# Allocating Talent in Organizations: Evidence from Quasi-Random Exits of Physicians*

Danyelle Branco Bladimir Carrillo Luiz Felipe Fontes Leticia Nunes

April 27, 2026

**Abstract**

How well can organizations allocate tasks when individual performance is only imperfectly observed? We develop a simple framework that exploits quasi-random personnel exits to infer how much organizations attend to differences in worker quality when assigning tasks. We apply this framework to health care, a setting with life-and-death stakes where quality is particularly difficult to assess. Using rich administrative data from Brazil, we show that hospitals react to the unexpected death of higher-performing physicians by shifting patients toward higher-quality physicians. This adjustment is sharply asymmetric: losses from high-performer exits are largely offset, whereas exits of low performers lead to nearly random reassignment. Quantitatively, targeted reassignment prevents a 20 percent increase in mortality that would otherwise arise under random reassignment after the exit of a top physician. The pattern is driven almost entirely by shifts among incumbent physicians rather than external hiring, becomes stronger with the deceased physician's tenure, and is most pronounced in hospitals with formal oversight structures and labor slack. These findings suggest that organizations attend to differences in quality, but that organizational capacity is crucial for responding effectively to such differences.

**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: 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). This research is funded by the Becker Friedman Institute at the University of Chicago. We also gratefully acknowledge financial support from the Brazilian National Council for Scientific and Technological Development (CNPq —# 04/2021). We thank Luis Araujo, Chad Syverson, Emanuel Ornelas, Esteban Rossi-Hansberg, Fabio Miessi, Jan Stuhler, Jose Tessada, Joseph Doyle, Lucas Finamor, Rudi Rocha, Rogerio Santarrosa, and Vladimir Ponczek, as well as participants at various seminars, 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

A central challenge for any organization is how to allocate work efficiently across workers who differ in productive ability (Holmstrom and Milgrom, 1991; Holmström, 2017). This is a difficult task because firms typically observe aggregate outcomes that reflect the contribution of multiple workers and varying environmental conditions, making it unclear which individuals drive performance. If organizations are unable to reliably discern an individual's true value-added, they may systematically misallocate labor and face persistent performance losses when a top performer leaves. Yet direct empirical evidence on whether organizations can overcome these frictions and target their best workers remains scarce.

Credibly answering this question has proven difficult precisely because observed outcomes and task allocation may reflect a variety of constraints and confounding forces beyond individual performance. In this paper, we show that quasi-random personnel exits, combined with observational measures of individual performance, can be used to evaluate an organization's ability-targeting behavior. The idea is simple. Consider an organization whose performance is given by the weighted sum of each worker's intrinsic productivity (i.e., $\sum \alpha_j \mu_j$), such as defect-free rates for units assembled, revenue per client assignment in a salesforce, win rates in court cases, or a teacher's value-added to test scores. When someone leaves, what happens to weighted-mean performance depends on how good that worker was, their task share, and who takes over their tasks. Hence, comparing the realized change in expected performance to that implied by random reallocation provides a simple test of how "blind" organizations are to differences in worker quality. As discussed in Section 3, this relationship can be parsimoniously summarized by a single response parameter that tends toward -1 under random reallocation and toward 0 under perfect quality targeting. We show that when the exit event is plausibly exogenous, one can obtain consistent estimates of this parameter using observational measures of performance even if they are imperfect forecast of true quality. This logic is related to recent work using quasi-experimental changes in choice sets to validate observational measures of health-care quality (Abaluck et al., 2021).

We apply this framework in the context of health care, where performance differences have immediate life-and-death consequences, and where both task assignments and outcomes are measured with unusual precision at population scale. Our analysis focuses on physicians in charge of inpatient care, the central decision-makers responsible for diagnosis, treatment decisions, and the overall management of patients during a hospital stay. Physicians differ in their clinical effectiveness, and their caseloads are shaped by a combination of formal and informal allocation practices that can result in substantial variation in the number of patients they see.[1] This setting is often cited as a canonical example in which quality is especially difficult to infer (Arrow, 1963). When a patient

[1] Differences in physicians' clinical effectiveness can arise for a variety of reasons, including variation in formal training, accumulated experience, diagnostic skill, procedural proficiency, and the ability to synthesize complex clinical information under time pressure. Even holding constant observable credentials such as specialty or years of practice, physicians may differ in how they tailor treatment to patient-specific signals, anticipate complications, and coordinate care with other providers. A growing empirical literature documents persistent heterogeneity in physician performance across a range of settings and outcomes, consistent with meaningful and stable differences in clinical effectiveness (Ginja et al., 2025; Chan and Chen, 2022).

dies, is the outcome attributable to the physician's clinical decisions, downstream execution by other professionals, or simply to the severity of the case the physician happened to receive? Using rich administrative data from the universal health system in Brazil, we find strong evidence that even in this highly noisy environment providers attend to differences in clinical performance and manage to redirect patient caseload selectively to mitigate losses in care quality.

Our starting point is to document that there is a great deal of variation in observational value-added across physicians practicing in the same area of care, measured as patient survival rates net of differences in case mix, hospital fixed effects, and sampling error. As documented in Section 2, the standard deviation in the observational survival is approximately 7.7 percentage points, equivalent to three quarters of raw survival gap between patients over 60 and under 40 years old. For each area of care within a hospital, we then aggregate these estimates into a patient-weighted average, yielding our observational measure of expected performance.[2] Note that since a physician's observational value-added is time-invariant by construction, changes in expected performance reflect shifts in which physicians see the patients or a change in who is on the medical staff—not the type of patient.[3]

We estimate the parameter of interest using a difference-in-differences design that exploits unexpected physician deaths—defined as those occurring during active clinical practice—as a source of variation in the timing and magnitude of the quality shock implied by the observed patterns of survival. While variation in survival may reflect unobserved determinants of case allocation as well as the unforecastable component of physician survival effects, this approach yields consistent estimates of the parameter of interest as long as such unobservables evolve similarly across more- and less-affected areas. Put simply, identification requires that physician deaths are not systematically timed with concurrent changes in either how hospitals rank their physician pool or how cases are allocated conditional on those rankings. In principle, it is unclear why this would not be the case given our focus on deaths that are, by definition, sudden and outside the hospital's control. We provide empirical evidence consistent with these conditions in the data.

Since the timing of physician deaths differs across hospitals, we estimate a stacked difference-in-differences to avoid potential biases in two-way fixed effects models (Wing et al., 2024; Butters et al., 2022; Cengiz et al., 2019).[4] The key findings of this paper are presented in Section 4. We find that following the death event, areas with a larger expected loss in care quality induce stronger

[2] An area of care corresponds to a clinically coherent service line within the hospital, defined by grouping admissions according to the primary diagnosis subchapter of the 10-ICD classification. This level of aggregation reflects the fact that hospitals are organized around semi-specialized units that treat similar patient populations using related clinical technologies and physician expertise. Aggregating performance within areas of care allows us to compare physicians facing comparable diagnostic tasks and to interpret reallocations as meaningful organizational responses.

[3] To see this more clearly, let $\mu_j$ denote physician $j$'s time-invariant intrinsic clinical ability and $\alpha_{jt}$ the share of patients assigned to $j$ in period $t$. Expected performance in period $t$ is defined as $\bar{\mu}_t = \sum_j \alpha_{jt} \mu_j$. Because $\mu_j$ does not vary over time by construction, any change in $\bar{\mu}_t$ across periods reflects only changes in patient allocation ($\alpha_{jt}$) or in the set of active physicians. This holds true regardless of how $\mu_j$ is measured or estimated.

[4] In the presence of staggered treatment adoption and heterogeneous effects, two-way fixed effects estimators may be biased due to comparisons that use already-treated units as controls. For formal discussions and solutions, see De Chaisemartin and d'Haultfoeuille (2020), Callaway and Sant'Anna (2021), Sun and Abraham (2021), and Goodman-Bacon (2021).

reallocation of patients toward similarly performing physicians. As a result, the overall decline is much smaller than what the quality shock alone would imply. We estimate a statistically significant response parameter of -0.38, indicating that approximately two thirds of the shock is absorbed through endogenous reallocation. This average effect masks substantial heterogeneity with respect to the direction of the shock. The point estimate is nearly zero and statistically insignificant when the deceased physician's observational survival exceeds the peer average, but substantially more negative when it falls below it. This asymmetry is consistent with the intuition that hospitals aim to preserve quality but reallocation toward higher-quality physicians is costly. When hospitals face no immediate threat to quality, they have limited incentives to incur the cost of redirecting patients toward its most overburdened top performers and hence the reallocation response becomes more diffuse.

Estimates from an event-study specification supports the validity of our statistical approach by showing that areas exposed to larger and smaller turnover-induced quality shocks exhibit similar trends prior to the death event. Moreover, the results are robust to the inclusion of a comprehensive set of flexible controls, including interactions between pre-event characteristics and time effects as well as specifications that account for potential spillovers within the hospital. As a falsification exercise, we construct analogous predictors of a physician's performance based on observable characteristics such as experience and specialization but residualize them with respect to survival outcomes. These predictors capture dimensions of physician skill that are unrelated to observed survival differences by construction, and thus serve as a proxy for the forecast-error component in using observational survival to predict true physician quality.[5] There is no evidence of systematic changes in overall performance based on these placebo predictors, suggesting that our estimates are not confounded by differential changes in the informational content of observational survival around physician deaths.

Our exploration of mechanisms in Section 5 points to the joint role of internal information and organizational capacity. First, the adjustment operates almost entirely through internal reallocation among incumbent physicians, rather than through external hiring. Second, the strength of the mitigating response increases sharply with the departed physician's tenure, consistent with accumulated learning sharpening managerial beliefs and enabling more effective targeting. Third, we continue to find a similarly large reallocation response even after accounting for experience and specialization. This suggests that hospitals target reallocation based on richer, physician-specific signals of clinical effectiveness rather than relying solely on coarse, salient credentials. Finally, hospitals with more formal monitoring structures and greater labor slack are substantially more effective at mitigating negative quality shocks. Read together, these findings indicate that effective adjustment requires not only information but the organizational capacity to act on it.

Reallocating patients across physicians is not the only possible margin through which hospitals

[5]This falsification exercise is designed to test for differential changes in the use of information about physicians that is orthogonal to observed survival outcomes. It does not test for omitted-variable bias in survival itself, but rather assesses whether hospitals adjust patient allocation based on non-survival-based signals of physician quality around the death event.

may mitigate the effects of losing a top physician. Although expanding capital capacity is limited in the short run, hospitals may adjust by changing both the number and the composition of admitted patients. In practice, however, we find no evidence of changes in the number of patients admitted or in the selection of healthier or sicker patients. This absence of adjustment is institutionally plausible in our setting, where hospital activity is closely monitored by local authorities and restricting admissions could be interpreted as limiting access to care unless the deceased physician accounted for an unusually large share of the service's effective capacity.

Our finding that hospitals tend to replace high-quality physicians with other high-quality physicians naturally raises an important question: how much does this reallocation matter for patient mortality outcomes? The answer depends on the extent to which the observed variation in patient survival reliably maps into differences in physicians' true clinical ability. This predictive reliability can be estimated from a regression of actual patient survival on expected survival given the physician mix, using the turnover-induced quality shock as an instrumental variable. Our estimates indicate that observational survival is a highly reliable predictor of true survival effects, with a reliability coefficient of approximately one.[6] This implies that the sizable reallocation response averts a large share of deaths. A back-of-the-envelope calculation indicates that this endogenous reorganization avoids roughly a 20 percent increase in mortality that would otherwise occur following a one–standard deviation increase in the severity of a negative quality shock. This effect is comparable in scale to the impact of a major increase in hospital spending (Doyle et al., 2015).

**Related Literature.** The findings of this paper contribute to a large literature on firm behavior and organizations. When output is noisy, multidimensional, and jointly produced, theory emphasizes that performance-based responses are potentially counterproductive (Holmstrom, 1982; Holmstrom and Milgrom, 1991; Lazear, 1986; Baker, 1992). A natural implication is that what matters for organizations is not only whether performance signals exist, but whether they are sufficiently actionable to guide core production decisions rather than merely compensation or retention. Empirically, much of what we know about how firms learn about worker ability comes from indirect tests using wage and employment trajectories (Altonji and Pierret, 2001; Lange, 2007; Kahn and Lange, 2014; Schönberg, 2007). We add to the literature by providing more direct evidence on how organizations use noisy performance information to reorganize production itself. In doing so, we connect learning about heterogeneous ability to a growing literature that views task assignment as a first-order driver of firm performance (Garicano, 2000; Garicano and Rossi-Hansberg, 2006; Bandiera et al., 2007; Hjort, 2014; Adhvaryu et al., 2022). A major implication of our findings is that organizations can cope with highly disruptive events by reallocating tasks, even without structurally adjusting technology or factor inputs. By focusing on death-induced exits as a source of variation, our findings also offer

[6]The reliability coefficient is a reduced-form parameter that summarizes how informative observed survival is about true clinical effectiveness in our setting. A value close to one does not imply that observational survival is noise-free or perfectly measured. Individual realizations may deviate from this relationship due to idiosyncratic shocks, sampling variation, or measurement error. More generally, the forecast equation characterizes the best linear projection of true quality on the empirical signal. Deviations from the projection are mean-zero and cancel out in expectation, so that the relationship between observed survival and true clinical effectiveness holds only on average.

a new perspective on how and when labor turnover affects performance (Sauvagnat and Schivardi, 2024; Jäger et al., 2025).

Our findings also relate to a growing literature estimating value added in health care, including work documenting large and persistent differences in physician value added (Currie et al., 2016; Chan and Chen, 2022; Ginja et al., 2025), as well as sizable heterogeneity in the quality of helthcare institutions and health insurance plans (Abaluck et al., 2021; Doyle et al., 2015; Hull, 2018; Einav et al., 2025). While previous work use quasi-experimental designs to validate observational measures of care quality, we ask a distinct question: how do organizations allocate services given that providers differ in quality? A key implication of our results is that organizational responses are an important margin through which individual performance differences affect aggregate outcomes. This matters for the use of value-added metrics in evaluation and policy, as the consequences of replacing individual physicians need not map one-for-one into changes in outcomes when organizations buffer losses (or fail to exploit potential gains). This is not unique to health care. In the well-developed literature on value-added in education (Chetty et al., 2014a,b; Bau and Das, 2020), for example, the welfare interpretation of widely debated policies, such as "deselecting" low-quality teachers, may similarly be confounded by how principals endogenously reallocate students.

# 2 Setting and Data

This section describes the institutional setting and the data sources, samples, and variable definitions.

## 2.1 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.[7]

**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

[7]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 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.

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.

### 2.2 Data

**Overview.** We assemble a new dataset combining information from several administrative registries managed by the Ministry of Health, Regional Medical Council, and National Registry Office, described in detail in Appendix B. The data cover all hospital admissions within the entire Brazil's Unified Health System between 2008 and 2024, including granular information on patients and providers.[8] 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. There are approximately 189.2 million inpatient admissions between 2008 and 2024.

For each hospital admission, we observe the identity of the main physician responsible for the patient's care. 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.

[8]Our analysis focuses on the 2008-2024 period. This choice was determined on the front end by the availability of physician data that can be linked to patients.

**Turnover After Death.** We follow 315,902 physicians, of whom 8,198 die during the period. Many of these deaths are unlikely to have been sudden from the perspective of colleagues. Physicians with serious or terminal conditions often experience prolonged absences and intensive treatment, and may reduce their workload or resign well before death to manage their health. 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 1,021 deaths, or approximately 12 percent of all deaths.[9]

**Alive versus Deceased Physicians.** Appendix Figure A.2 shows that alive and deceased physicians are very similar in terms of the observational value-added measure described below. A formal test of equality in mean value-added yields a $p$-value above 0.8, providing no evidence of a statistically significant difference between the two groups. Nevertheless, deceased physicians are, on average, more experienced and more likely to be male than their surviving counterparts (Appendix Table A.1). This pattern is largely mechanical. The main unexpected causes of death, such as heart attacks, heart failure, and strokes, are more prevalent at older ages and among men. As a result, the differences between deceased and alive physicians primarily reflect that they are at different stages of the life cycle, rather than systematic differences in underlying performance.

