104138
Physician-quarter observations	70171	269102	11848428	2661036

Notes. This table shows summary statistics for deceased and alive physicians. These statistics are calculated by averaging all Physician-hospital-quarter observations in the dataset, with the exception of the age at death. Standard deviations in brackets.

Table 2: Effects on Incumbent Physicians' Workload

	Dependent variable is:		
	Log (number of patients)	Number of distinct diagnoses treated	Indicator for nonclinical activities
	(1)	(2)	(3)
Exposure × short run	0.428	3.714	0.069
	[0.106]***	[0.699]***	[0.011]***
Exposure × long run	0.203	2.188	0.096
	[0.119]*	[0.805]***	[0.014]***
Pre-event control mean	2.35	8.726	0.015
Basic controls	✓	✓	✓
Observations	5,390,827	5,390,827	5,390,827

Notes. This table presents estimates based on a reduced version of model (5), in which the full set of quarter-since-event indicators is replaced by two indicators: one for the two quarters immediately following the event (*short run*) and another for the period from the third quarter onward (*long run*). Incumbent physicians correspond to those who were affiliated with the hospital before the death event. Column (1) considers changes in the number of patients managed. Column (2) looks at the number of distinct diagnoses. Column (3) examines an indicator for whether the physician has nonclinical responsibilities attributed. Diagnoses are defined according to 3-digit ICD-10 codes, which roughly corresponds to a disease. Standard errors are clustered at the hospital level. * $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

Table 3: Effect on Allocation of Physicians across Areas of Care

	Dependent variable is:				
	Log (number of FTE physicians on service) (1)	Log (number of physicians on service) (2)	Number of physicians on service (3)	Number of physicians on service (4)	Share of patients attended by new entrants (5)
Exposure × short run	-0.057	-0.018	0.005	0.005	0.084
	[0.060]	[0.037]	[0.024]	[0.024]	[0.016]***
Exposure × long run	-0.031	0.045	0.046	0.046	0.011
	[0.062]	[0.045]	[0.028]	[0.028]	[0.006]*
Pre-event control mean	1.019	1.662	8.723	8.723	0.019
Specification	OLS	OLS	Poisson	Neg. Bin	OLS
Basic controls	✓	✓	✓	✓	✓
Number of observations	6,500,630	6,500,630	6,500,630	6,500,630	6,500,630

Notes. This table presents estimates based on the model (3). Column (1) investigates the log number of distinct FTE physicians providing care in a given disease area. Columns (2)-(4) explore the robustness of these results to alternative specifications. Column (2) uses simple headcounts of physicians in logs as the dependent variable (rather than FTE units). Columns (3) and (4) consider headcounts of physicians in levels and employ count data models (Poisson and Negative Binomial, respectively) instead of OLS. Column (5) examines the patient share of new entrants in a disease area in the quarter they start operations in the hospital. Standard errors are clustered at the hospital level.

* $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

Table 4: Effect on Volume

	Dependent variable is:				
	log(number of patients) (1)	number of patients (2)	number of patients (3)	log(number of patients/ number of beds) (4)	log(number of patients/ number of physicians) (5)
Exposure × short run	-0.093	-0.098	-0.098	-0.082	-0.088
	[0.020]***	[0.028]***	[0.028]***	[0.023]***	[0.019]***
Exposure × long run	-0.117	-0.117	-0.117	-0.097	-0.121
	[0.021]***	[0.034]***	[0.034]***	[0.023]***	[0.019]***
Pre-event control mean	1.065	8.689	8.689	-3.803	0.489
Specification	OLS	Poisson	Neg. Bin.	OLS	OLS
Basic controls	✓	✓	✓	✓	✓
Observations	6,500,630	6,500,630	6,500,630	6,500,630	6,500,630

Notes. This table presents estimates based on the model (3). Column (1) investigates the log number of patients in a disease area. Columns (2) and (3) use patient headcounts in levels and employ count data models (Poisson and Negative Binomial, respectively) instead of OLS. Columns (4) and (5) examine the ratio (in logs) of patients to beds and patients to physicians, respectively. Standard errors are clustered at the hospital level.

* $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

Table 5: Effects on Selection of Patients and Intensity of Care

	Dependent variable is:				
		Share patients above percentile		Intensity of care	
	Predicted mortality (1)	75th mortality risk distrubtion (2)	90th mortality risk distrubtion (3)	Length of stay (4)	number of procedures (5)
Exposure × short run	0.001	0.005	0.015	-0.035	-0.058
	[0.001]	[0.008]	[0.007]**	[0.069]	[0.188]
Exposure × long run	0.002	0.011	0.017	-0.077	-0.171
	[0.001]*	[0.008]	[0.008]**	[0.082]	[0.219]
Pre-event control mean	0.049	0.213	0.085	5.58	8.707
Risk-adjustment				✓	✓
Basic controls	✓	✓	✓	✓	✓
Number of observations	6,500,630	6,500,630	6,500,630	6,500,630	6,500,630

Notes. This table presents estimates based on the model (3). Column (1) investigates the predicted in-patient mortality rate. Columns (2) and (3) look at the log number of at-risk and non-at-risk patients, respectively. Column (4) focuses on the total number of procedures per patient, and column (5) considers length of stay. Outcomes in columns (4) and (5) are risk-adjusted, using the two-step procedure described in Section 3.4. We predict mortality by regressing it on patient characteristics, including demographic and indicators for a comprehensive set of admission-level comorbidities. At-risk patients are defined as those in the top quartile of the predicted mortality distribution. The observations in columns (1), (4), and (5) are weighted by cell size to account for differences in the precision with which the cell means are estimated. Standard errors are clustered at the hospital level.
* $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

Table 6: Effect on Inpatient Mortality:
At-Risk Patients

	Dependent variable is:	
	Predicted mortality	Actual mortality
	(1)	(2)
Exposure × short run	0.002	0.009
	[0.001]	[0.006]
Exposure × long run	-0.000	0.009
	[0.002]	[0.005]
Pre-event control mean	0.134	0.128
Risk-adjustment		✓
Basic controls	✓	✓
Observations	2,790,873	2,790,873

Notes. This table presents estimates based on the model (3), with the sample restricted to at-risk patients. Column (1) analyzes the predicted inpatient mortality rate. Columns (2) and (3) consider the inpatient mortality rate in a disease area; the outcome in column (3) is risk-adjusted using the two-step procedure described in Section 3.4. At-risk patients are defined as those in the top quartile of the predicted mortality distribution. The observations are weighted by cell size to account for differences in the precision with which the cell means are estimated. Standard errors are clustered at the hospital level.