### 2.3 Sample and Matched Controls

There are 7734 hospitals in the period of analysis. Of these hospitals, 1159 had at least one physician who passed away on duty.[10] 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 A.2). 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

[9] 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. Hence, we focus on our higher-powered and more parsimonious strategy.

[10] 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. 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.

of emergency department, 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 A.2 documents that this approach essentially eliminates differences between treated and untreated hospitals.

### 2.4 Areas of Care and Observational Value-Added

Our analysis exploits variation across areas of care within hospitals. The idea of an area of care is motivated by the fact that hospitals operate through semi-autonomous clinical service lines, each with its own patient population, and technologies, much like the specialized production lines of a factory. We define areas of care based on patients' primary diagnoses, grouping admissions according to the corresponding subchapter of the ICD classification. Examples of subchapters that define distinct areas of care in our analysis include ischemic heart diseases and cerebrovascular diseases.

With this conceptualization in place, we now estimate differences in observed survival rates across physicians operating within the same area of care, accounting for differences in baseline case mix as well as hospital and time effects. We do so by estimating the following patient-level regression:

$$\text{Survival}_{it} = \theta_{js} + \mathbf{X}'_{it}\Theta + \varepsilon_{it},$$

where $Survival_i$ is an indicator for whether the patient survived the hospitalization, and $\mathbf{X}_{it}$ is a vector of pre-determined covariates. We include patient demographic and clinical characteristics, including age, gender, race, residential zipcode, history of recent hospital admissions, and a complete set of comorbidity indicators. We also control for hospital and quarter-by-year fixed effects to absorb time-invariant differences across hospitals and common shocks affecting all hospitals in a given period. The parameter of interest is $\theta_{js}$, which represents risk-adjusted physician survival rates specific to each area of care. We normalize $\theta_{js}$ to have a mean of zero within each area of care to facilitate interpretation and within area comparability. To account for differences due to sampling error, we construct an empirical Bayes estimator that adjusts the estimated physician effects according to the standard shrinkage formula, as described in Appendix C.[11]

We refer to $\theta_{js}$ as *observational value-added* or related variations of this term. It is described as observational because it reflects differences in observed outcomes based on available data, not necessarily differences in causal clinical effectiveness across physicians. Our approach described below does not require that the observational value-added be an unbiased predictor of true value-

[11] An empirical Bayes estimator corrects for the fact that some estimated physician effects are measured with greater precision than others. Intuitively, estimates based on few patient observations are noisier and more likely to overstate true differences in performance, while those based on many patients are more reliable. The empirical Bayes adjustment applies a standard shrinkage correction that pulls each physician's estimated effect toward the overall mean, with the degree of shrinkage inversely related to the precision of the estimate (Morris, 1983). Empirical Bayes estimators of this type are widely used in the teacher value-added literature to correct for measurement error in estimated teacher performance (e.g., Kane and Staiger, 2008; Chetty et al., 2014c).

added, but it must provide a reasonably strong signal of true quality differences. In Section 4.5, we will provide evidence that the observational survival is indeed a highly reliable predictor of true causal effects. Panel A of Figure 1 documents that there is a great deal of variation in this measure. The standard deviation in the observational survival is approximately 7.7 percentage points, equivalent to three quarters of raw survival gap between patients over 60 and under 40 years old. As one can infer from the figure, the empirical Bayes adjustment makes little practical difference.[12]

# 3 Conceptual Framework and Design

This section presents our empirical approach to quantifying the impact of the quality shock generated by physician deaths. We begin by setting a simple model of health production to motivate our identification strategy.

## 3.1 Physicians and the Production of Quality

We model a hospital service line as a production unit that transforms clinical inputs and physician expertise into patient health output.[13] Following Chandra et al. (2016), the core output, health production $Y_t$, is defined as the exponential of in-patient survival, with the underlying production technology:

$$Y_t = A_t \left[ \prod_{\omega} Z_{\omega t}^{\phi_w} \right] \left[ \prod_{k} R_{kt}^{\rho_k} \right], \tag{1}$$

where $Z_{\omega t}$ denotes a set of case-mix adjusters, such as patient age, sex, and comorbidities, while $R_{kt}$ represents a vector of hospital-wide resource endowments such as the number of nursing staff, installed capital, and available equipment. The term $A_t$ functions as a Hicks-neutral efficiency shifter, capturing the clinical value-added of the physician staff. It determines how effectively a given combination of available resources and patient severity is translated into health output.

Patients are assigned to physicians who differ in their clinical value-added, with each physician $j$ having a skill parameter $\mu_j$ and contributing to aggregate output in proportion to their patient share $\alpha_{jt}$. The log-productivity, $\bar{\mu}_t = \ln A_t$, is therefore the patient-share-weighted average of physicians' individual clinical effectiveness:

$$\bar{\mu}_t \equiv \sum_{j=0}^{N} \alpha_{jt} \mu_j.$$

While a physician's clinical ability is time-invariant, overall staff effectiveness can vary over time because of either shifts in the distribution of patients across physicians or because physicians enter

[12] In our data, the median shrinkage factor exceeds 0.95, meaning that most physician-specific estimates are adjusted by less than 5 percent toward the overall mean.

[13] To lighten notation, we omit subscripts for the area of care and hospital.

and exit the staff. When a physician $j = 0$ leaves, the expected change in average quality is:

$$\begin{aligned}\Delta\bar{\mu} &\equiv \bar{\mu}_{\text{post}} - \bar{\mu}_{\text{pre}} \\ &= \left[\sum_{j=1}^{N} \alpha_{j\text{post}}\mu_j + \alpha_k\mu_k\right] - \left[\sum_{j=1}^{N} \alpha_{j\text{pre}}\mu_j + \alpha_{0\text{pre}}\mu_0\right]\end{aligned}$$

The term $\alpha_k\mu_k$ denotes the contribution of a potential new hire $k$ who enters the team in the post-turnover period. When no external hire takes place, this term mechanically equals zero ($\alpha_k = 0$).

Intuitively, the change in overall performance depends not only on the quality shock induced by the departing physician but also on the hospital's active response to that shock. To see this formally, consider the counterfactual in which the departing physician's patients are redistributed randomly among the remaining incumbents. In that case, the pre-period quality of incumbents can be written as:

$$\bar{\mu}_{\text{pre}}^{-0} = \sum_{j=1}^{N} \frac{\alpha_{j\text{pre}}}{1 - \alpha_{0\text{pre}}}\mu_j$$

Adding and subtracting this counterfactual term from $\Delta\bar{\mu}$ yields:

$$\Delta\bar{\mu} = \underbrace{\sum_{j=1}^{N}\left(\alpha_{j\text{post}} - \frac{\alpha_{j\text{pre}}}{1 - \alpha_{0\text{pre}}}\right)\mu_j}_{\text{caseload reallocation}} + \underbrace{\alpha_k\mu_k}_{\text{external replacement}} - \underbrace{\alpha_{0\text{pre}}\left(\mu_0 - \bar{\mu}_{\text{pre}}^{-0}\right)}_{\text{quality shock}} \tag{2}$$

The first two components on the right-hand of this expression capture the hospital's active response to the physician's departure. The *caseload reallocation* term reflects how the caseload left by the departing physician is redistributed among incumbent physicians relative to the random reallocation benchmark. The overall impact of the shock will be attenuated to the extent that the hospital reallocates cases to physicians whose clinical quality closely matches that of the departing physician. The *external replacement* term represents the change in performance due to a new physician hired externally, with its magnitude reflecting both the replacement's relative quality and the share of the departing physician's caseload that is absorbed. The final component captures the *quality shock* arising from the departing physician's deviation from the incumbents' average clinical value-added.

Multiplying and dividing equation (2) by the magnitude of the quality shock, $\alpha_{0\text{pre}}(\mu_0 - \bar{\mu}_{\text{pre}}^{-0})$, we obtain the following parsimonious expression:

$$\Delta\bar{\mu} = \left(\gamma - 1\right) \cdot \alpha_{0\text{pre}}\left(\mu_0 - \bar{\mu}_{\text{pre}}^{-0}\right) \tag{3}$$

Here, $\gamma$ captures the hospital's overall endogenous response, scaled by the magnitude of the quality shock. When $\gamma = 0$, the hospital takes no mitigating action, and performance changes in proportion to the quality shock. At the other extreme, when $\gamma = 1$, the hospital's response fully offsets the turnover-induced change in clinical quality, leaving overall performance unchanged. Where $\gamma$ falls between these two extremes identifies the relative importance of the hospital's response to the quality shock.[14]

[14] In theory, $\gamma$ is not necessarily bounded between 0 and 1. Values below 0 suggest that the hospital's response exacerbates the performance loss from a negative quality shock, such as by reallocating disproportionately patients to incum-

### 3.2 Mapping to Estimation

**Estimating Equation.** Equation (3) shows that, for a given hospital–area, the change in average clinical quality before and after a physician's death is proportional to the size of the quality shock, scaled by the hospital's endogenous response parameter $\gamma$. In the data, we observe a panel of hospital–area cells over time, together with the timing and intensity of these turnover-induced shocks. This structure naturally lends itself to a difference-in-differences design that compares areas that are more versus less exposed to a given death, before and after the event. Let $\tau$ denote the period in which a physician death occurs in hospital $h$ and area of care $s$. The empirical version of equation (3) takes the form:

$$\bar{\mu}_{sht} = \beta \cdot \alpha_{0sh,\text{pre}} \left( \mu_{0s} - \bar{\mu}^{-0}_{sh,\text{pre}} \right) \cdot \mathbb{1}\{t \geq \tau\} + \mathbf{X}'_{sht}\Omega + \zeta_{st} + \psi_{sh} + \epsilon_{sht} \tag{4}$$

The treatment variable of interest is given by the interaction between the post-death indicator, $\mathbb{1}\{t \geq \tau\}$, and the intensity of the quality shock, $\alpha_{0sh,\text{pre}}(\mu_{0s} - \bar{\mu}^{-0}_{sh,\text{pre}})$. It varies across hospital-areas based on both the timing of physician deaths and the pre-existing composition of physician expertise. Matched control hospitals represent a particular case in which there is no quality shock and $\alpha_{0sh,pre} = 0$. Area-time fixed effects, $\zeta_{st}$, absorb common shocks across areas, while hospital-area fixed effects, $\psi_{sh}$, account for time-invariant differences in performance across hospital-areas. The vector $\mathbf{X}_{sht}$ includes interactions between time fixed effects and the deceased physician's baseline patient share to account for any direct impact of the turnover event itself.

The parameters of equation (4) are not directly estimable because a physician's latent quality $\mu_{js}$ is unobserved. Instead, we observe the empirical Bayes estimate $\theta_{js}$, which is a noisy signal of $\mu_{js}$ constructed from risk-adjusted survival patterns across physicians. We model latent quality as the best linear projection of true performance on its empirical signal:

$$\mu_{js} = \lambda\theta_{js} + \eta_{js} \tag{5}$$

The disturbance term $\eta_{js}$ is a forecast-error component of true quality, orthogonal to $\theta_{js}$. The forecast coefficient $\lambda$ quantifies the reliability of the observed measure but carries no structural or behavioral interpretation. It is shaped by the data-generating process governing the joint distribution of true physician effects and observational differences in survival rates in a service area. If the observed survival rate is an average unbiased predictor of physician quality, then $\lambda = 1$, and the expected difference in observed survival between two randomly drawn physicians closely mirrors the difference in their underlying quality. Values of $\lambda$ above and below one imply that observed survival rates understate and overestimate differences in true quality, respectively.[15]

Substituting equation (5) into the unobserved-quality specification in (4) yields an estimable expression in terms of the empirical signal. After substituting and reparameterizing, we arrive at

bents with lower quality. Conversely, values above 1 would imply an overcompensating response, where the hospital's reorganization or hiring more than offsets the initial loss in physician quality.

[15] If differences in observed survival rates primarily reflect that patients with worse health conditions are allocated to certain physicians, then variation in the empirical signal will overstate true differences in physician quality, yielding $\lambda < 1$. Alternatively, if selection were such that better physicians treat more complex cases, observed survival differences would understate true differences in quality, leading to $\lambda > 1$.

our baseline empirical specification:

$$\bar{\theta}_{sht} = \beta \cdot \alpha_{0sh,\text{pre}} \left( \theta_{0s} - \bar{\theta}^{-0}_{sh,\text{pre}} \right) \cdot \mathbb{1}\{t \geq \tau\} + \mathbf{X}'_{sht}\tilde{\Omega} + \tilde{\zeta}_{st} + \tilde{\psi}_{sh} + v_{sht} + \tilde{\epsilon}_{sht}, \tag{6}$$

where

$$v_{sht} = \tfrac{\beta}{\lambda} \cdot \alpha_{0sh,\text{pre}} \left( \eta_{0s} - \bar{\eta}^{-0}_{sh,\text{pre}} \right) \cdot \mathbb{1}\{t \geq \tau\} - \tfrac{1}{\lambda} \bar{\eta}_{sht}.$$

In this specification, the parameter of interest $\beta$ retains the same structural interpretation as in equation (4). What is different is that this reparameterization introduces a scaling factor that depends on the forecast coefficient, effectively dividing all structural terms by $\lambda$, and adds an additional disturbance term, $v_{sht}$, that depends on the residual forecast.

Using the deceased physician's patient share within each service area and hospital, along with their observational value-added, we construct the "treatment" intensity measure of interest, $\alpha_{0sh,\text{pre}} \times (\theta_{0s} - \bar{\theta}^{-0}_{sh,\text{pre}})$. While this measure is based on the observational value-added, we henceforth refer to it as the "quality shock" to ease exposition. Panel B of Figure 1 shows that there is substantial variation in the intensity of the quality shock, with negative shocks occurring about as frequently as positive ones. We next formalize the identifying conditions under which $\beta$ can be consistently estimated.

**Identification.** This model is a difference-in-differences setup with heterogeneity in treatment timing, since the timing of physician deaths differs across hospitals. Our approach to avoid biases due to differences in the timing of death events is a stacked difference-in-differences design (Wing et al., 2024; Butters et al., 2022; Cengiz et al., 2019).[16] We first construct a separate panel dataset for each hospital that experienced a death in quarter $q$ and then append all these sub-samples into a single file to form the final stacked sample. We estimate the parameter of interest using this stacked sample and control for a full set of event fixed effects interacted with hospital-area and area-time fixed effects.

As in any difference-in-differences design, the key identifying assumption is a parallel trends condition. To understand its plausibility, it is important to note that this assumption combines two conditions with distinct economic meaning.

**Condition 1** (Common Trends in Residual Quality)**.** *Conditional on fixed effects and controls, idiosyncratic performance shocks, $\epsilon_{sht}$ are mean–independent of the timing and intensity of the quality shock:*

$$\epsilon_{sht} \perp \alpha_{0sh,pre}(\theta_{0s} - \bar{\theta}^{-0}_{sh,pre}) \cdot \mathbb{1}\{t \geq \tau\} \mid \mathbf{X}_{sht}, \zeta_{st}, \psi_{sh}$$

Condition 1 essentially requires that unobserved determinants of average clinical value-added evolve similarly across areas that are differentially exposed to the quality shock. It could be violated if, for example, hospital managers were concurrently or previously reallocating better physicians into (or out of) more or less exposed areas for reasons unrelated to the death, such as the staggered

[16] It is now well established that conventional estimators based on two-group/two-period before-after comparisons may be biased in the presence of heterogeneous treatment effects. See De Chaisemartin and d'Haultfoeuille (2020), Callaway and Sant'Anna (2021), Sun and Abraham (2021), and Goodman-Bacon (2021).

introduction of a new cardiac program, targeted specialty hiring, a residency rotation change, or a specialty-specific resource squeeze. There is little reason to believe that the death of higher- or lower-quality physicians would systematically coincide with such structural reorganizations.

Of particular concern is the possibility that hospitals anticipate the death of a physician and undertake organizational decisions in advance. Such anticipatory effects are unlikely in our setting because the sample is restricted to physicians who were actively practicing at the time of death, thus excluding those on prolonged medical leave due to chronic or otherwise foreseeable conditions. In any case, we can estimate a flexible event-study version of the model and check directly for antecipatory effects. If more and less exposed areas exhibit similar trends prior to the death and begin to diverge only afterward, this would provide strong evidence in favor of the identifying condition. As we show below, this is indeed what we find in the data.

**Condition 2** (Common Forecast Shocks)**.** *Conditional on fixed effects and controls, the evolution of residual forecast errors is mean-independent of the timing and intensity of the quality shock:*

$$v_{sht} \perp \alpha_{0sh,pre}(\theta_{0s} - \bar{\theta}^{-0}_{sh,pre}) \cdot \mathbb{1}\{t \geq \tau\} \mid \mathbf{X}_{sht}, \zeta_{st}, \psi_{sh}$$

Given that $\theta_{js}$ is orthogonal to $\eta_{js}$ by construction and that we control for any differential trends in baseline shares $\alpha_{0sh,pre}$, condition 2 reduces to requiring independence of the quality shock treatment with respect to $\bar{\eta}_{sht}$.[17]This requirement is similar to the fallback condition introduced by Abaluck et al. (2021) in the context of consumers' health plan choices. In the simplest of terms, it requires that more- and less-exposed areas around the death event do not experience differential shocks to the unobserved component of true quality that is invisible in $\theta_{js}$. By construction, the forecast error represents a relatively stable, slow-moving aspect of physicians' effectiveness and of how hospitals perceive them. This includes factors such as internal reputation, soft skills or other dimensions that hospitals learn over a long period through daily interaction, committee work, and informal feedback. The sudden death of a colleague does not mechanically alter these attributes, nor does it generate new information about them in a systematic way. Any within-physician or within-area change in this component exactly at the death date is likely to be negligible. Even if economically meaningful changes exist, it is unclear why they would be systematically stronger in more exposed areas than in less exposed ones.

It is important to note that our baseline specification is deliberately demanding and already incorporates a rich set of hospital, area, and time controls. Hospital-by-time and area fixed effects absorb broad hospital-wide changes in protocols, staffing, or evaluation practices, as well as smooth area-specific evolutions in unobserved quality. This leaves a relatively narrow source of identifying

[17]To see this, let $\mathcal{F}$ collect the controls and fixed effects, under which $\alpha_{0sh,\text{pre}}$, $\theta_{0s}$, and $\bar{\theta}^{-0}_{sh,\text{pre}}$ are predetermined. Because the forecast equation $\mu_{js} = \lambda\theta_{js} + \eta_{js}$ implies $\mathbb{E}[\theta_{js}\eta_{js}] = 0$ by construction, it follows that the aggregated residual component is mean-independent of any function of $\theta_{js}$, i.e. $\mathbb{E}[\eta_{0s} - \bar{\eta}^{-0}_{sh,\text{pre}} \mid \theta_{0s}, \bar{\theta}^{-0}_{sh,\text{pre}}, \mathcal{F}] = 0$. Under the assumption of exogeneity of the timing of physician deaths, this orthogonality extends to the treatment term, yielding

$$\text{Cov}\Big(\alpha_{0sh,\text{pre}}(\theta_{0s} - \bar{\theta}^{-0}_{sh,\text{pre}})\mathbb{1}\{t \geq \tau\},\ \alpha_{0sh,\text{pre}}(\eta_{0s} - \bar{\eta}^{-0}_{sh,\text{pre}})\mathbb{1}\{t \geq \tau\} \ \Big|\ \mathcal{F}\Big) = 0.$$

variation. A meaningful violation of condition 2 would therefore require finely targeted, within-hospital changes in the unobserved component of quality that: *i*) occur exactly around death events, *ii*) differ sharply between more- and less-exposed areas, and *iii*) are unrelated to $\theta_{js}$ yet aligned with our exposure measure. We view this constellation of requirements as empirically unlikely. While this assumption cannot ultimately be formally tested, we will conduct several falsification tests designed to assess its plausibility. We return to this discussion in Section 4.3.

# 4 Effects of Quality Shocks

## 4.1 Quality Composition and Endogenous Adjustment

Figure 2 reports estimates from an event-study specification that replaces the single post indicator with a full set of event-time dummies (leads and lags) relative to the death quarter, each interacted with the pre-period quality-shock intensity. Coefficients are normalized to zero in the quarter immediately preceding the event. The navy series plots estimates for our key outcome of interest, the patient-weighted average of physician quality. The figure shows that in the quarters leading up to a physician's death, more-exposed and less-exposed areas follow identical trends, with pre-event coefficients close to zero, statistically insignificant, and without any clear visual pattern of improvement or deterioration in physician quality composition. In sharp contrast, there is a discrete decline in expected performance immediately after the death event. The estimated coefficient falls to roughly -0.27 in the event quarter and stabilizes around -0.38 from the following quarter onward. These estimates indicate that hospitals offset approximately two-thirds of the quality loss associated with a physician's death through internal reallocation or new hires. This magnitude suggests a substantial, but incomplete, organizational adjustment in response to the shock.

The lack of pre-existing differential trends is consistent with the validity of our research design. For selection to threaten identification, unobserved factors would need to change at the time of the death in proportion to the intensity of the quality shock. This rules out stories of endogeneity that involve changes in slow-moving factors, such as gradual capacity expansions, phased quality-improvement rollouts, or budget-driven staffing drifts. The fact that the event-time profile displays a sharp drop followed by a flat trajectory suggests that hospitals already possess reliable information about physicians' relative abilities and can immediately reorganize their operations. If hospital managers had to learn about quality only after the death event, one would expect a more gradual and noisy recovery, with coefficients converging slowly over several quarters.

As a sanity check, the maroon series plots estimates for the same event-time dummies using as outcome the patient-weighted average quality among alive physicians, which excludes the deceased from the aggregation in the entire period. In practice, this mechanically removes the direct quality-shock term from the estimation equation. In our framework, this corresponds to estimating $\gamma$ directly rather than $\beta = \gamma - 1$. If our estimates of $\beta$ are not an artifact of the data, the maroon coefficients should trace the mirror image of the navy ones, converging to a value around $+0.64$ in the post-event period. This is precisely what we observe, confirming the internal consistency of

our findings. The only exception is the event quarter, where estimates are mechanically attenuated because some deaths occur late in the quarter and thus generate only partial exposure during that period.

We next assess the sensitivity of the estimated coefficient $\beta$ to alternative specifications, focusing on the average treatment effect to summarize the dynamics more parsimoniously. To avoid contaminating post-treatment estimates with partial exposure, we exclude the quarter of the death event from the estimation sample. Column 1 reports results from the most basic specification, which includes the full set of stacked fixed effects as well as interactions between time fixed effects and the deceased physician's baseline patient share. The remaining columns introduce interactions between time effects and a series of pre-period characteristics. Column 2 considers pre-event physician attributes, including general experience, logged area-specific caseload, specialty, medical school of training, and measures of practice style (risk-adjusted length of stay and spending).[18] Column 3 examines pre-event patient characteristics (age, gender, race, and comorbidities), while Column 4 focuses on hospital structural features (ownership, complexity level, operating beds, and infrastructure indices).[19] Column 5 adds all of these covariates simultaneously. If our estimates were driven by differential trends in forecast errors or omitted determinants of performance correlated with baseline conditions, the inclusion of these rich controls should substantially alter the estimated effect. In practice, however, the coefficient of interest remains almost unchanged at -0.38.

Overall, we find that hospitals quickly offset most of the quality shock from a physician's death. In light of the remarkable stability of this finding across a wide range of alternative specifications, we focus in what follows on the computationally efficient and parsimonious specification in column 1 as our baseline.

### 4.2 Positive versus Negative Quality Shocks

The results presented so far capture the average effect of turnover-induced changes in physician quality. We next explore heterogeneity with respect to the deceased physician's value added relative to that of their colleagues. Figure 3 reveals a striking asymmetry. When a relatively high-performing physician departs and thus generates a expected negative quality shock, hospitals actively mitigate the loss, with estimates of $\beta$ much closer to zero and in general statistically insignificant. In contrast, the estimated effect is much more negative when a low-performing physician exits, with point estimates fluctuating around -0.6 and statistically significant. These patterns suggest that hospitals actively reorganize and mobilize resources to cushion the impact of losing valuable talent, but

[18] The risk-adjusted of length of stay and spending are constructed using the same empirical logic as the observational survival measure. They are specific to each physician and clinical area of practice within the hospital. To construct controls based on medical school of training and specialty, we first regress observational survival on a full set of medical-school fixed effects and separately on a full set of specialty fixed effects, and then assign to each physician the corresponding predicted value from these regressions. These predicted components capture the portion of observed performance attributable to coarse, salient credentials rather than physician-specific effectiveness.

[19] The infrastructure indices summarize hospitals' physical and technological capacity across broad categories of equipment and facilities. Each index is constructed by standardizing the availability of individual infrastructure items and then aggregating them into category-specific averages at the hospital–time level. This approach captures differences in the scale and scope of hospital infrastructure in a parsimonious way while avoiding reliance on any single piece of equipment.

exhibit little managerial action when turnover is beneficial by construction.

This asymmetry is consistent with the intuition that hospitals aim to preserve quality but reallocation toward higher-quality physicians is costly. Physicians with better clinical outcomes may already operate near capacity, precisely because hospitals have incentives to concentrate baseline caseloads on them to maximize expected survival. When such a physician departs, the potential loss in quality is large enough that hospitals find it worthwhile to bear the congestion and any other costs required to shift patients toward the best available substitutes. But when a low-performing physician departs, overall quality is expected to improve even under random reallocation. In this case, hospitals have weaker incentives to incur the additional costs of further concentrating patients on the very best physicians. The fact that the confidence intervals still exclude -1 following a positive quality shock suggests that reallocating patients to below-average-quality physicians entails lower costs than random reallocation in such instances.

### 4.3 Validation of Baseline Design

We now subject our main findings to a series of specification checks designed to address potential threats to identification.

**Testing for Differential Forecast Shocks.** To assess the plausibility of the common forecast shock assumption, we construct several proxies for the residual forecast component $\eta_{js}$ based on observable physician attributes and use them as placebo outcomes. This placebo strategy is analogous in spirit to the test in Abaluck et al. (2021) in the context of health plan choice. We first residualize each observable attribute on the risk-adjusted survival effect $\theta_{js}$ and then construct patient-weighted averages across hospital-areas and quarters, as in our main analysis. We consider a variety of physician characteristics, including medical school, specialty, experience, caseload, practice style (risk-adjusted spending and length-of-stay), as well as a composite index that averages these standardized components.[20] Intuitively, these residualized variables capture the component of each observable physician characteristic that is *unrelated* to the empirical survival signal, and therefore serve as observable proxies for the slow-moving, unobserved forecast-error term in our framework. If more-exposed areas undertake reallocation decisions based on signals orthogonal to observed survival but correlated with other observable dimensions of quality, our difference-in-differences specification would predict changes in these residualized characteristics. Figures 4 and A.3 show no evidence that this is the case. The point estimates are statistically insignificant and extremely small in magnitude, allowing us to rule out effects larger than 0.01 standard deviations. Overall, the evidence supports the plausibility of our statistical design.

[20] For categorical attributes such as medical school and specialty, we first construct continuous proxies by assigning to each physician the average observational survival associated with their category. We then residualize these category-level averages with respect to physician-level observational survival. The resulting residual captures the component of coarse professional credentials that is not explained by how well physicians in that category perform on average. While survival differences across categories may still exist, the residualized component isolates variation that is orthogonal to measured physician effectiveness.

**Spillovers.** When a physician unexpectedly leaves, the hospital may respond by drawing not only on physicians already working in the directly exposed area but also on physicians from other service lines, with potential negative spillovers. Spillovers within hospitals are concerning because shifting physicians from less-exposed to more-exposed areas lowers measured quality in the donor areas, making the comparison between these two areas look as if the original shock was absorbed more than it actually was (i.e., $\hat{\beta}$ toward 0). It is important to emphasize that because the baseline specification includes a rich set of event-by-area-by-time fixed effects, the parameter of interest is identified mainly from comparing the same areas of care across different hospitals. This should largely limit the scope for mechanical biases due to within-hospital spillovers.

As a robustness check, we replace these fixed effects with a full set of event-by-hospital-by-time fixed effects, a specification that relies entirely on within-hospital, across-area variation and therefore maximizes any bias from spillover effects. One might argue that any potential spillovers inside the hospital are more local, occurring between clinically adjacent areas, such as heart and cerebrovascular services. We therefore go step further and consider a specification that includes event-by-hospital-by-chapter-by-time fixed effects. A "chapter" here refers to the broadest ICD-10 level of aggregation and contains all subchapters we use to define areas of care. If spillover effects were quantitatively important, these demanding specifications should substantially move estimates of $\beta$ closer to zero. In practice, the estimates are statistically and economically indistinguishable from our baseline ones (Appendix Figure A.4). This suggests that, insofar as hospitals adjusted in response to the quality shock, they were able to do so without materially affecting other areas within the hospital.[21]

A related question is that the death event may represent a quality shock to not only the hospital hit by the event but also other hospitals if treated hospitals poach replacements from nearby control hospitals. The fact that the results are almost unaffected when we use the specification that exploits within-hospital variation suggests that such source of spillovers is unlikely to be relevant. As a further test, we repeat the matching algorithm but with the restriction that the control hospitals must be located *outside* the state of the treated hospital and then rerun the baseline specification. Unsurprisingly, this makes no difference for the results (Appendix Figure A.5).

**Concurrent Shocks.** One might be worried with concurrent events that coincide with physician deaths and directly affect hospital operations. By their very nature, plausible relevant candidates are extreme, system-wide events, including pandemics, wars, natural disasters, or large-scale public health crises. Because of their rarity and episodic timing, such events are unlikely to introduce an important bias in our estimates. In Appendix Table A.3, we show that our results are unaffected when excluding the period of the COVID-19 pandemic—one of the most extreme global shocks of the century—as well as hospitals experiencing multiple physician deaths in the same period, which

[21] Hospitals might rely on sources of capacity that do not require taking physicians away from other active service lines, such as hiring new physicians or expanding the hours of physicians already attached to the affected area of care. Even if the hospital draws physicians from other areas, it may do so from doctors who were operating below capacity, shifting them toward the shocked area without materially lowering performance in their original area.

could plausibly be driven by extreme events.

**Other Robustness Checks.** Appendix Tables A.3-A.5 report a number of additional robustness checks. One set of specifications recomputes our baseline estimates using alternative clustering levels for inference, including clustering on matched pairs, municipalities, and microregions. A second set of exercises explores alternative constructions of the observational survival by removing the empirical Bayes shrinkage and, in a further step, abstracting from risk adjustment altogether. We also reestimate the model excluding hospitals that experience multiple physician deaths with overlapping pre and post periods, so that identification is not driven by facilities facing unusually dense sequences of shocks. Across all of these alternative specifications, the estimated effects are very similar to our baseline results and the main conclusions of the paper are unchanged.

### 4.4 No Evidence of Adjustment in Patient Volume or Composition

Reallocating patients across physicians is not the only possible margin through which hospitals may mitigate the effects of losing a top physician. Hospitals might attempt to mitigate the consequences of the quality shock by changing the number of patients admitted or by altering the composition of patients they treat. However, such adjustments may be difficult to justify institutionally. In the absence of a clear capacity constraint, tightening admission criteria could be interpreted as restricting access to care, potentially exposing hospitals to legal repercussions. It is important to emphasize that our baseline specification controls for the deceased physician's pre-event patient share interacted with time effects, so the quality shock is not capturing mechanical differences in service capacity that could otherwise make changes in admissions practices defensible.

We next explore empirically this question directly. We begin by repeating our baseline specification using the number of patients admitted as the outcome of interest. As shown in Appendix Figure A.6, we find no statistically meaningful changes in patient volume associated with the quality shock. We then turn to the observable composition of patients, including age, race, gender, and comorbidities. The quality shock is uncorrelated with changes in these pre-determined characteristics. The same pattern emerges when we examine predicted survival as a function of these observables, an approach that improves power by weighting patient characteristics by their relevance for mortality (Appendix Figure A.8). Overall, there is no evidence that hospitals adjust their admitting behavior in response to the turnover-induced quality shock.