* $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

Online Appendix to "Organizational Responses to Labor Turnover: Evidence from Hospitals"

Danyelle Branco
Sao Paulo School of Economics FGV

Bladimir Carrillo
Sao Paulo School of Economics FGV

Luiz Fontes
Institute of Education and Research - INSPER

Leticia Nunes
Institute of Education and Research - INSPER

February 13, 2025

List of Figures 1

List of Tables 2

A Descriptive Statistics 3

B Additional Tables and Figures 6

C Data Sources and Construction 27

References 28

List of Figures

A.1 Timing of Deaths . . . 3
A.2 Exposure to Deceased Physicians Across Disease Areas . . . 4
A.3 Mortality Rate of Deceased Relative to ICD-specific Mean . . . 5
B.1 Effect on Log(number of new affiliates): Robustness to Alternative Calibrations of Extensive Margin . . . 6
B.2 Effect on Log (Number of Separations): Exclude Deceased . . . 7

Contact information:

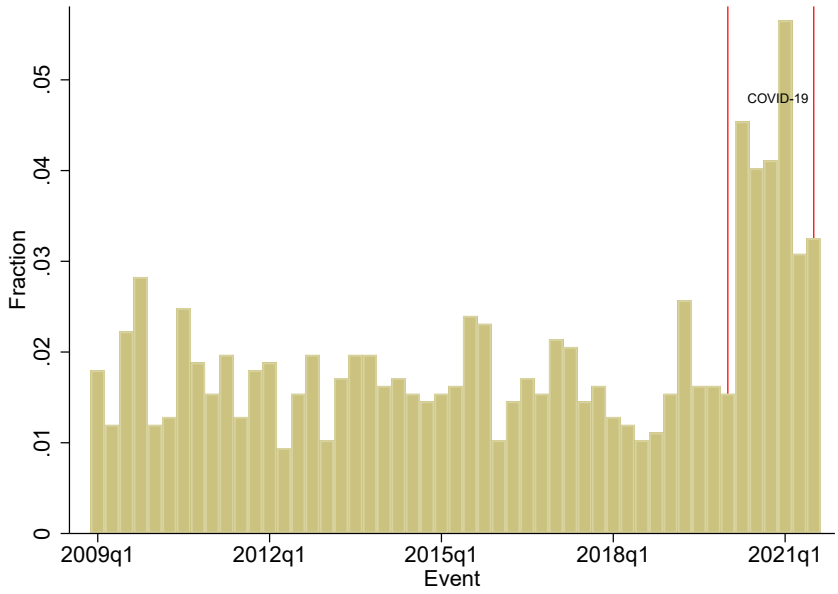
B.3 Effect on Physician-Patient Mismatch . . . 8
B.4 Effect on New Hires/Total Physicians . . . 9
B.5 Effect on Log (number of Patients): hospital versus care-area level estimates . . . 10
B.6 Effect on Inpatient Mortality: Alternative Definitions of At-risk Patients . . . 11
B.7 Effect on Inpatient Mortality: Control Hospitals Outside State (At-risk Patients) . . . 12
B.8 Effect on Predicted Inpatient Mortality: Heterogeneity by Mortality Prediction of Deceased Physician (At-risk patients) . . . 13
B.9 Effect on Predicted Inpatient Mortality: Post-period Decomposition (At-risk patients) . . . 14

List of Tables

B.1 Summary Statistics: Hospital-Level . . . 15
B.2 Characteristics of Alive and Deceased Physicians . . . 16
B.3 Effect on Characteristics of Physicians . . . 17
B.4 Effect on Incumbent Physicians: Other Outcomes . . . 18
B.5 Effects on Nurses . . . 19
B.6 Top 10 Causes of Admission in the At-risk Sample . . . 20
B.7 Effect on Inpatient Mortality: hospital-level exposure . . . 21
B.8 Assessing External Validity . . . 22
B.9 Effect on Inpatient Mortality: Post-period Decomposition . . . 23
B.10 Robust Inference . . . 24
B.11 Robustness to Sample Restrictions . . . 25
B.12 Pre-Trend Tests . . . 26

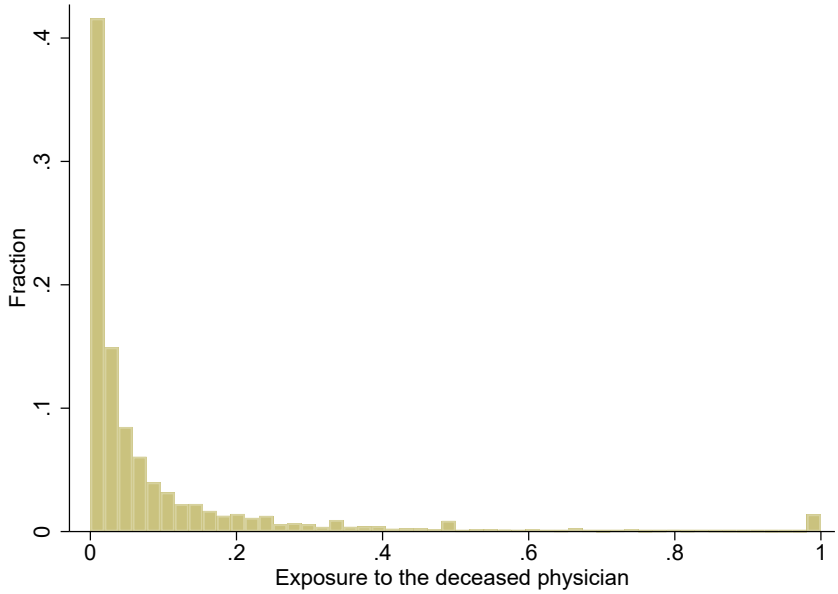
A Descriptive Statistics

Figure A.1: Timing of Deaths



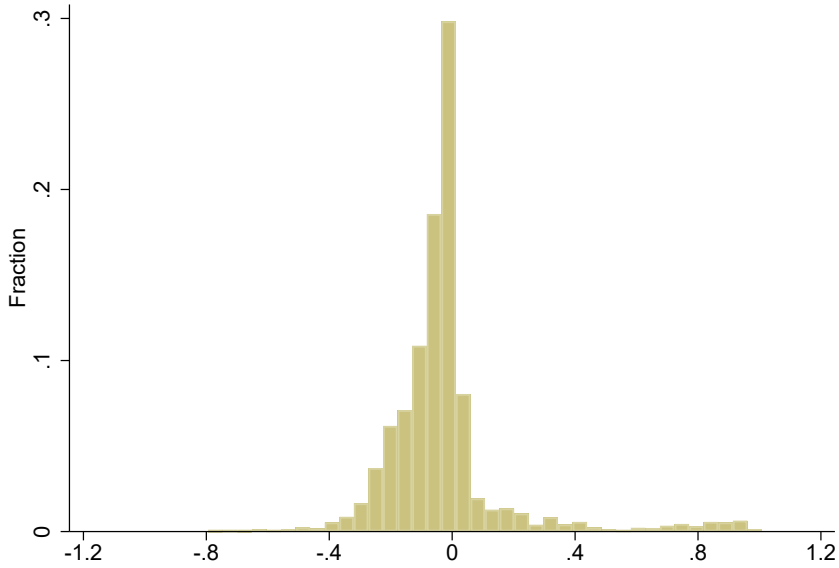
Notes. This figure plots the distribution of physician deaths over time.