### 4.5 Signal Reliability and Impacts on Survival

**The Forecast Coefficient.** We next examine how informative the observed variation in survival rates is about true clinical ability. Our design does not require the forecast coefficient $\lambda$ equal to one. However, the interpretation of our estimates is complicated from a statistical and economic perspective when $\lambda$ approaches 0, since a completely uninformative signal can dramatically magnify biases from even negligible violations of the identification conditions. To measure the reliability of

signal, we combine the projection of physician value added in (5) to the empirical representation of the log-linearized production function in (1):

$$y_{sht} = \lambda \bar{\theta}_{sht} + \sum_{\omega} \phi_w z_{sht} + \pi_{st} + \delta_{sh} + \bar{\eta}_{sht} + \xi_{sht}, \tag{7}$$

This model relates realized survival $y_{sht}$ to expected physician effects $\bar{\theta}_{sht}$. The term $\xi_{sht}$ is an error term that captures unobserved determinants of mortality. We can exploit the variation in the observational survival induced by physician deaths in an instrumental variable (IV) strategy. Under this approach, the stacked difference-in-differences specification (6) would be the first stage and equation (7) is the corresponding second stage. The fact that the net effect of the quality shock documented in Table 1 is, on average, negative and highly significant implies that we have a strong first stage for the purposes of this exercise. In addition to conditions 1 and 2, identification of $\lambda$ using this IV approach requires that the turnover-induced quality shock be orthogonal to the error term $\xi_{sht}$. This assumption could be violated if, for example, areas experiencing more pronounced quality shocks were also admitting healthier or sicker patients following a physician's death. The evidence in Section 4.4 suggests that this is unlikely to be the case.

Table 2 reports the key findings of this subsection. Columns 1-2 present the reduced-form estimates, showing that a one-unit larger quality shock lowers survival by about 0.42 percentage points. The corresponding first stage is indeed strong, with a $F$-statistics above the relevant threshold for weak instrument of 100 (Lee et al., 2022). Using this variation as an instrument for observational survival, the forecast coefficient is estimated at 1.13 and is not statistically different from 1. The point estimate remains unchanged when we add patient demographics and comorbidities, consistent with the balance evidence that the shock is not picking up changes in patient mix. Read together, the results indicate that the observational value added is a highly reliable predictor of true physician clinical effectiveness.

**Adverted Deaths.** Given the high reliability of observational value added and its substantial cross-sectional dispersion, the sizable hospital response to turnover-induced quality shocks averts a large share of deaths that would otherwise follow the exit of high-performing physicians. To gauge magnitudes, consider a one–standard deviation negative quality shock of $\sigma = 0.012$. Absent any targeted reallocation (i.e., $\gamma = 0$, so $\beta = -1$), survival would fall by:

$$100 \times \hat{\lambda} \times 1 \times \sigma \approx 1.36$$

In the data, for negative shocks we estimate $\hat{\beta}_{\text{neg}} \approx -0.14$, so the realized change is

$$100 \times \hat{\lambda} \times |\hat{\beta}_{\text{neg}}| \times \sigma \approx 0.19$$

Hence, targeted reallocation *averts* about $1.36 - 0.19 \approx 1.17$ percentage points, or roughly $1 - 0.14 = 86$ percent of the deaths that would have occurred under a purely random reassignment of the deceased's caseload. This corresponds to avoiding roughly a 20 percent increase in mortality relative to the pre-event period.

**What Do High-Survival Physicians Do?** Physicians may impact patient health both through *which* treatments they choose and through *how* they deliver a given treatment. On the *which* margin, better physicians may systematically differ in their treatment decisions relative to the average physician if they are better at detecting case-specific clinical signals and acting on information not encoded in guidelines. On the *how* margin, physicians may differ in the productivity with which they apply a given set of inputs. High-survival physicians may be more willing to order additional diagnostics, increase bedside surveillance, or move earlier to invasive therapy, which can rescue patients on the brink of death. At the same time, they may use the available inputs with greater precision and timeliness, potentially preventing complications at the outset.

To investigate these questions, we re-estimate equation (7), replacing survival with outcomes that capture physicians' use of clinical inputs. We examine a case-level clinical deviation score that measures how each admission's set of delivered procedures departs from the typical procedure pattern for the same 3-digit ICD condition.[22] We also examine patients' length of stay, an indicator for intensive care unit (ICU) admission, and total spending. ICU admission may reflect a range of factors, including pre-existing severity, but we treat it as a marker for acute deterioration and escalation of care during the hospitalization given the absence of systematic changes in case mix correlated with the quality shock.

The results are presented in Appendix Table A.6. We find that observational survival is associated with higher clinical deviation scores, suggesting that higher-survival physicians make treatment choices that depart more from typical practice patterns. The evidence also indicates that patients treated by higher-survival physicians are less likely to require ICU care and incur lower total spending. We detect no statistically meaningful change in length of stay, so the spending decline is unlikely to be driven by shorter admissions and instead points to differences in the intensive margin of care. Taken in its entirety, these patterns are consistent with high-survival physicians delivering more selective, case-tailored care that avoids low-value clinical add-ons and prevents acute deterioration that would otherwise trigger intensive rescue care.

## 5 Mechanisms

The preceding results demonstrate that hospitals substantially offset the quality shock from a physician's death. In this section, we investigate the mechanisms enabling this adjustment. As we show below, the evidence points to the use of internal information supported by flexible organizational capacity as key factors shaping the hospital response.

[22] For each 3-digit primary ICD diagnosis $d$, we form the modal national distribution of procedure codes $r_d$ and compute the cosine similarity between an admission's realized procedure set and $r_d$. The deviation score is $Dev_{id} = 1 - Sim_{id}$, where higher values indicate greater departure from modal practice. Because cosine similarity is normalized by the number of distinct procedures in the admission, this measure captures differences in procedure *composition* rather than mechanically reflecting procedure intensity. Appendix Section D provides further details.

### 5.1 Internal versus External Replacements

A first key question is whether hospitals absorb the quality shock primarily by reallocating patients across incumbent physicians or by hiring replacements from outside. The discrete, one-time pattern shown in Figure 2 is consistent with an internal reallocation mechanism that relies on information available at the time of the event. In contrast, an adjustment relying on external hiring, where information about candidates is inherently more limited, would likely lead to a more gradual recovery. To investigate this question directly, we first decompose the observable value added in a service among alive physicians as follows:

$$\bar{\theta}^{\text{alive}}_{sht} = \underbrace{\sum_{j \in \mathcal{I}_{sht}} \alpha_{j,sht}\theta_{js}}_{\text{Incumbent contribution}} + \underbrace{\sum_{k \in \mathcal{N}_{sht}} \alpha_{k,sht}\theta_{ks}}_{\text{New hire contribution}}$$

where $\mathcal{I}_{sht}$ and $\mathcal{N}_{sht}$ denote the sets of incumbent and newly hired physicians, respectively, and $\sum_{j \in \mathcal{I}_{sht}} \alpha_{j,sht} + \sum_{k \in \mathcal{N}_{sht}} \alpha_{k,sht} = 1$. This decomposition simply writes overall observational value added as a convex combination of the average value added from incumbent and newly hired physicians, weighted by their respective patient shares. Table 3 reruns our baseline specifications using the incumbent and new hire contributions as dependent variables. As one can infer from the table, the adjustment is driven by incumbents, with an estimated coefficient that is statistically identical to that for the overall pool of alive physicians. Following turnover after death, hospitals reallocate the vacant caseload to incumbents rather than to new hires, arguably leveraging their pre-existing knowledge of the incumbents' clinical effectiveness and thereby reducing the costs of managerial decision making.

An alternative interpretation is that hospitals rely on internal reallocation not because information on incumbents is inherently less noisy but because external hiring is particularly difficult. To investigate this question, Appendix Table A.7 examines whether the adjustment margin differs between thin and thick physician labor markets. We classify labor markets as thin and thick according to whether the number of physicians per capita in the microrregion is below or above the median, respectively.[23] The data reveal that internal reallocation remains the dominant adjustment channel regardless of market thickness, with the contribution of external hires close to zero and statistically insignificant in both thin and thick markets. Even when external replacements should be relatively easier to find, hospitals do not adjust to quality shocks through this margin.

### 5.2 Heterogeneity by Deceased's Tenure

While hospitals appear to reorganize effectively after unexpected physician exits, these results do not clarify the extent to which accumulated learning about physicians matter. This is important it because provides insights on the role of managerial attention in shaping the compensatory responses. Figure 5 shows that the mitigating response to quality shocks is strongly increasing in the tenure

[23] A microrregion (or *microrregião* in Portuguese) in Brazil is a statistical geographic division comprising a contiguous group of neighboring municipalities with similar economic and social characteristics. It is broadly analogous to a U.S. "commuting zone" in that it captures local labor-market and service-area geography.

of the departing physician. The effect for when the deceased physician had less than one year of tenure is more than twice as large in magnitude as that for those with over three years, a difference that is statistically significant at conventional levels. This finding is consistent with a learning mechanism. Managers receive increasingly precise signals as physicians accumulate tenure, sharpening posterior beliefs and enabling more effective allocation decisions.

### 5.3 Acting Upon General Credentials

A natural, relatively simple interpretation of our findings is that hospitals do care about preserving clinical quality, but do so in ways that do not require fine-grained knowledge of each physician's individual value added. Instead, managers may rely on coarse, readily observable credentials to make reallocation decisions. When a highly experienced physician leaves, hospitals may simply reassign patients to colleagues with similar tenure or seniority with the idea that experience is a fundamental predictor of clinical performance. Similarly, if physicians in a given specialty tend, on average, to deliver better outcomes in a given area of care, hospitals may shift the departing physician's caseload to others in the same field. For example, cardiologists may systematically achieve better outcomes for myocardial infarction cases than general internists, making specialty affiliation a simple yet informative signal for managers. Under this view, targeting quality-preserving reallocations is straightforward.

For this mechanism to rationalize our results, however, most of the cross-physician dispersion in observational value added would have to be explained by these general credentials alone. In practice, this is far from true. A regression of the observational valued-added against a full set of fixed effects for a physician's experience and specialty-by-area-of-care yields a $R^2$ of only 0.12, implying that approximately 90 percent of the variation occurs among physicians with the same observable credentials. Two cardiologists can display very different survival value added, just as two general internists or two highly experienced physicians can differ markedly in their clinical performance. Consistent with this interpretation, Appendix Table A.8 shows that our estimates of $\beta$ are not materially affected when we use a measure of observational value added purged of the variation attributable to experience and specialty. Hospital managers appear to act on richer signals of individual physicians' effectiveness than those captured by experience and specialty alone.

### 5.4 Management Quality

The large mitigating response documented in Table 1 can be viewed as evidence that managerial attention matters following turnover events. To dig deeper, we next examine how the response varies across hospitals that differ in their degree of attentiveness based on the adoption of formal monitoring and quality-control structures in the pre-event period. We rely on administrative data that record whether hospitals have established key oversight committees, collected by the Ministry of Health to verify that institutions maintain formal structures for routine review, information flow, and control monitoring. We classify hospitals by whether they report having established all, some,

or none of the formal committees responsible for mortality/biopsy review, infection control, and cost accounting.

The results of this exercise are reported in Table 5. For negative quality shocks, a clear pattern emerges. Hospitals with established formal practices exhibit essentially complete adjustment in response to the exit of high-value added physicians, while those with limited practices experience meaningful and statistically significant losses. Hospitals with only partial structures fall in between. By contrast, differences are far less pronounced when the shock is positive. All types of hospitals offset only part of the mechanical improvement that would arise from the exit of a low-performing physician, and the magnitudes are broadly similar across established, partial, and limited groups. Taken together, these results are consistent with the hypothesis that managerial attention is scarce and deployed more intensively to avoid losses than to pursue incremental improvement. More institutionalized monitoring capacity facilitates more effective, targeted reallocations when performance is at risk.

### 5.5 Labor Slack

The hospital's ability to carry out targeted reallocations, especially toward its best physicians, may depend on pre-existing workloads. Scheduling rigidities, capacity constraints, and coordination frictions could be easier to overcome when baseline workloads are lighter. Motivated by this reasoning, Table 5 re-estimates the baseline specification separately for hospitals below and above the median pre-event patient caseload per physician. Consistent with labor slack enabling targeted reallocations, we find that areas experiencing negative quality shocks exhibit complete adjustment when baseline workload is low but experience sizable and statistically significant losses when workloads are high. For positive quality shocks, the estimate is closer to -1 when baseline workload is low than when it is high (-0.74 versus -0.44). This pattern suggests that when workloads are high, hospitals face greater difficulty fully realizing the mechanical quality gain from the exit of a low-performing physician, thereby routing some patients to similarly below-average physicians. Overall, the evidence suggests that effective reallocation requires spare capacity.

## 6 Discussion and Interpretation

### 6.1 Co-existence of Bad and Good Physicians

Our findings are consistent with a framework where employers learn workers' value added over time and use this information to actively organize production. But if this is true, why do lower-ability physicians receive any patient assignments to begin with? While this might seem counterintuitive at first glance, it arises naturally in a model with the most basic assumptions on production and costs. Consider an environment in which hospitals must allocate a fixed mass of patients across physicians. Patient volume is exogenously determined by community health needs and coverage mandates that preclude selectively treating fewer patients. Hospitals start with a set $\mathcal{J}$ of physicians.

Each physician $j$ possesses a time-invariant clinical ability, which is heterogeneous and unknown to the hospital. The hospital has beliefs about each $\mu_j$ and updates them over time using a vector of noisy, physician-specific signals $\boldsymbol{\sigma}_t \equiv \{\sigma_{jt}\}_{j\in\mathcal{J}}$ observed at the end of each period $t$. The hospital's posterior belief about $j'$ clinical effectiveness in period $t$ is denoted by $\hat{\mu}_{jt}$. For simplicity, we assume that effort costs are homogeneous across physicians and that managerial actions are costless (no reassignment or monitoring frictions).

Physicians are compensated on a fee-for-service basis. However, treating patients imposes an effort cost $c(q_{jt})$ on physician $j$ when handling a caseload of $q_{jt}$ patients. We assume this cost function is twice differentiable, increasing, and strictly convex, which implies that the marginal disutility of effort is increasing in caseload. To accept an additional patient, a physician must be compensated for their marginal effort cost, making it increasingly costly for the hospital to concentrate patients in a small subset of physicians. This intuitive formulation is consistent with more general models of physician behavior (Ma and McGuire, 1997; Clemens and Gottlieb, 2014; Chan, 2018).

We also assume that the hospital is subject to a minimum-quality standard—e.g., clinical governance or regulatory requirements—that constrain the average expected quality of care to be at least $\bar{\mu}^* > 0$. Formally, the hospital's objective is:

$$\max_{\{q_{jt}\}_j} \sum_j \hat{\mu}_{jt}(\sigma_t)\, q_{jt} - \sum_j c(q_{jt}) \quad \text{s.t.} \quad \sum_j q_{jt} = Q_t,\ q_{jt} \geq 0, \quad \frac{1}{Q_t}\sum_j \hat{\mu}_{jt}(\sigma_t)\, q_{jt} \geq \bar{\mu}^*.$$

In this framework, the allocation decision is driven by the trade-off between the expected marginal product of a physician and the marginal cost of compensating that physician for effort, tilted by the requirement that the average expected quality cannot fall below $\bar{\mu}^*$. If a physician's expected marginal survival gain net of marginal cost exceeded the corresponding margin for another physician, a feasible reallocation would raise the objective, contradicting optimality unless the quality restriction is binding and renders the move infeasible. When the quality restriction is slack, then the net marginal benefit of assigning an additional patient is equalized across all physicians so that for any two physicians $j$ and $k$:

$$\hat{\mu}_{jt} - c'(q_{jt}) \;=\; \hat{\mu}_{kt} - c'(q_{kt}).$$

This condition implies a "fill-the-top-first" allocation rule that always gives the next patient to the physician with the highest expected net marginal benefit. Start with the highest-expected-quality doctor and increases that physician's caseload until rising marginal cost pushes their net marginal benefit down to the next-best.[24]Figure 6 illustrates visually the intuition of the equilibrium in a

[24]When the minimum–quality requirement binds, the equalization condition takes a weighted form. Let $\varphi_t \geq 0$ denote the shadow value of the quality floor in period $t$. For any two active physicians $j, k$ (i.e., $q_{jt}, q_{kt} > 0$), optimality requires

$$(1+\varphi_t)\,\hat{\mu}_{jt} - c'(q_{jt}) \;=\; (1+\varphi_t)\,\hat{\mu}_{kt} - c'(q_{kt}),$$

so the common margin is computed after reweighting perceived ability by the factor $(1+\varphi_t)$. Intuitively, $\varphi_t > 0$ raises the effective value of quality at the margin, tilting assignments toward higher-ability physicians and potentially inducing corner solutions for sufficiently low-ability types (e.g., $q_{jt} = 0$ whenever $(1+\varphi_t)\hat{\mu}_{jt}$ is too small to justify a positive load given marginal effort costs).

setting where there are only two physicians.

An immediate implication of the fill-the-top-first condition is that it leaves lower-ability doctors with positive caseload and this can occur even under perfect learning on quality. This is not because the hospital is indifferent to quality, but because concentrating everything in a few good doctors becomes prohibitively costly. When the quality restriction is binding, there could be corner solutions for sufficiently low-ability physicians. If meeting $\bar{\mu}^*$ requires concentrating cases on higher-ability doctors, some physicians will see their caseload reduced. Since quality is learned from noisy signals rather than observed directly, another implication is that some lower-ability physicians may temporarily carry a relatively high caseload when idiosyncratic signals inflate their perceived value added.

In a perfectly fluid labor market with no search costs and no regulatory or contractual frictions, persistently low-performing physicians would be rare. Hospitals would simply dismiss low-ability doctors and hire higher-ability replacements from outside. However, this is a highly sensitive point, especially in a profession where labor supply is tightly regulated and entry requires extensive credentialing and onboarding. Typical job-protection laws make dismissal costly, and even with precise signals, punitive actions based on them are error-prone and expose hospitals to legal risk. In equilibrium, "bad" physicians co-exist alongside "good" ones even when hospitals have perfectly learned to differentiate them.

### 6.2 Are (Observationally) Better Physicians Rewarded?

In classical theories of employer learning, organizations that identify talent and recognize its value invest in retention by offering better working conditions or opportunities for advancement (Jovanovic, 1979; Harris and Holmstrom, 1982). While a detailed formal treatment of these models is beyond the scope of this paper, we examine whether basic correlations in the data are consistent with the most fundamental prediction of the theory. To do so, we construct a physician-level analysis dataset with information on average compensation and job duration. Total compensation is computed as the sum of labor earnings reported by all employing establishments in Brazil's employer–employee matched registry over 2003–2018.[25] Job duration is defined as the average length in quarters of a physician's employment across all hospitals observed in the administrative records of medical practice. Our measure of observational value added is originally estimated at the physician-service area level. For this analysis, we aggregate it to a single physician-level value by taking a patient-weighted average across all areas. To ease interpretation and mitigate the influence of outliers, we transform the observational value added into percentiles based on 50 equal-sized bins.

Figure 7 presents plots the labor market outcomes against percentiles of the observational value-

[25] We use the RAIS (*Relação Anual de Informações Sociais*), which is an administrative census of formal employment in Brazil. While the RAIS wages and other contract attributes at the establishment–worker level, the definition of an "employer" in the context of physicians is somewhat diffuse. Many healthcare establishments, including hospitals across different regions in the country, are generally grouped under a single employer entity in the RAIS, such as the Ministry of Health or a state-level health department. This administrative structure prevents us from precisely identifying the specific hospital-establishment where a physician works, precluding a more granular analysis at the physician-hospital level. We provide institutional details, variable construction, and standard cleaning steps in Appendix B.

added distribution. Panel A shows a clear, positive relationship between observational survival and log hourly earnings. Physicians at the top of the value-added distribution earn significantly more than those at the bottom, with a slope that is highly significant. Moving from the 25th to the 75th percentile of the value-added distribution is associated with an earnings difference of approximately 4.1 percent. Panel B reveals a similarly positive gradient for job duration. Higher-value-added physicians tend to have longer-lasting employment spells at a given hospital. This relationship is not only statistically significant but quantitatively meaningful. The gap in job duration between physicians at the top and bottom of the distribution amounts to 3.3 quarters or 8.3 percent of the average job duration in our sample.

The patterns in the data are consistent with the interpretation that clinical effectiveness is valued in the labor market. Hospitals that accurately perceive physician quality appear to retain and promote higher performers. At this point, it is important to emphasize that this evidence is tentative and obviously we are not seeking to make causal claims. Our estimates should be interpreted as descriptive correlations. It is possible that more effective physicians have personal characteristics, such as better networking skills and ambition, that lead to better labor market outcomes, independent of any active retention efforts by their employers. Nevertheless, at a broad level, these findings can be taken as a basic stylized fact that can help guide future research.

# 7 Conclusion

This paper provides systematic evidence that organizations can extract actionable information from inherently noisy performance signals and use it to actively reorganize production. Exploiting quasi-random physician exits caused by unexpected deaths, we find that hospitals selectively reallocate patients across physicians to mitigate losses in care quality. The reallocation response is strongly sharp and targeted when high-quality physicians exit but diffuse when low-quality physicians leave. This asymmetry is consistent with the intuition that hospitals aim to preserve quality but reallocation toward higher-quality physicians is costly. When hospitals face no immediate threat to quality, they have limited incentives to incur the cost of redirecting patients toward its most overburdened top performers.

The evidence further shows that effective adjustment hinges on organizational capabilities rather than on information alone. Reallocation occurs almost exclusively among incumbent physicians, intensifies with tenure, and is strongest in hospitals with greater managerial capacity and available slack. Together, these findings imply that inference about quality and the ability to act on that knowledge are distinct constraints. Even when performance differences are well understood, organizations may be unable or unwilling to reoptimize task assignment without sufficient flexibility. More broadly, the results highlight a central role for internal organization in shaping how individual-level performance differences translate into aggregate outcomes. Beyond its immediate implications for healthcare, our framework offers a template for assessing how firms in other noisy-performance settings discern, target, and redeploy human capital when it matters most.

# Bibliography


**Abaluck, Jason, Mauricio Caceres Bravo, Peter Hull, and Amanda Starc**, "Mortality effects and choice across private health insurance plans," *The quarterly journal of economics*, 2021, *136* (3), 1557–1610.

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

**Altonji, Joseph G and Charles R Pierret**, "Employer learning and statistical discrimination," *The quarterly journal of economics*, 2001, *116* (1), 313–350.

**Arrow, Kenneth J.**, "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review*, 1963, *53* (5), 941–973.

**Baker, George P**, "Incentive contracts and performance measurement," *Journal of political Economy*, 1992, *100* (3), 598–614.

**Bandiera, Oriana, Iwan Barankay, and Imran Rasul**, "Incentives for managers and inequality among workers: Evidence from a firm-level experiment," *The Quarterly Journal of Economics*, 2007, *122* (2), 729–773.

**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.

**Bau, Natalie and Jishnu Das**, "Teacher value added in a low-income country," *American Economic Journal: Economic Policy*, 2020, *12* (1), 62–96.

**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.

**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.

**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**, "The efficiency of slacking off: Evidence from the emergency department," *Econometrica*, 2018, *86* (3), 997–1030.

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

**Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson**, "Health care exceptionalism? Performance and allocation in the US health care sector," *American Economic Review*, 2016, *106* (8), 2110–2144.

**Chetty, Raj, John N Friedman, and Jonah E Rockoff**, "Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates," *American economic review*, 2014, *104* (9), 2593–2632.

_ **,** _ **,** _ **, and** _ , "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood," *American economic review*, 2014, *104* (9), 2633–2679.

_ **,** _ **,** _ **, and** _ , "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood," *American Economic Review*, 2014, *104* (9), 2633–79.

**Clemens, Jeffrey and Joshua D Gottlieb**, "Do physicians' financial incentives affect medical treatment and patient health?," *American Economic Review*, 2014, *104* (4), 1320–1349.

**Currie, Janet, W Bentley MacLeod, and Jessica Van Parys**, "Provider practice style and patient health outcomes: The case of heart attacks," *Journal of health economics*, 2016, *47*, 64–80.

**Doyle, Joseph J, John A Graves, Jonathan Gruber, and Samuel A Kleiner**, "Measuring returns to hospital care: Evidence from ambulance referral patterns," *Journal of Political Economy*, 2015, *123* (1), 170–214.

**Einav, Liran, Amy Finkelstein, and Neale Mahoney**, "Producing health: measuring value added of nursing homes," *Econometrica*, 2025, *93* (4), 1225–1264.

**Garicano, Luis**, "Hierarchies and the Organization of Knowledge in Production," *Journal of political economy*, 2000, *108* (5), 874–904.

_ **and Esteban Rossi-Hansberg**, "Organization and inequality in a knowledge economy," *The Quarterly journal of economics*, 2006, *121* (4), 1383–1435.

**Ginja, Rita, Julie Riise, Barton Willage, and Alexander Willén**, "Does Your Doctor Matter?," *Journal of Political Economy Microeconomics*, 2025, *3* (3), 497–538.

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

**Harris, Milton and Bengt Holmstrom**, "A theory of wage dynamics," *The Review of Economic Studies*, 1982, *49* (3), 315–333.

**Hjort, Jonas**, "Ethnic divisions and production in firms," *The Quarterly Journal of Economics*, 2014, *129* (4), 1899–1946.

**Holmstrom, Bengt**, "Moral hazard in teams," *The Bell journal of economics*, 1982, pp. 324–340.

**Holmström, Bengt**, "Pay for performance and beyond," *American Economic Review*, 2017, *107* (7), 1753–1777.

**Holmstrom, Bengt and Paul Milgrom**, "Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design," *The Journal of Law, Economics, and Organization*, 1991, *7* (special_issue), 24–52.

**Hull, Peter**, "Estimating hospital quality with quasi-experimental data," *Available at SSRN 3118358*, 2018.

**Jäger, Simon, Jörg Heining, and Nathan Lazarus**, "How substitutable are workers? evidence from worker deaths," *American Economic Review*, 2025, *Forthcoming*.

**Jovanovic, Boyan**, "Job matching and the theory of turnover," *Journal of political economy*, 1979, *87* (5, Part 1), 972–990.

**Kahn, Lisa B and Fabian Lange**, "Employer learning, productivity, and the earnings distribution: Evidence from performance measures," *The Review of Economic Studies*, 2014, *81* (4), 1575–1613.

**Kane, Thomas J. and Douglas O. Staiger**, "Estimating teacher impacts on student achievement: An experimental evaluation," Working Paper w14607, National Bureau of Economic Research 2008.

**Kane, Thomas J, Daniel F McCaffrey, Trey Miller, and Douglas O Staiger**, "Have we identified effective teachers? Validating measures of effective teaching using random assignment," *Research Paper. MET Project. Bill & Melinda Gates Foundation*, 2013.

**Lange, Fabian**, "The speed of employer learning," *Journal of Labor Economics*, 2007, *25* (1), 1–35.

**Lazear, Edward P**, "Salaries and piece rates," *Journal of business*, 1986, pp. 405–431.

**Lee, David S, Justin McCrary, Marcelo J Moreira, and Jack Porter**, "Valid t-ratio Inference for IV," *American Economic Review*, 2022, *112* (10), 3260–90.

**Ma, Ching-To Albert and Thomas G McGuire**, "Optimal health insurance and provider payment," *The American Economic Review*, 1997, pp. 685–704.

**Morris, Carl N.**, "Parametric empirical Bayes inference: theory and applications," *Journal of the American Statistical Association*, 1983, *78* (381), 47–55.

**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.

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

**Schönberg, Uta**, "Testing for asymmetric employer learning," *Journal of Labor Economics*, 2007, *25* (4), 651–691.

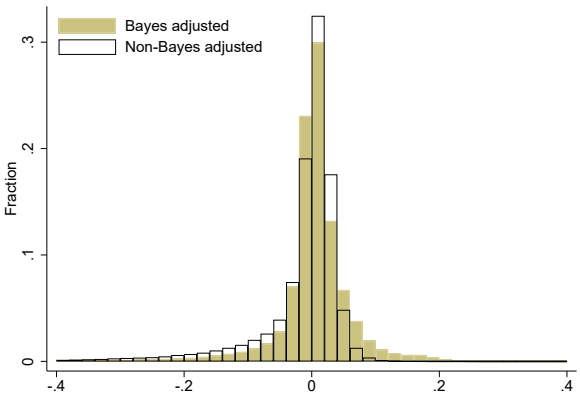
**Silver, David**, "Haste or Waste? Peer pressure and productivity in the emergency department," *The Review of Economic Studies*, 2021, *88* (3), 1385–1417.

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

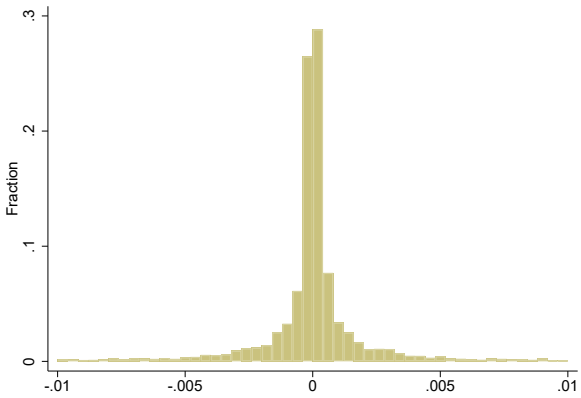
**Wing, Coady, Seth M Freedman, and Alex Hollingsworth**, "Stacked difference-in-differences," Technical Report, National Bureau of Economic Research 2024.