Figure A.2: Exposure to Deceased Physicians Across Disease Areas



Notes. This figure plots the distribution of the exposure to deceased physicians across areas where they provided care.

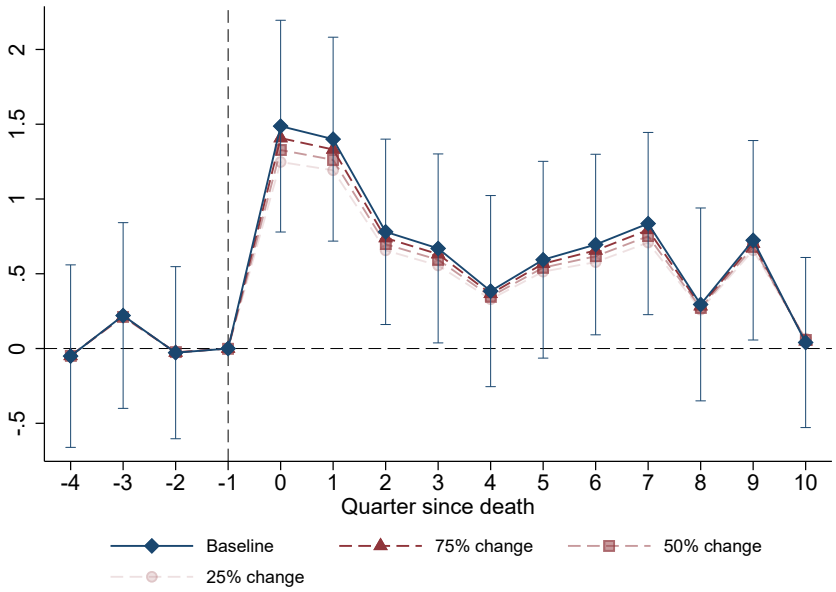
Figure A.3: Mortality Rate of Deceased Relative to ICD-specific Mean



Notes. This figure plots the distribution of the deceased physicians' risk-adjusted inpatient mortality rate in a disease area, using the subsample of patients at risk of mortality. The variable is normalized by subtracting its area-specific mean. At-risk patients are defined as those in the top quartile of the predicted mortality distribution.

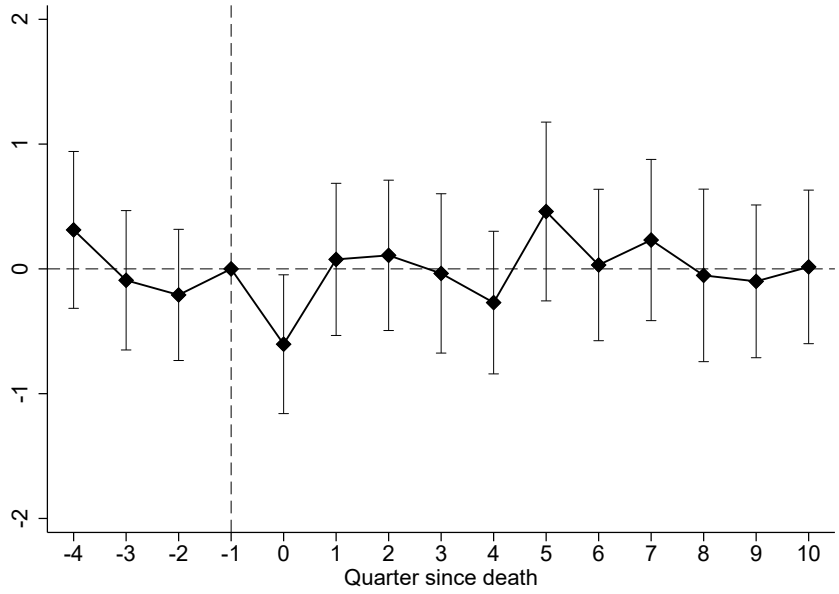
B Additional Tables and Figures

Figure B.1: Effect on Log(number of new affiliates):
Robustness to Alternative Calibrations of Extensive Margin



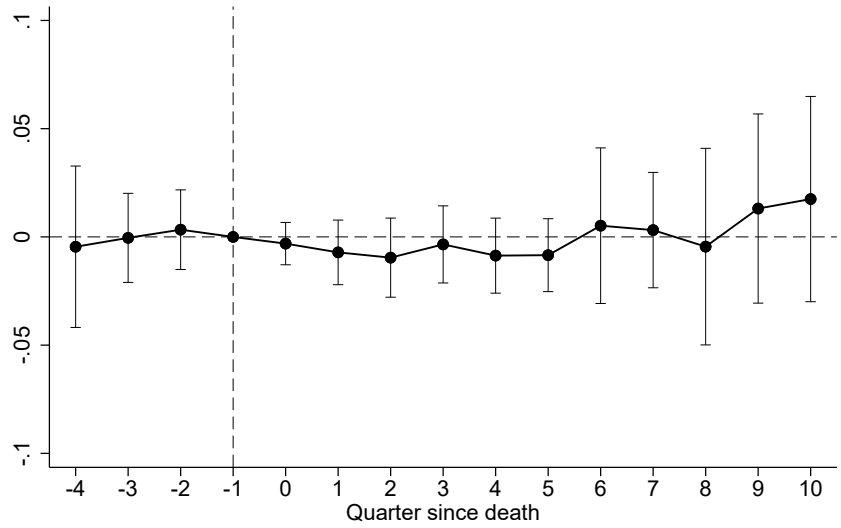
Notes. This figure presents estimates based on the model (4) with respective 95 percent confidence. The dependent variable is the log number of newly affiliated physicians. To account for the fact that many hospital quarters have no hiring activity, the extensive margin in our baseline results (the same from Figure 1, panel B) is calibrated following Chen and Roth (2024): we assign a value of 100 log points to an extensive margin change between 0 and the minimum non-zero value of the dependent variable. This figure additionally reports the results when we assign a value of 25, 50, and 75 log points for the same change. Standard errors are clustered at the hospital level.

Figure B.2: Effect on Log (Number of Separations): Exclude Deceased



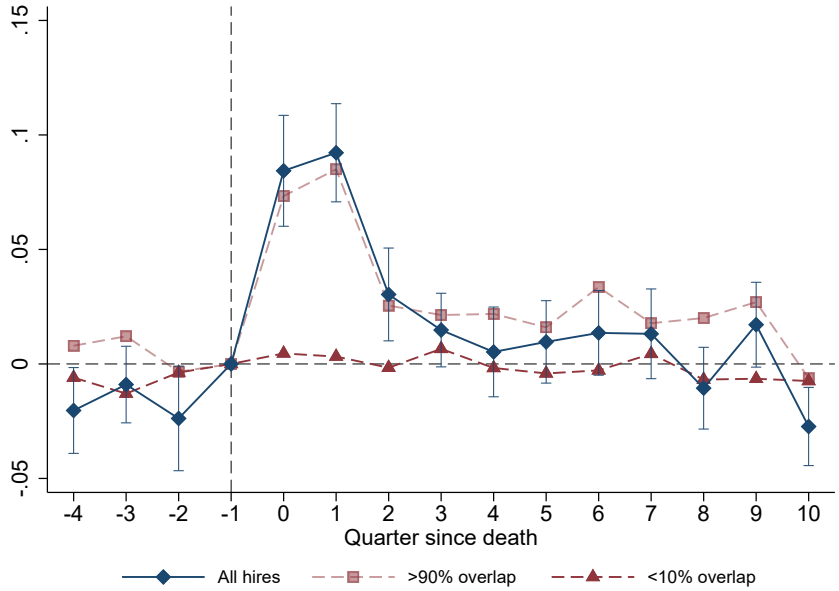
Notes. This figure presents estimates based on the model (4) with respective 95 percent confidence intervals. The dependent variable is the log number of FTE separations among physicians (excluding the deceased). Standard errors are clustered at the hospital level.

Figure B.3: Effect on Physician-Patient Mismatch



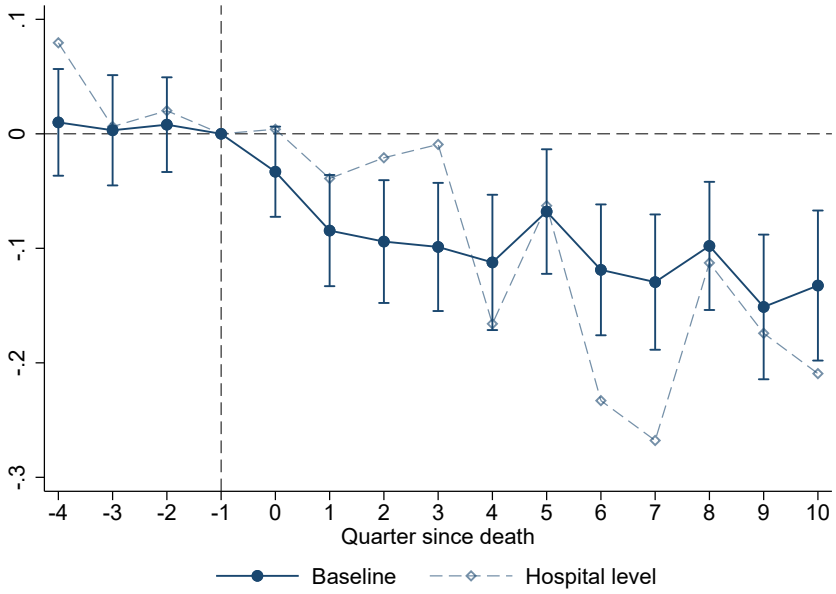
Notes. This figure presents estimates based on the model (2) with respective 95 percent confidence intervals. The dependent variable is the average mismatch between physicians and the medical needs of their patients across all patient-physician pairs in a given disease area. We define mismatch based on data from the Ministry of Health indicating the list of specialties considered clinically relevant for each procedure, as outlined in Section 5.2.B. The observations are weighted by cell size to account for differences in the precision with which the cell means are estimated. Standard errors are clustered at the hospital level.

Figure B.4: Effect on New Hires/Total Physicians



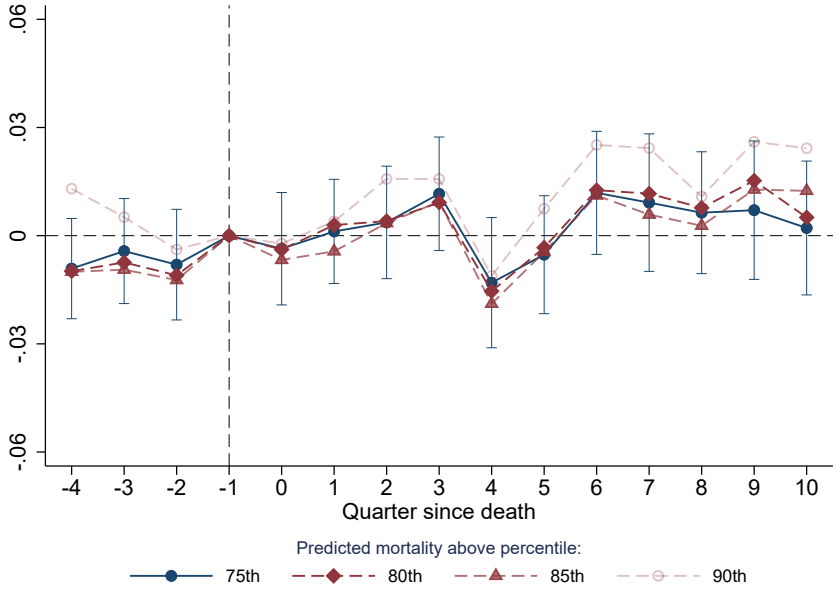
Notes. This figure presents estimates based on the model (4) with respective 95 percent confidence. The dependent variable is the ratio between the number of new affiliated physicians to the total number of affiliated physicians, distinguishing between new hires with high and low overlap with the deceased in the medical procedures that they are qualified to perform. To compute the overlap in tasks, we first identify all medical procedures performed by the deceased in the pre-event period. We compare this list of procedures with the medical procedures that new hires are qualified to perform given their specialties. Standard errors are clustered at the hospital level.

Figure B.5: Effect on Log (number of Patients): hospital versus care-area level estimates



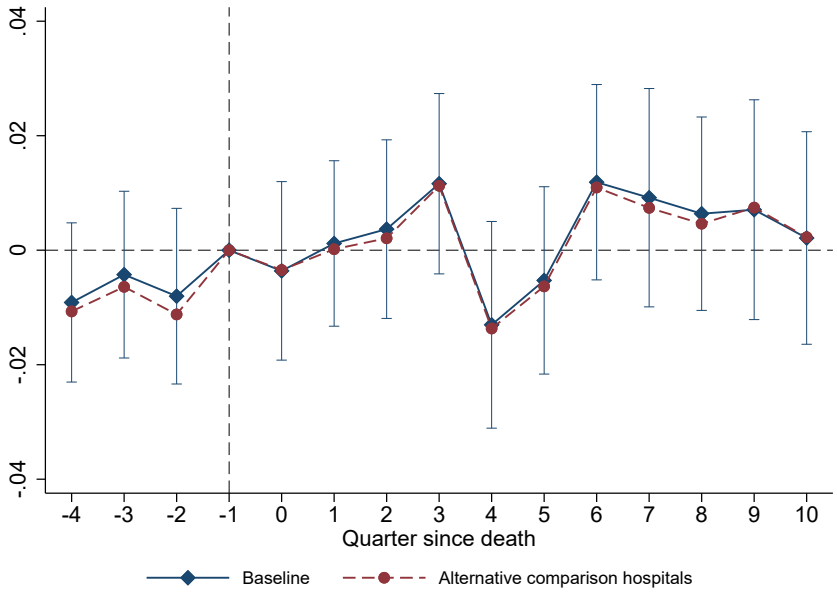
Notes. This figure presents estimates based on the models (2) and (4) with respective 95 percent confidence intervals. The dependent variable is the log number of patients. We estimate our baseline results (the same from Figure 4) using model (2). This figure additionally plots hospital-level results based on model (4). Standard errors are clustered at the hospital level.