**Figure 1:** Observational Value Added and Quality Shocks

**Panel A.** Risk-Adjusted Survival

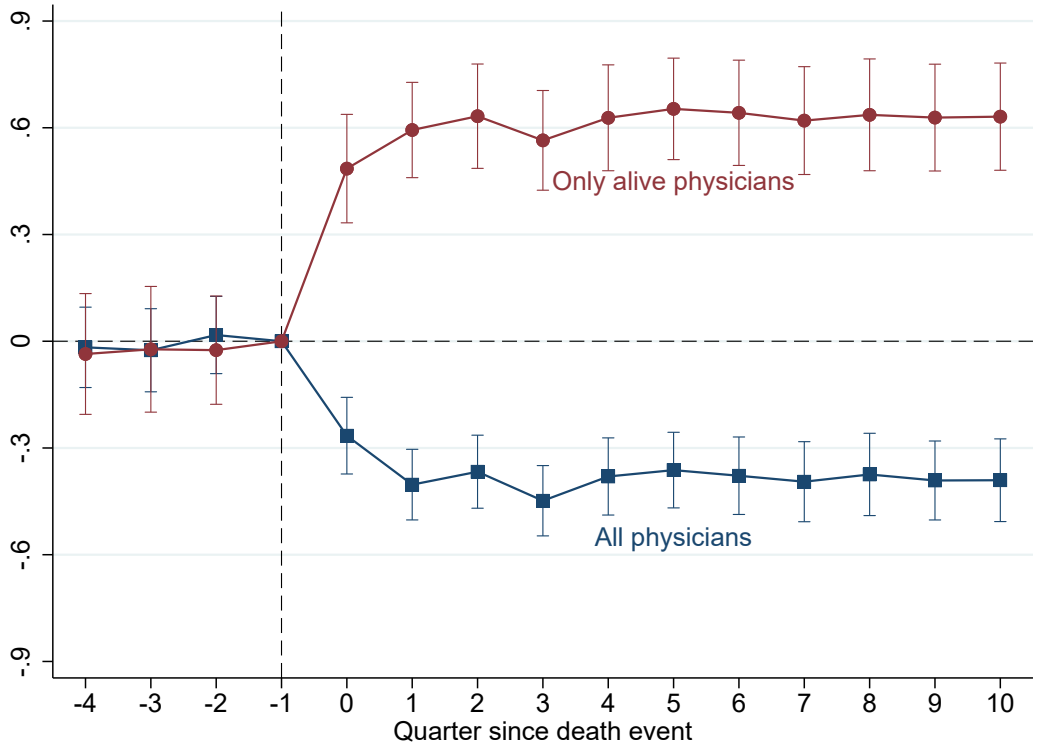



**Panel B.** Quality Shock



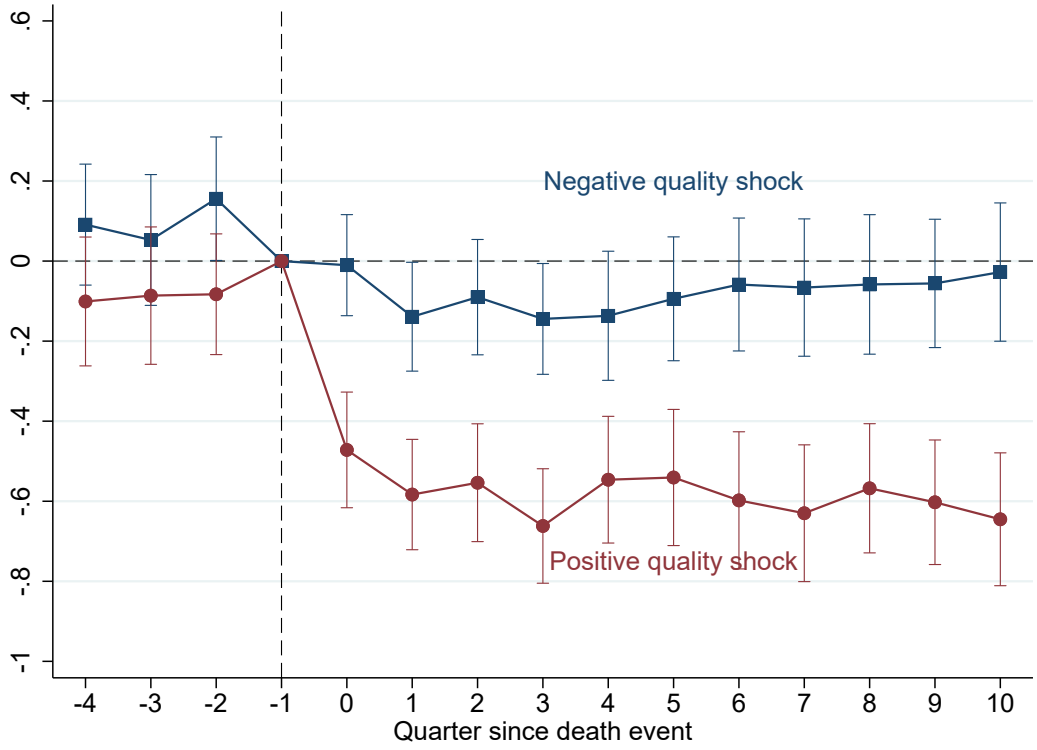

*Notes.* Panel A plots the distribution of physicians' observational value added measured by risk-adjusted patient survival. The non-Bayes-adjusted measure corresponds to raw risk-adjusted survival differences across physicians, while the Bayes-adjusted measure applies empirical Bayes shrinkage to account for sampling variation. Panel B shows the distribution of turnover-induced quality shocks, as described in Section 3.

**Figure 2:** Effects on Quality Composition of Physicians



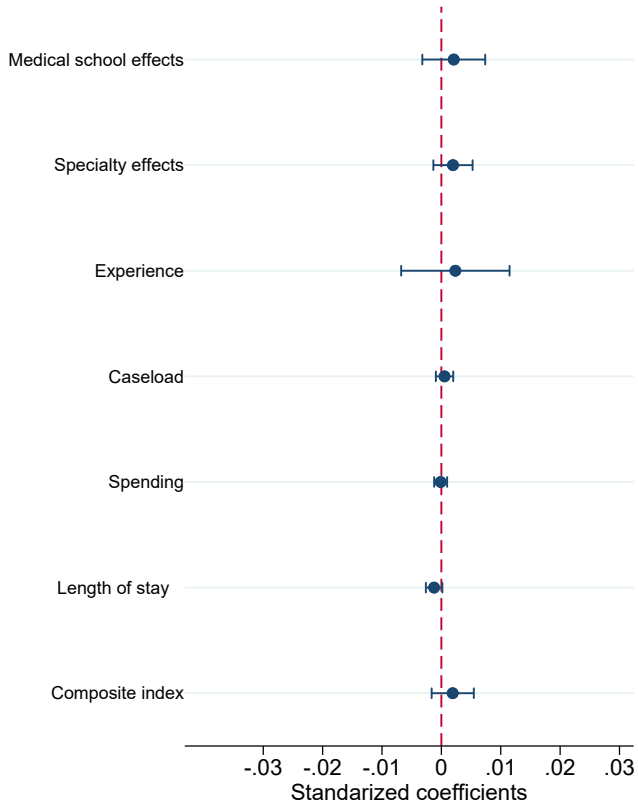

*Notes:* This figure plots event-study estimates of changes in the quality composition of physicians in a given hospital–area of care following an unexpected physician death. The outcome is the patient-weighted average of physicians' observational survival in that hospital–area–quarter, where each physician's contribution is weighted by their share of patients treated in that period. The blue series ("All physicians") computes this patient-weighted average using the full set of treating physicians. The red series ("Only alive physicians") computes the same patient-weighted average after excluding the physician whose death defines the event. The horizontal axis indexes quarters relative to the death event, with coefficients normalized to zero in the quarter immediately prior to the event. All specifications include hospital-area-by-event and area-by-quarter-by-event fixed effects, as well as interactions between the deceased physician's baseline patient share and time fixed effects. Vertical bars indicate 95 percent confidence intervals based on standard errors clustered at the hospital level.

**Figure 3:** Effects on Quality Composition of Physicians:
(Negative versus Positive Quality Shocks)



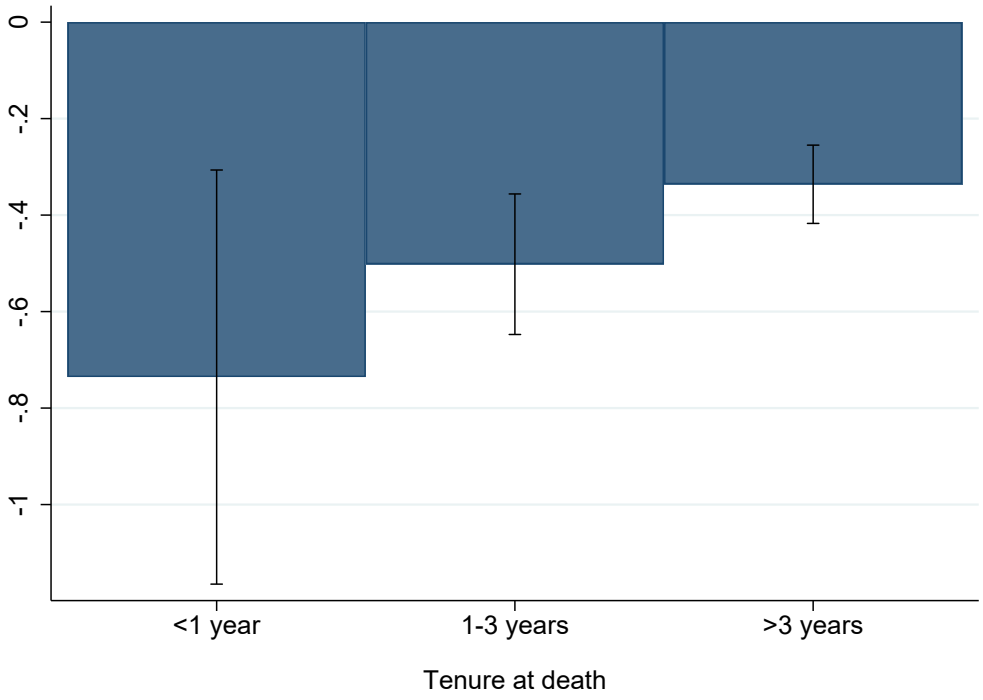

*Notes:* This figure plots event-study estimates of changes in the quality composition of physicians following an unexpected physician death, separately by the sign of the implied quality shock. The outcome is the patient-weighted average of physicians' observational survival in that hospital-area-quarter, where each physician's contribution is weighted by their share of patients treated in that period. Negative quality shocks correspond to the death of a physician whose observational survival exceeds the hospital–area average, while positive quality shocks correspond to the death of a physician whose observational survival falls below that average. The horizontal axis indexes quarters relative to the death event, with coefficients normalized to zero in the quarter immediately prior to the event. All specifications include hospital-area-by-event and area-by-quarter-by-event fixed effects, as well as interactions between the deceased physician's baseline patient share and time fixed effects. Vertical bars indicate 95 percent confidence intervals based on standard errors clustered at the hospital level.

**Figure 4:** Falsification Test for Common Forecast Shocks:
Effects on Residualized Characteristics of Physicians



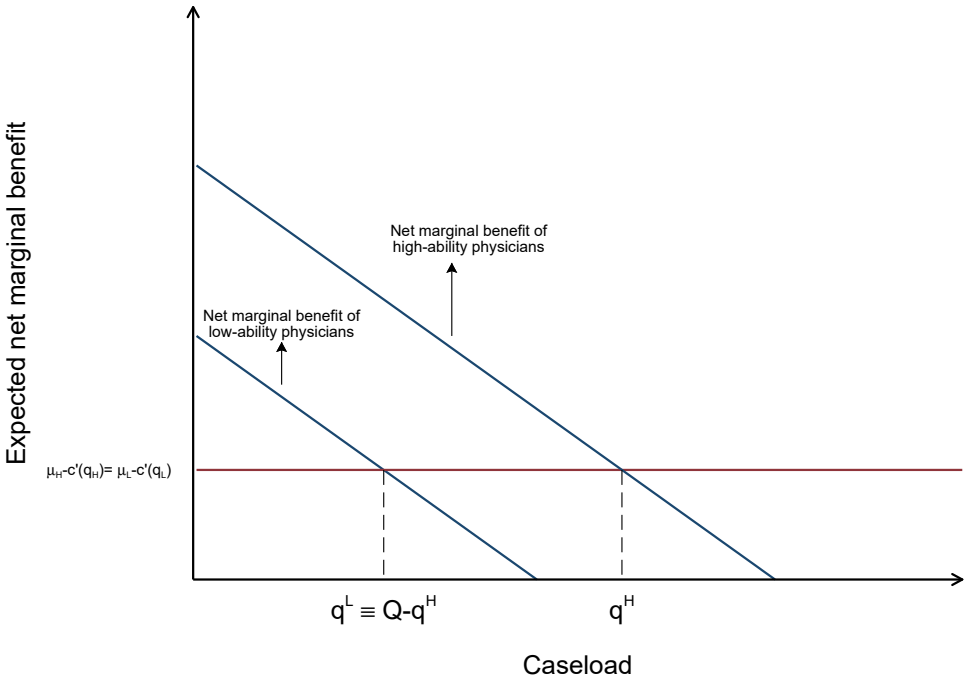

*Notes:* This figure reports difference-in-differences estimates from a falsification exercise that tests for differential forecast shocks unrelated to observed survival. Each point corresponds to the estimated effect of the turnover-induced quality shock on a residualized physician characteristic. Observable physician attributes are first residualized with respect to observational survival and then aggregated into patient-weighted averages at the hospital-area-quarter level, as in the main analysis. For categorical attributes such as medical school and specialty, the residualized measures capture variation in coarse credentials that is orthogonal to physician-level survival. The composite index averages these standardized residualized characteristics. All specifications include hospital-area-by-event and area-by-quarter-by-event fixed effects, as well as interactions between the deceased physician's baseline patient share and time fixed effects. The dashed vertical line denotes zero. Vertical bars indicate 95 percent confidence intervals based on standard errors clustered at the hospital level.

**Figure 5:** Effects on Quality Composition of Physicians:
(Heterogeneity by Deceased's Tenure )




*Notes:* This figure reports difference-in-differences estimates on the quality composition of physicians separately by the deceased physician's tenure at the hospital. The outcome is the patient-weighted average of physicians' observational survival in that hospital-area-quarter, where each physician's contribution is weighted by their share of patients treated in that period. Tenure is measured as the number of years the physician had been practicing at the hospital at the time of death and grouped into the indicated categories. All specifications include hospital-area-by-event and area-by-quarter-by-event fixed effects, as well as interactions between the deceased physician's baseline patient share and time fixed effects. Bars indicate point estimates, and vertical lines denote 95 percent confidence intervals based on standard errors clustered at the hospital level.

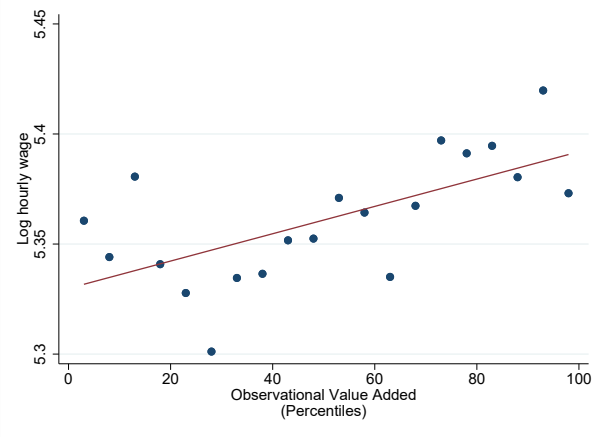
**Figure 6:** A Model of Patient Allocation with Two Physicians
(Non-binding Quality Constraint)



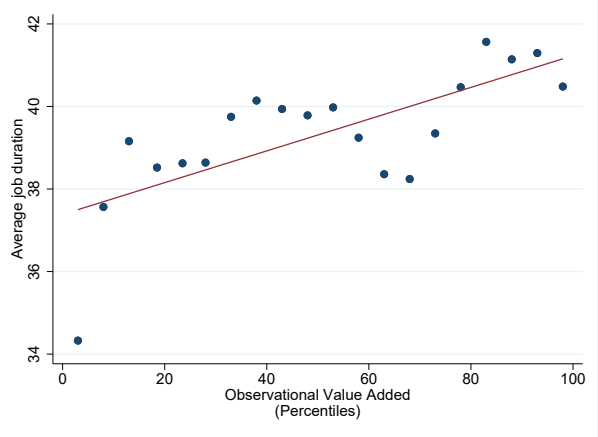

*Notes:* This figure provides a stylized illustration of the allocation problem described in Section 6. It plots the net marginal benefit of assigning an additional patient to a high-ability physician and to a low-ability physician as a function of caseload. Net marginal benefits decline with caseload due to convex effort costs. When the minimum-quality constraint is non-binding, optimal allocation equalizes net marginal benefits across physicians, yielding positive caseloads for both high- and low-ability types. The figure illustrates the resulting "fill-the-top-first" rule: patients are allocated to higher-ability physicians until rising marginal costs make further concentration unattractive. The horizontal dashed lines mark the equilibrium caseloads for each physician.

**Figure 7:** Observational Value Added and Labor Market Outcomes

**Panel A.** Earnings




**Panel B.** Job Duration




*Notes:* This figure plots physician-level labor market outcomes against percentiles of the observational value-added distribution. Observational value added is aggregated to the physician level by taking a patient-weighted average across all areas of care and then transformed into percentiles based on 50 equal-sized bins. Panel A plots average log hourly earnings, where earnings are computed as total labor income across all employing establishments in Brazil's employer–employee matched registry divided by total hours worked. Panel B plots average job duration, defined as the mean length (in quarters) of a physician's employment spells across hospitals. Each point represents the mean outcome within a value-added percentile bin, and the solid line shows the fitted linear relationship.

**Table 1:** Effects on Quality Composition of Physicians

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Quality shock $\times$ $\mathbb{1}${post event} | -0.381 | -0.384 | -0.381 | -0.380 | -0.385 |
| | [0.037]*** | [0.037]*** | [0.035]*** | [0.036]*** | [0.035]*** |
| Quarter-year FE | | | | | |
| $\times$ Pre-period physician composition | | ✓ | | | ✓ |
| $\times$ Pre-period case mix | | | ✓ | | ✓ |
| $\times$ Pre-period hospital characteristics | | | | ✓ | ✓ |
| Basic controls | ✓ | ✓ | ✓ | ✓ | ✓ |
| Pre-period mean | -0.001 | -0.001 | -0.001 | -0.001 | -0.001 |
| Observations | 2,469,143 | 2,468,823 | 2,469,143 | 2,468,210 | 2,467,890 |

*Notes:* The dependent variable is the patient-weighted average of physicians' observational survival in a given hospital–area–quarter, computed over the full set of treating physicians. The key regressor is the interaction between the turnover-induced quality shock and an indicator for the post-event period. All specifications exclude the quarter of the death event to avoid partial exposure. Column (1) includes the basic controls: hospital-area-by-event and area-by-quarter-by-event fixed effects, as well as interactions between the deceased physician's baseline patient share and time fixed effects. Columns (2)–(4) add interactions between time effects and pre-period physician, patient, and hospital characteristics, respectively, while Column (5) includes all such interactions simultaneously. Standard errors are clustered at the hospital level.