Figure B.6: Effect on Inpatient Mortality:
Alternative Definitions of At-risk Patients



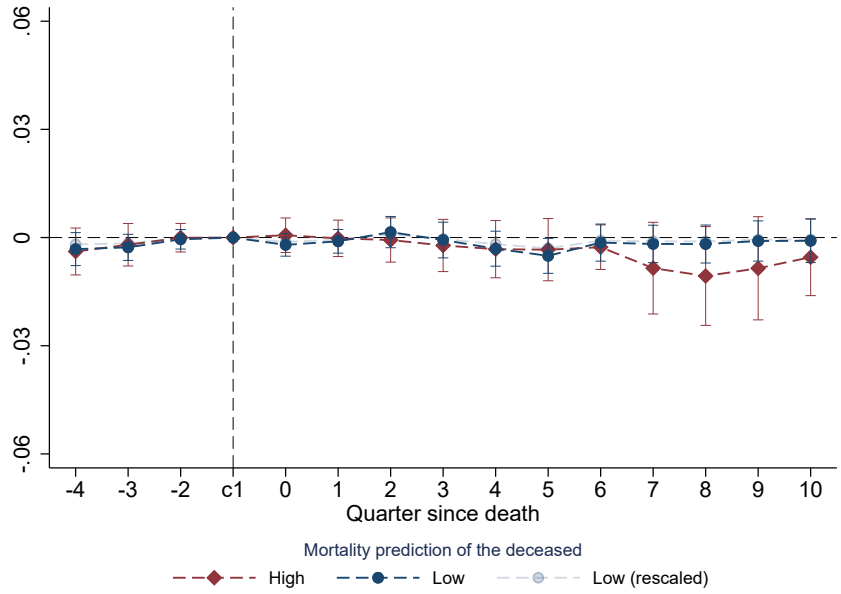
Notes. This figure presents estimates based on the model (2) with respective 95 percent confidence intervals. The dependent variable is the risk-adjusted inpatient mortality rate, restricted to patients in the at-risk subsample. In our baseline results (the same from Figure 5), at-risk patients are defined as those above the 75th percentile of the predicted mortality distribution. This figure additionally show results for thresholds at the 80th, 85th, and 90th percentiles. Standard errors are clustered at the hospital level.

Figure B.7: Effect on Inpatient Mortality:
Control Hospitals Outside State
(At-risk Patients)



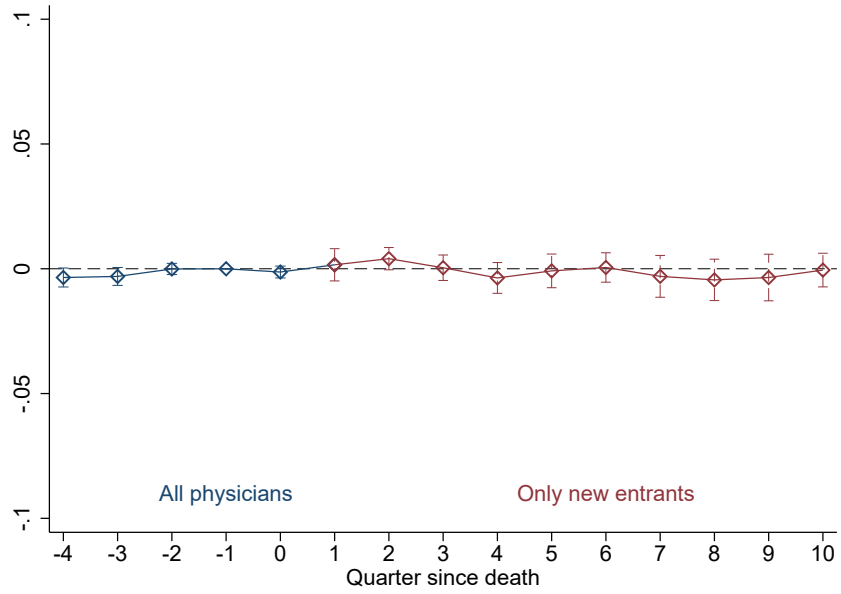
Notes. This figure presents estimates based on the model (2) with respective 95 percent confidence intervals. The dependent variable is the risk-adjusted inpatient mortality rate, restricted to patients in the at-risk subsample. In our baseline results (the same from Figure 5), we use as the comparison group hospitals selected following the two-step matching algorithm outlined in Section 3.3. This figure additionally reports the results when we use an alternative comparison group, selected from the same matching algorithm but with the restriction that the control hospitals must be located outside the state of the treated hospital. Standard errors are clustered at the hospital level.

Figure B.8: Effect on Predicted Inpatient Mortality:
Heterogeneity by Mortality Prediction of Deceased Physician
(At-risk patients)



Notes. This figure presents estimates based on the model (2) with respective 95 percent confidence intervals. The dependent variable is the predicted inpatient mortality rate in a disease area among patients in the at-risk subsample. We predict mortality by regressing it on patient characteristics, including demographic and indicators for a comprehensive set of admission-level comorbidities. At-risk patients are defined as those in the top quartile of the predicted mortality distribution. The mortality prediction of a deceased physician is measured by their risk-adjusted inpatient mortality rate in a disease area, using the subsample of at-risk patients. The mortality prediction is normalized by subtracting its disease-area-specific mean. The high- and low-mortality groups correspond to treated observations where the deceased physician's normalized mortality prediction is above and below the median, respectively. The dynamic treatment effects are estimated separately for each group, using the full set of control hospitals in both subsamples to maximize precision. To account for differences in expected mortality changes across groups, the low-mortality coefficients are adjusted by scaling them with the ratio of the average mortality prediction in the high-mortality subsample to that in the low-mortality subsample. The observations are weighted by cell size to account for differences in the precision with which the cell means are estimated. Standard errors are clustered at the hospital level.

Figure B.9: Effect on Predicted Inpatient Mortality:
Post-period Decomposition
(At-risk patients)



Notes. This figure presents estimates based on the specifications (2) with respective 95 percent confidence intervals. The dependent variable is the predicted inpatient mortality rate in a disease area among patients in the at-risk subsample. The dependent variable in the post-event observations is computed from at-risk patients attended only by new entrants. We predict mortality by regressing it on patient characteristics, including demographic and indicators for a comprehensive set of admission-level comorbidities. At-risk patients are defined as those in the top quartile of the predicted mortality distribution. The dependent variable for the quarter of the event or the preceding quarters is computed from all patients in the at-risk subsample. All the observations are weighted by cell size to account for differences in the precision with which the cell means are estimated. Standard errors are clustered at the hospital level.

Table B.1: Summary Statistics:
Hospital-Level

	Treated hospitals	Matched hospitals	All untreated
	(1)	(2)	(3)
Male	0.589	0.614	0.627
	[0.492]	[0.487]	[0.484]
Age	45.75	46.11	46.35
	[15.11]	[12.91]	[14.23]
Experience	17.86	18.47	17.89
	[13.34]	[13.06]	[14.61]
Number of specialties	2.471	2.626	2.753
	[1.536]	[1.463]	[1.588]
Non-clinical duties	0.0133	0.0157	0.0278
	[0.115]	[0.124]	[0.165]
Inpatient mortality rate	0.0466	0.0412	0.0395
	[0.0423]	[0.0336]	[0.0818]
Volume	1595.1	1326.9	360.6
	[1756.0]	[1099.3]	[556.1]
Number of FTE physicians	93.60	78.43	11.17
	[172.0]	[99.52]	[26.83]
Number of FTE nurses	67.40	54.59	13.82
	[109.2]	[82.11]	[31.72]
Numeber of beds	184.3	171.9	61.28
	[200.0]	[148.9]	[71.76]
Specialized hospital	0.107	0.166	0.110
	[0.310]	[0.372]	[0.313]
Teaching hospital	0.277	0.314	0.0680
	[0.448]	[0.464]	[0.252]
Emergency department	0.916	0.926	0.815
	[0.278]	[0.262]	[0.388]
Operating rooms	0.944	0.944	0.759
	[0.231]	[0.229]	[0.428]
Hospital-event pairs	1546	1546	5743

Notes. This table shows summary statistics for hospitals in our data. In columns (1) and (2), variables are measured a quarter before the death. In column (3), variables are measured during the entire study period, but each quarter is weighted by the share of hospitals that experienced the death of a physician in that quarter. Standard deviations in parentheses.

Table B.2: Characteristics of Alive and Deceased Physicians

	Dependent variable is:							
	Indicator for male		Experience		Number of specialties		Indicator for nonclinical activities	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Deceased	0.177	0.089	11.962	-0.294	0.154	0.041	0.008	0.004
	[0.006]***	[0.006]***	[0.232]***	[0.145]**	[0.020]***	[0.021]**	[0.002]***	[0.002]**
Sample mean	0.618	0.618	18.142	18.142	2.680	2.680	0.017	0.017
Age FE		✓		✓		✓		✓
Hospital× quarter-year FE	✓	✓	✓	✓	✓	✓	✓	✓
Observations	11,976,202	10,705,760	11,494,550	10,615,323	11,976,202	10,705,760	11,909,052	10,643,120

Notes. This table presents results from OLS regressions of physician characteristics on a binary variable that equals 1 if the physician died and hospital×time fixed effects. Even-numbered columns consider an augmented specification that adds age fixed effects. * $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

Table B.3: Effect on Characteristics of Physicians

	Dependent variable is:			
	Number of specialties (1)	Experience since medical school (2)	Age (3)	Male (4)
Exposure × short run	-0.088	-4.445	-5.910	-0.035
	[0.058]	[0.616]***	[1.155]***	[0.014]**
Exposure × long run	-0.161	-6.445	-7.114	-0.048
	[0.073]**	[0.815]***	[1.181]***	[0.017]***
Pre-event control mean	2.57	17.917	45.247	0.611
Number of observations	6,102,635	6,102,635	6,102,635	6,102,635

Notes. This table presents estimates based on a reduced version of model (4), in which the full set of quarter-since-event indicators is replaced by two indicators: one for the two quarters immediately following the event (*short run*) and another for the period from the third quarter onward (*long run*). Columns (1) through (6) investigate changes in the average characteristics of physicians in a given hospital.Standard errors are clustered at the hospital level.
* $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

Table B.4: Effect on Incumbent Physicians: Other Outcomes

	Dependent variable is:				
	Number of distinct diagnosis sub-chapters (1)	High-hour contract (2)	Low-hour contract (3)	Hours worked (4)	Log(number of patients / hours worked) (5)
Exposure × short run	1.864	0.031	0.026	-0.124	0.479
	[0.301]***	[0.020]	[0.020]	[0.640]	[0.114]***
Exposure × long run	0.882	0.038	0.030	-0.217	0.274
	[0.330]***	[0.026]	[0.028]	[0.804]	[0.134]**
Pre-event control mean	3.094	0.18	0.207	20.376	-0.528
Basic controls	✓	✓	✓	✓	✓
Observations	5,390,827	5,390,827	5,390,827	5,390,827	5,390,827

Notes. This table presents estimates based on a reduced version of model (5), in which the full set of quarter-since-event indicators is replaced by two indicators: one for the two quarters immediately following the event (*short run*) and another for the period from the third quarter onward (*long run*). Incumbent physicians correspond to those who were affiliated with the hospital before the death event. Column (1) considers the number of distinct diagnosis categories treated, where categories are defined at the ICD-10 sub-chapter level rather than 3-digit ICD groups. Column (2) examines an indicator for whether the physician has a high-hour position, defined as physicians above the percentile 75th of the working hours distribution. Column (3) considers an indicator for whether the the physician has a low-hour position, defined as physicians below the percentile 25th of the working hours distribution. Column (4) measures total hours worked per week. Column (5) looks at the log number of patients managed per hour. Standard errors are clustered at the hospital level.

* $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

Table B.5: Effects on Nurses

	Dependent variable is:		
	Log(number of FTE separations) (1)	Log(number of new FTE nurses) (2)	Log(number of FTE nurses) (3)
Exposure × short run	-0.252	-0.037	-0.033
	[0.235]	[0.214]	[0.049]
Exposure × long run	-0.711	-0.460	0.058
	[0.174]***	[0.163]***	[0.060]
Pre-event control mean	-2.985	-2.137	2.372
Basic controls	✓	✓	✓
Observations	37,560	37,560	37,560

Notes. This table presents estimates based on a reduced version of model (4), in which the full set of quarter-since-event indicators is replaced by two indicators: one for the two quarters immediately following the event (*short run*) and another for the period from the third quarter onward (*long run*). Column (1) examines the log number of FTE separations among nurses. Column (2) considers the log number of new affiliated FTE nurses. Column (3) measures the log number of FTE nurses. Standard errors are clustered at the hospital level.
* $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

Table B.6: Top 10 Causes of Admission in the At-risk Sample

	Inpatient	Share of cases, %	
	mortality rate, %	At-risk sample	Full sample
Cause of admission:	(1)	(2)	(3)
Heart failure	13.302	3.176	0.127
Pneumonia, unspecified organism	19.226	2.498	0.1
Stroke, unspecified	17.857	2.472	0.099
Viral infection of unspecified site	27.287	1.988	0.08
Other chronic obstructive pulmonary disease	12.45	1.655	0.066
Acute myocardial infarction	16.242	1.537	0.062
Fracture of femur	5.332	1.489	0.06
Other sepsis	60.099	1.271	0.051
Angina pectoris	3.211	1.217	0.049
Inguinal hernia	.593	1.2	0.048

Notes. This figure displays the top 10 admission causes in the subsample of at-risk patients. Column (1) reports the inpatient mortality rate of these causes in the full sample. Columns (2) and (3) present their prevalence among cases in the subsample of at-risk patients and in the full sample, respectively. At-risk patients are defined as those in the top quartile of the predicted mortality distribution.