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

**Table 2:** Forecast Coefficient: Effects on Relized Survival

| | Reduced-Form | | IV | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Observational survival | | | 1.133 | 1.140 |
| | | | [0.095]*** | [0.090]*** |
| Quality shock $\times$ $\mathbb{1}$\{post event\} | -0.432 | -0.434 | | |
| | [0.050]*** | [0.048]*** | | |
| Kleibergen-Paap $F$-statistics | | | 107.49 | 107.49 |
| Risk adjustment | | ✓ | | ✓ |
| Basic controls | ✓ | ✓ | ✓ | ✓ |
| Pre-period mean | 0.948 | 0.948 | 0.948 | 0.948 |
| Observations | 2,470,122 | 2,470,122 | 2,415,174 | 2,415,174 |

*Notes:* The dependent variable is realized patient survival at the hospital–area–quarter level. Columns 1–2 report reduced-form estimates of the effect of the turnover-induced quality shock on realized survival. Columns 3–4 report instrumental-variables estimates of the forecast coefficient, where the patient-weighted average of physicians' observational survival is instrumented using the turnover-induced quality shock. All specifications exclude the quarter of the death event to avoid partial exposure. The basic controls include hospital-area-by-event and area-by-quarter-by-event fixed effects, as well as interactions between the deceased physician's baseline patient share and time fixed effects. The regressions are weighted by the number of patients in each hospital-area-quarter cell. The Kleibergen–Paap $F$-statistic reports the strength of the first stage. Standard errors are clustered at the hospital level.

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

**Table 3:** Decompositing Estimated Adjustment:
Internal versus External Replacement

| | Total adjustment (1) | Internal reallocation (2) | External replacements (3) |
|---|---|---|---|
| Quality shock $\times$ $\mathbb{1}$\{post event\} | 0.648 | 0.644 | 0.006 |
| | [0.053]*** | [0.056]*** | [0.019] |
| Basic controls | ✓ | ✓ | ✓ |
| Pre-period mean | -0.001 | -0.001 | 0 |
| Observations | 2,461,168 | 2,461,168 | 2,461,168 |

*Notes:* This table decomposes the mitigating response to turnover-induced quality shocks into internal reallocation among incumbent physicians and external replacement through newly hired physicians. The dependent variable in column (1) is the total adjustment in observational value added among alive physicians in a hospital–area–quarter following an unexpected physician death. Column (2) uses as the outcome the contribution of incumbent physicians to observational value added, while Column (3) uses the analogous contribution of newly hired physicians. All specifications exclude the quarter of the death event to avoid partial exposure. The basic controls include hospital-area-by-event and area-by-quarter-by-event fixed effects, as well as interactions between the deceased physician's baseline patient share and time fixed effects. Standard errors are clustered at the hospital level.
* $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

**Table 4:** Effects on Value Added:
(Heterogeneity by Adoption of Formal Quality Practices)

| | Negative quality shocks | | | Positive quality shocks | | |
|---|---|---|---|---|---|---|
| | Formal quality control practices? | | | Formal quality control practices? | | |
| | Established | Partial | Limited | Established | Partial | Limited |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Quality shock $\times$ 𝟙{post event} | 0.111 | -0.151 | -0.217 | -0.529 | -0.585 | -0.493 |
| | [0.191] | [0.066]** | [0.097]** | [0.157]*** | [0.088]*** | [0.112]*** |
| Basic controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Pre-period mean | 0.003 | -0.003 | -0.002 | 0.004 | -0.003 | -0.002 |
| Observations | 385,055 | 1,542,211 | 445,907 | 379,365 | 1,524,446 | 443,682 |

*Notes:* This table reports difference-in-differences estimates of the effect of turnover-induced quality shocks on the quality composition of physicians, allowing the response to vary by hospitals' adoption of formal quality-control practices in the pre-event period. Hospitals are classified as having established, partial, or limited formal practices based on whether they report the presence of key oversight committees for mortality and biopsy review, infection control, and cost accounting prior to the event. Columns (1)–(3) focus on negative quality shocks arising from the exit of relatively high-performing physicians, while columns (4)–(6) focus on positive quality shocks arising from the exit of relatively low-performing physicians. The dependent variable is the patient-weighted average observational survival of physicians in a hospital–area–quarter. All specifications exclude the quarter of the death event to avoid partial exposure. The basic controls include hospital-area-by-event and area-by-quarter-by-event fixed effects, as well as interactions between the deceased physician's baseline patient share and time fixed effects. Standard errors are clustered at the hospital level.

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

**Table 5:** Effects on Value Added:
(Heterogeneity by Baseline Caseload)

| | Negative quality shocks | | Positive quality shocks | |
|---|---|---|---|---|
| | Baseline caseload | | Baseline caseload | |
| | Low | High | Low | High |
| | (1) | (2) | (3) | (4) |
| Quality shock $\times$ $\mathbb{1}$\{post event\} | 0.012 | -0.244 | -0.677 | -0.427 |
| | [0.117] | [0.053]*** | [0.079]*** | [0.075]*** |
| Basic controls | ✓ | ✓ | ✓ | ✓ |
| Pre-period mean | -0.002 | 0 | -0.002 | 0 |
| Observations | 1,194,863 | 1,178,310 | 1,177,812 | 1,169,681 |

*Notes:* This table reports difference-in-differences estimates of the effect of turnover-induced quality shocks on the patient-weighted average physician value added, separately by baseline workload. Hospitals are classified as having low or high baseline caseload based on whether pre-event patient volume per physician in the hospital–area is below or above the median. Columns (1)–(2) report results for negative quality shocks, corresponding to the exit of high-performing physicians, while Columns (3)–(4) report results for positive quality shocks, corresponding to the exit of low-performing physicians. All specifications exclude the quarter of the death event to avoid partial exposure. The basic controls include hospital-area-by-event and area-by-quarter-by-event fixed effects, as well as interactions between the deceased physician's baseline patient share and time fixed effects. Standard errors are clustered at the hospital level.
* $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

# Online Appendix to "Allocating Talent in Organizations: Evidence from Quasi-Random Exits of Physicians"

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

April 27, 2026

**List of Figures** 1

**List of Tables** 2

**A Additional Tables and Figures** 3

**B Data Sources and Construction** 19

**C Empirical Bayes Correction** 21

**D Clinical Deviation Score** 22

**References** 23

## List of Figures

A.1 Timing of Deaths . . . . . 3
A.2 Observational Value-Added: Deceased versus Alive Physicians . . . . . 4
A.3 Falsification Test for Common Forecast Shocks: Effects on Residualized Index of Physician Characteristics . . . . . 5
A.4 Effects on Quality Composition of Physicians: Exploiting Within Hospital Variation . . . . . 6

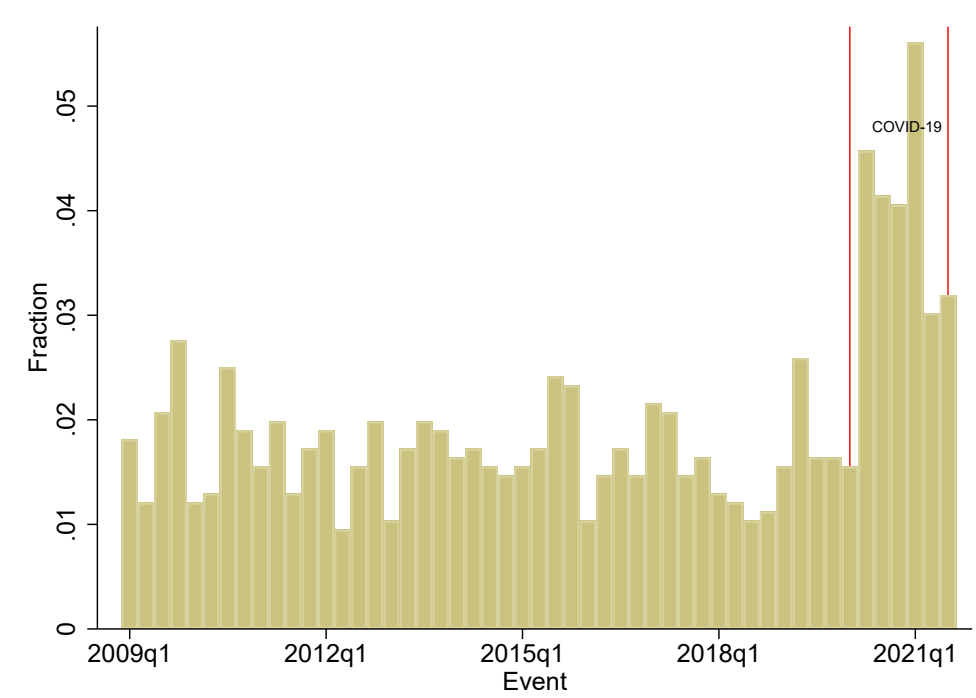
A.5 Effects on Quality Composition of Physicians: Matched Controls Outside State . . 7
A.6 Effects on Overall Volume . . . . . 8
A.7 Effects on Patient Characteristics: (No Evidence of Compositional Changes) . . . . 9

## List of Tables

A.1 Alive Versus Deceased Physicians . . . . . 11
A.2 Summary Statistics: Hospital-Level . . . . . 12
A.3 Effects on Quality Composition of Physicians: Further Sample Restrictions . . . . . 13
A.4 Effects on Quality Composition of Physicians: Robust Inference . . . . . 14
A.5 Effects on Quality Composition of Physicians: Alternative Constructions of Observational Value-Added . . . . . 15
A.6 IV Estimates of Observational Value-Added on Clinical Inputs . . . . . 16
A.7 Effects on Incumbents' and New Hires Contributions: (heterogeneity by Labor Market Thickness) . . . . . 17
A.8 Effects on Quality Composition of Physicians: Observational Value after Removing Specialty and Experience Effects . . . . . 18

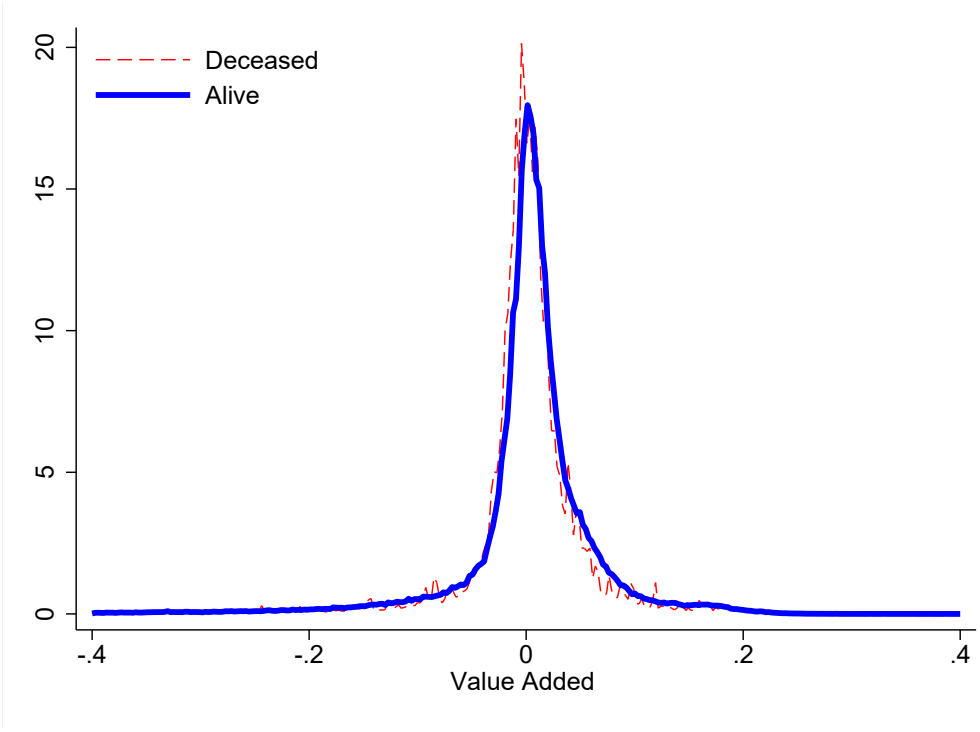
# A Additional Tables and Figures

**Figure A.1:** Timing of Deaths



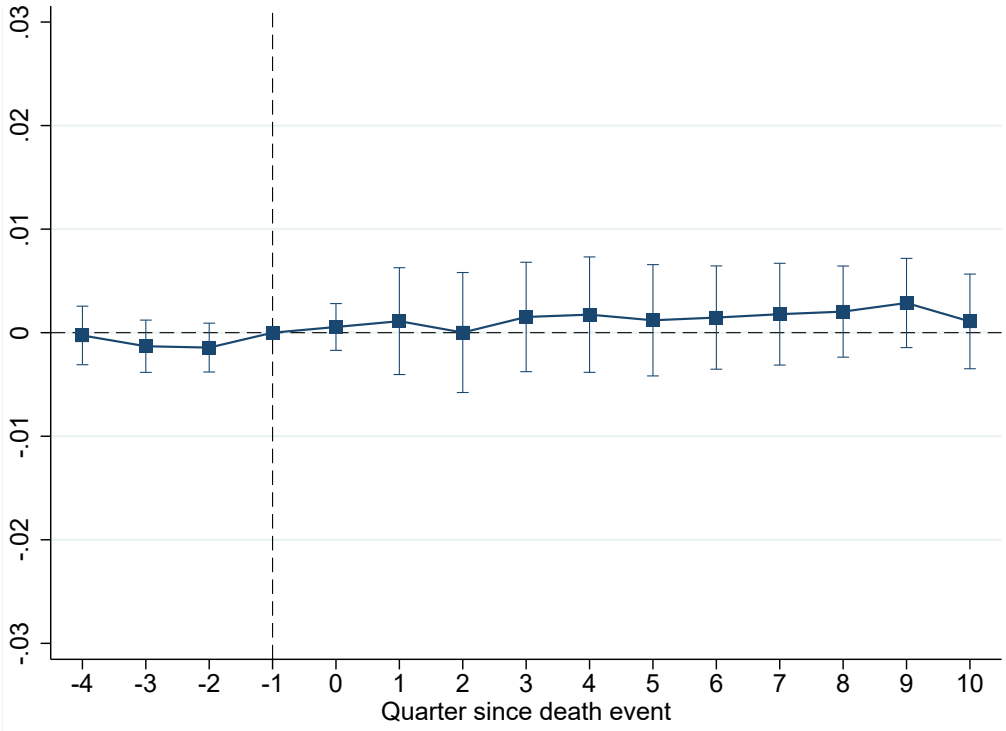

*Notes:* This figure plots the distribution of unexpected physician deaths over time at the quarterly level. Vertical red lines indicate the onset and duration of the COVID-19 pandemic. The figure shows that death events are broadly distributed over time, with a visible spike during the pandemic period.