Table B.7: Effect on Inpatient Mortality: hospital-level exposure

	Dependent variable is inpatient mortality	
	Baseline	Hospital-level exposure
	(1)	(2)
Exposure × short run	0.0087	0.0064
	[0.0057]	[0.0076]
Exposure × long run	0.0088	0.0046
	[0.0054]	[0.0072]
Pre-event control mean	0.128	0.128
Risk-adjustment	✓	✓
Basic controls	✓	✓
Observations	2,790,873	2,790,873

Notes. This table presents estimates based on the model (3), with the sample restricted to at-risk patients. Columns (1) reports our baseline results (the same from Table 6). Columns (2) re-runs the basic specification but using our hospital-level measure of exposure to the deceased. We compute risk-adjusted mortality rates using the two-step procedure described in Section 3.4. At-risk patients are defined as those in the top quartile of the predicted mortality distribution. The observations are weighted by cell size to account for differences in the precision with which the cell means are estimated. Standard errors are clustered at the hospital level.

* $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

Table B.8: Assessing External Validity

	Dependent variable is:					
	Log number of patients			Inpatient mortality		
	Baseline	Observable adjusted	Bounds $\Phi \in [1,2]$	Baseline	Observable adjusted	Bounds $\Phi \in [1,2]$
	(1)	(2)	(3)	(4)	(5)	(6)
Exposure × short run	-0.0926	-0.1175	[-.1425, -.1175]	0.0087	0.0083	[.0079, .0083]
	[0.0202]***	[0.0202]***		[0.0057]	[0.0057]	
Exposure × long run	-0.1174	-0.0986	[-.0986, -.0797]	0.0088	0.0105	[.0105, .0123]
	[0.0207]***	[0.0207]***		[0.0054]	[0.0054]*	
Observations	6,500,630	6,500,630		2,790,873	2,790,873	

Notes. This table assesses the external validity of our main findings, following the approach developed by Andrews and Oster (2019). They propose a procedure that adjusts the point estimates by accounting for heterogeneity in treatment effects across observables influencing participation in the design and provides bounds under different assumptions on the relative importance of unobservables. Columns (1) and (4) report our baseline results from model (3) for two key outcomes: the log number of patients in a given area and the risk-adjusted inpatient mortality rate (restricted to the at-risk subsample), respectively. Columns (2) and (5) present the results adjusted for heterogeneity in treatment effects across the characteristics of the deceased physicians from Table 1, excluding experience (which is nearly perfectly correlated with age). Columns (3) and (6) provide bounds for these estimates when we assume that sample selection is also based on unobservables and that the importance of unobservables relative to observables in driving selection ranges from one to two times.

Table B.9: Effect on Inpatient Mortality:
Post-period Decomposition

	Dependent variable is patient mortality:	
	All new entrants in post-period (Baseline) (1)	Only early entrants in post-period (2)
Exposure × short run	0.035	0.040
	[0.013]***	[0.014]***
Exposure × long run	0.002	-0.006
	[0.009]	[0.016]
Pre-event control mean	0.128	0.128
Basic controls	✓	✓
Observations	3,973,814	3,973,814

Notes. This table presents estimates based on a reduced version of model (2), in which the full set of quarter-since-event indicators is replaced by two indicators: one for the two quarters immediately following the event (*short run*) and another for the period from the third quarter onward (*long run*). The dependent variable is the risk-adjusted inpatient mortality rate in a disease area among patients in the at-risk subsample. We compute risk-adjusted mortality rates using the two-step procedure described in Section 3.4. At-risk patients are defined as those in the top quartile of the predicted mortality distribution. In all columns, the dependent variable for the quarter of the event or the preceding quarters is computed from all at-risk patients. In column 1, the dependent variable in the post-event observations is computed from patients attended by physicians who affiliated to the hospital in the post-event period. In column 2, the dependent variable in the post-event observations is computed from patients attended by physicians who affiliated to the hospital in time event 0 or 1. All the observations are weighted by cell size to account for differences in the precision with which the cell means are estimated. Standard errors are clustered at the hospital level. * p < 0.1;** p < 0.05;*** p < 0.01.

Table B.10: Robust Inference

	Dependent variable is:			
	Log (number of FTE physicians on service) (1)	Share of patients attended by new entrants (2)	log(number of patients) (3)	Inpatient mortality (4)
Exposure × short run	-0.0573	0.0835	-0.0926	0.0087
baseline: clustering on hospital	[0.0604]	[0.0156]***	[0.0202]***	[0.0057]
clustering on pair matching	[0.0625]	[0.0177]***	[0.0201]***	[0.0057]
effective degrees of freedom adjustment	[0.0649]	[0.0168]***	[0.0251]***	[0.0061]
Exposure × long run	-0.0308	0.0115	-0.1174	0.0088
baseline: clustering on hospital	[0.0622]	[0.0062]*	[0.0207]***	[0.0054]
clustering on pair matching	[0.0650]	[0.0063]*	[0.0198]***	[0.0055]
effective degrees of freedom adjustment	[0.0668]	[0.0066]*	[0.0257]***	[0.0058]

Notes. This table presents estimates based on the model (3) under different inference procedures. Our baseline specification uses standard errors clustered at the hospital level. The table additionally reports the results from specifications that employ clustering by matching pair, and clustering by hospital (as in the baseline specification) with the degree-of-freedom corrections suggested by Young (2016). Column (1) investigates the log number of distinct FTE physicians providing care in a given disease area. Column (2) considers the patient share of new entreats in a disease area in the quarter they start operations in the hospital. Column (3) looks at the log number of patients in a disease area. Column (4) analyzes the risk-adjusted inpatient mortality rate in a disease area among patients in the at-risk subsample. We compute risk-adjusted mortality rates using the two-step procedure described in Section 3.4. At-risk patients are defined as those in the top quartile of the predicted mortality distribution. The observations in column (4) are weighted by cell size to account for differences in the precision with which the cell means are estimated.

* $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

Table B.11: Robustness to Sample Restrictions

	Dependent variable is:			
	Log (number of FTE physicians on service) (1)	Share of patients attended by new entrants (2)	log(number of patients) (3)	Inpatient mortality (4)
Panel A: *Baseline specification*				
Exposure × short run	-0.0573	0.0835	-0.0926	0.0087
	[0.0604]	[0.0156]***	[0.0202]***	[0.0057]
Exposure × long run	-0.0308	0.0115	-0.1174	0.0088
	[0.0622]	[0.0062]*	[0.0207]***	[0.0054]
Panel B: *Exclude COVID period*				
Exposure × short run	-0.0640	0.1019	-0.0957	0.0038
	[0.0751]	[0.0192]***	[0.0226]***	[0.0062]
Exposure × long run	-0.0660	0.0143	-0.1075	0.0084
	[0.0693]	[0.0068]**	[0.0232]***	[0.0064]
Panel C: *Exclude hospitals with multiple death events*				
Exposure × short run	-0.0250	0.0943	-0.0834	0.0108
	[0.0730]	[0.0213]***	[0.0229]***	[0.0067]
Exposure × long run	-0.0776	0.0086	-0.1197	0.0047
	[0.0681]	[0.0081]	[0.0242]***	[0.0061]

Notes. This table presents estimates based on the model (3) using different sample restrictions. Panel A shows the baseline results for ease of comparison. Panel B re-runs the baseline specification after excluding data from 2020 onward, while Panel C re-estimates it after discarding hospitals with multiple death events in the data. Column (1) investigates the log number of distinct FTE physicians providing care in a given disease area. Column (2) considers the patient share of new entreats in a disease area in the quarter they start operations in the hospital. Column (3) looks at the log number of patients in a disease area. Column (4) analyzes the risk-adjusted inpatient mortality rate in a disease area among patients in the at-risk subsample. We compute risk-adjusted mortality rates using the two-step procedure described in Section 3.4. At-risk patients are defined as those in the top quartile of the predicted mortality distribution. The observations in column (4) are weighted by cell size to account for differences in the precision with which the cell means are estimated. Standard errors are clustered at the hospital level.
* $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

Table B.12: Pre-Trend Tests

	Pre-periods: 4		Pre-periods: 8	
	F-statistic	*p*-value	*F* -statistic	*p*-value
Outcome variable:	(1)	(2)	(3)	(4)
Log(number of affiliated FTE physicians)	0.8397	0.4718	0.9798	0.4011
Log(number of new affiliated FTE physicians)	0.3495	0.7895	0.2311	0.8748
Incumbents: log(number of patients)	0.7273	0.5355	1.1431	0.3325
Incumbents: number of distinct diagnoses treated	0.4733	0.7009	1.3	0.2456
Incumbents: indicator for non-clinical activities	0.5552	0.6446	0.5908	0.764
Log(number of physicians on service)	0.43	0.7315	1.6055	0.1285
Share of patients attended by new entrants	0.5797	0.6283	1.3113	0.2401
Log(number of patients)	0.0961	0.9622	0.5999	0.7566
Inpatient mortality	0.7124	0.5444	1.2065	0.295

Notes. This table shows the results of pre-trend tests for key outcomes. We estimate pre-treatment coefficients using model (4) for hospital-level outcomes (the log number of total and new affiliated FTE physicians), model (5) for incumbent-level outcomes, and model (2) for the remaining disease-area outcomes. Columns (1) and (2) report the *F*-statistic and *p*-value from *F*-tests under the null hypothesis that all four pre-treatment coefficients in our baseline specifications are jointly zero. Columns (3) and (4) present the *F*-statistic and *p*-value from analogous tests under extended models that includes eight pre-treatment periods.

C Data Sources and Construction

We assemble a new dataset combining information from several administrative registries managed by the Ministry of Health, Regional Medical Councils, and National Registry Office. The final dataset contains detailed information on physicians' educational background, professional roles, and patient interactions. Our analysis focuses on the 2008-2024 period, a choice determined on the front end by the availability of physician data that can be linked to patients. To reduce noise in the data, we choose a quarter as the time unit of reference for the analysis.

Physician-patient data. We have access to administrative data covering all hospital admissions within the entire public health system —the vast majority of hospital care in the country. Importantly, we observe the identity of the main physician responsible for the care of a patient during their hospital stay. These physicians make key decisions regarding treatment, and manage the patient's overall care. Unique tax identifiers allow us to track physicians over time across patients. The data encompass approximately 189.2 million inpatient admissions between 2008 and 2024. At the individual patient level, we observe a set of clinical characteristics such as the cause of admission (as classified by the International Classification of Diseases, Tenth Revision, or ICD-10 code), type of admission, date of discharge, and date of death for those who passed away in the hospital. The data also include basic demographic characteristics such as date of birth, gender, race, date of birth, and place of residence at the zipcode level. Finally, there is detailed information on all medical procedures provided to each patient.

Physician background data. We obtain information on a physician's professional characteristics using the human resources module of the National Registry of Health Establishments. It is a database collected monthly and covers all professionals affiliated to health establishments. A major strength of these data is their universal and high-frequency nature. If a professional joins or leaves the staff, the health facility is required to report this change in the system by the following month. This allows us to examine sharp changes in a hospital's labor supply, even within intervals shorter than one year. Key information available in these data includes a physician's occupations, hours worked, and the identifier codes of all the establishments for which they provide any inpatient or outpatient services. Data on educational background come from a national registry of all medical personnel in the country maintained by the Federal Council of Medicine (CFM, by its acronym in Portuguese). The primary purpose of the CFM registry is to ensure compliance with regulatory standards for medical practice and to maintain accurate and up-to-date records of physicians. The database contains information on physicians' name, municipality of birth, date of birth, medical schools attended and the beginning and conclusion dates of physicians' training. With these data, we are able to infer a physician's date of birth, experience since registration with the CFM (the start of their medical practice), and any medical

specialty completed.

Physician deaths. The comprehensive registry provided by the CFM includes information on the vital status of each physician, indicating whether they are alive or deceased. Given the universal coverage and periodic updates of the CFM data, measurement error regarding a physician's mortality status is unlikely to be a major issue. However, a drawback of this dataset is that it does not record the exact date of death for any physician. Moreover, while the data are updated periodically, previous records become inaccessible once new updates are made. As a result, it is impossible to determine the precise date of death for physicians using these data alone. To overcome this limitation, we link the CFM data to the National Registry Office, which provides death certificates with critical details of deceased individuals —including the exact date of death. The laws governing the collection of the death certificates are national and no burial can be performed without one, ensuring a high level of accuracy in the data. We match both sources of data using unique tax identifiers.

Hospital-level data. We generated longitudinal data on hospitals using the health facility and infrastructure modules of the National Register of Health Establishments. These modules contains basic information about the characteristics of all health facilities, including number of beds, teaching status, type of services, physical resources available, and geographic location. We use these data to identify a similar control pair for each hospital experiencing the death of a physician, as described below.

References

Bavishi, Chirag, Sripal Bangalore, Dipen Patel, Saurav Chatterjee, Vrinda Trivedi, and Jacqueline E Tamis-Holland, "Short and long-term mortality in women and men undergoing primary angioplasty: A comprehensive meta-analysis," *International Journal of Cardiology*, 2015, *198*, 123–130.