**Figure A.2:** Observational Value-Added:
Deceased versus Alive Physicians



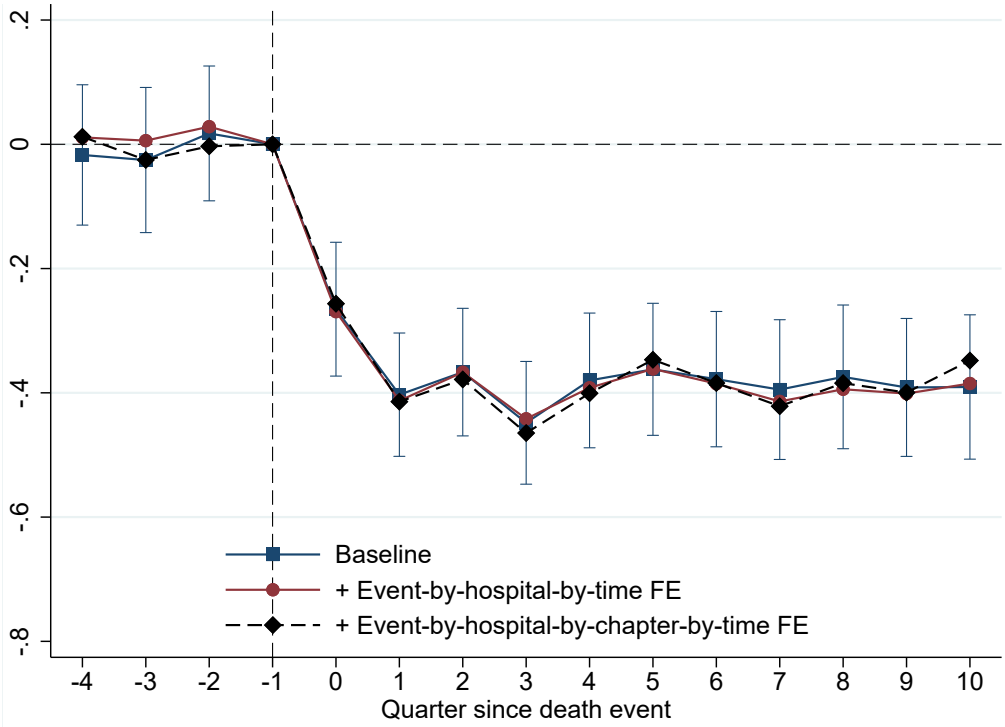

*Notes:* This figure compares the distribution of physicians' observational value added for deceased and non-deceased (alive) physicians. Observational value added is measured by risk-adjusted patient survival.

**Figure A.3:** Falsification Test for Common Forecast Shocks:
Effects on Residualized Index of Physician Characteristics



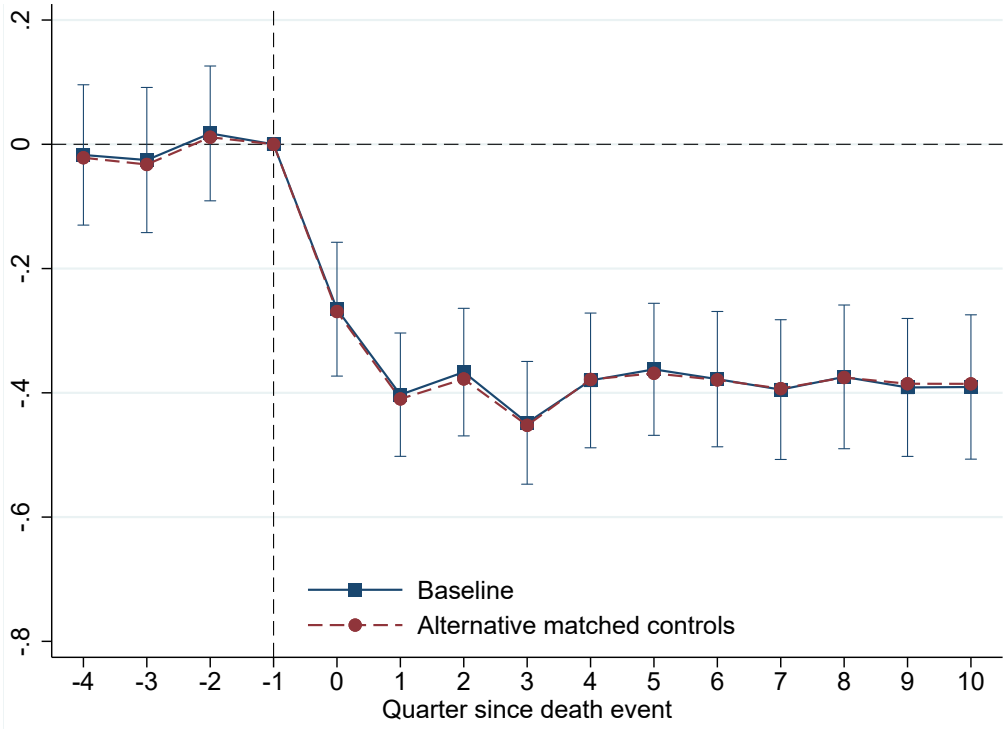

*Notes:* This figure reports event-study estimates from a falsification test for differential forecast shocks. The outcome is a patient-weighted average of a composite index of physician characteristics, constructed by first residualizing each component with respect to physicians' observational survival and then standardizing and averaging them. The index aggregates residualized measures of medical school, specialty, experience, caseload, and practice style (risk-adjusted spending and length of stay). The horizontal axis indexes quarters relative to the death event, with coefficients normalized to zero in the quarter immediately prior to the event. The specification includes hospital-area-by-event and area-by-quarter-by-event fixed effects, as well as interactions between the deceased physician's baseline patient share and time fixed effects. Vertical bars indicate 95 percent confidence intervals based on standard errors clustered at the hospital level.

**Figure A.4:** Effects on Quality Composition of Physicians: Exploiting Within Hospital Variation



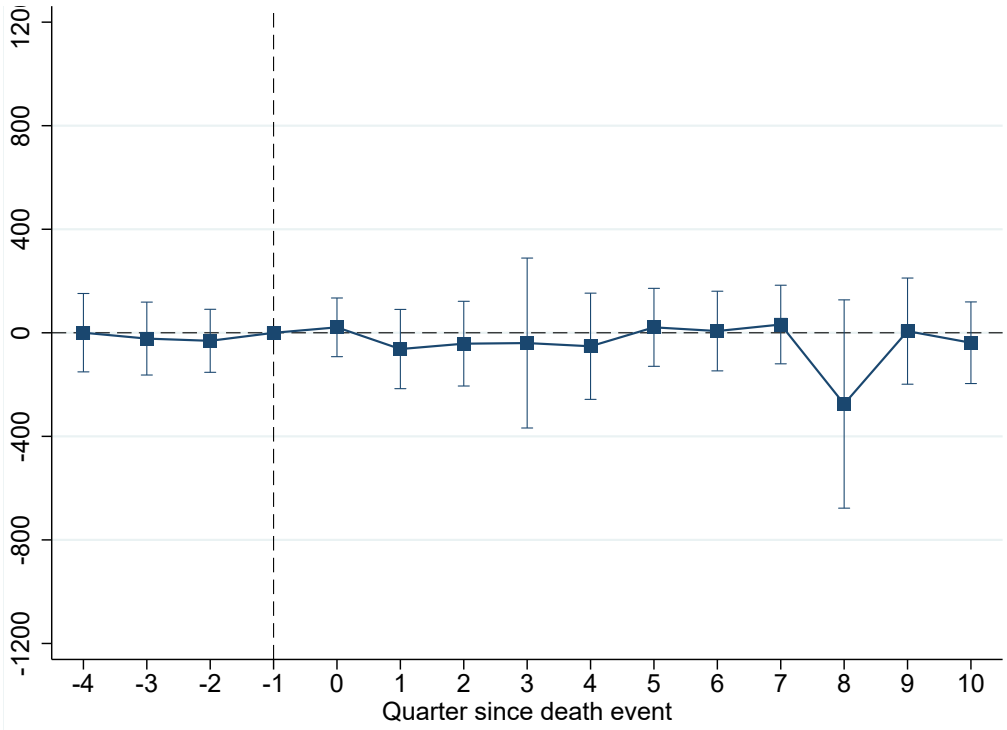

*Notes:* This figure reports robustness of the baseline event-study to alternative fixed-effects structures designed to assess within-hospital spillovers. The outcome is the patient-weighted average of physicians' observational survival in a given hospital–area–quarter. The baseline specification includes hospital–area–by–event and area-by-quarter–by–event fixed effects, as well as interactions between the deceased physician's base line patient share and time fixed effects. The second specification replaces the baseline set of fixed effects with a full set of hospital-by-quarter-by-event fixed effects. The third specification includes instead event-by-hospital-by-chapter-by-time fixed effects, where chapters correspond to the broadest ICD-10 classification grouping clinically related areas of care. Coefficients are normalized to zero in the quarter immediately prior to the event. Vertical bars indicate 95 percent confidence intervals based on standard errors clustered at the hospital level.

**Figure A.5:** Effects on Quality Composition of Physicians: Matched Controls Outside State



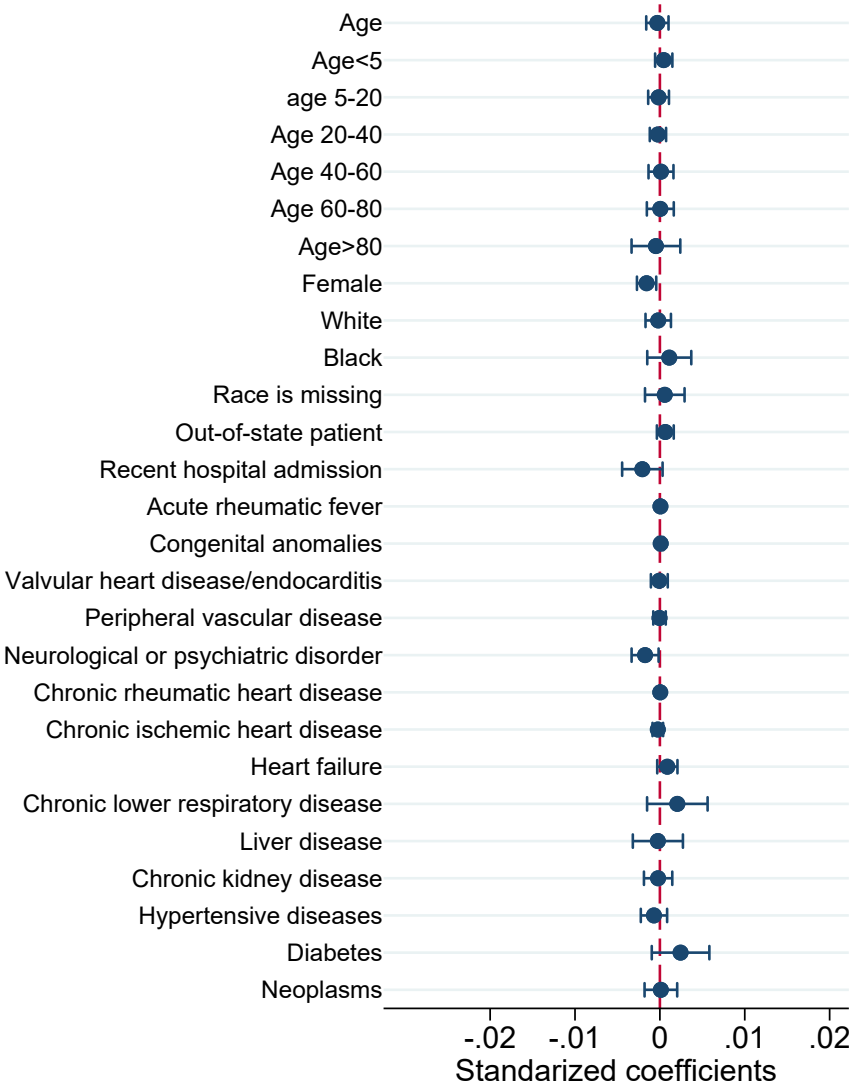

*Notes:* This figure replicates the baseline event-study specification using an alternative matched control group in which control hospitals are required to be located outside the state of the treated hospital. The solid line reports the baseline estimates using the original matched controls, while the dashed line reports estimates using the restricted out-of-state controls. The outcome is the patient-weighted average of physicians' observational survival in a hospital–area–quarter. Coefficients are normalized to zero in the quarter immediately prior to the death event. All specifications include the same fixed effects and controls as in the baseline analysis. Vertical bars indicate 95 percent confidence intervals based on standard errors clustered at the hospital level.

**Figure A.6:** Effects on Overall Volume



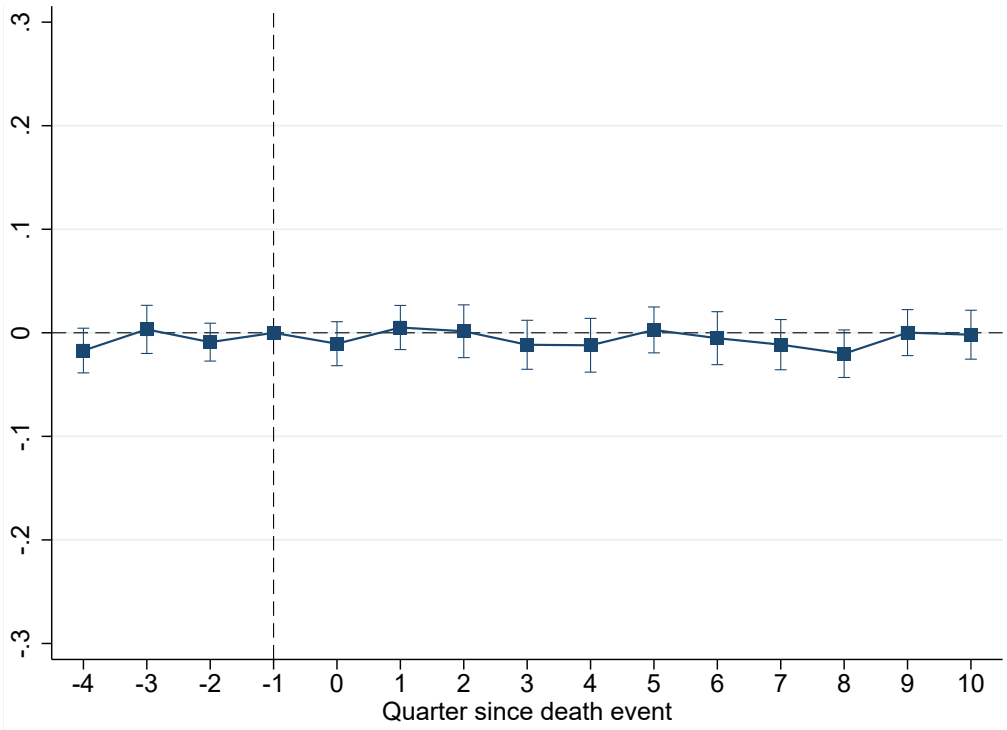

*Notes:* This figure plots event-study estimates of the effect of turnover-induced quality shocks on total patient volume in a hospital–area of care. The outcome is the total number of admitted patients per hospital–area–quarter. Coefficients are normalized to zero in the quarter immediately prior to the death event. All specifications include the same fixed effects and controls as in the baseline analysis. Vertical bars indicate 95 percent confidence intervals based on standard errors clustered at the hospital level.

**Figure A.7:** Effects on Patient Characteristics:
(No Evidence of Compositional Changes)




*Notes:* This figure reports difference-in-differences estimates of the effect of the turnover-induced quality shock on observable patient characteristics. Each point corresponds to a separate regression in which the dependent variable is a patient characteristic measured at the hospital–area–quarter level and standardized to have mean zero and unit variance. The specification includes the baseline set of controls and fixed effects used in the main analysis. Coefficients are normalized relative to the quarter immediately preceding the death event. The observations are weighted by the number of patients in a area-by-hospital-quarter-by-event cell. The dashed vertical line denotes zero. Vertical bars indicate 95 percent confidence intervals based on standard errors clustered at the hospital level.

**Figure A.8:** Effects on Predicted Survival




*Notes:* This figure plots event-study estimates of the effect of the turnover-induced quality shock on predicted patient survival. Predicted survival is constructed by projecting realized survival on observable patient characteristics (age, gender, race, and comorbidities) and aggregating the resulting predictions into patient-weighted averages at the hospital–area–quarter level. The horizontal axis indexes quarters relative to the death event, with coefficients normalized to zero in the quarter immediately prior to the event. All specifications include the same set of basic controls and fixed effects as in the main analysis. Vertical bars indicate 95 percent confidence intervals based on standard errors clustered at the hospital level.

**Table A.1:** Alive Versus Deceased Physicians

| | Deceased | | Alive | |
|---|---|---|---|---|
| | Final sample | Other | Final sample | Other |
| | (1) | (2) | (3) | (4) |
| Male | 0.865 | 0.841 | 0.631 | 0.613 |
| Experience | 29.52 | 31.73 | 17.70 | 17.45 |
| Top specialties | | | | |
| Internal Medicine | 0.349 | 0.367 | 0.182 | 0.322 |
| Obstetrics/Gynecology | 0.170 | 0.141 | 0.124 | 0.126 |
| Pediatrics | 0.067 | 0.107 | 0.119 | 0.118 |
| General Surgery | 0.121 | 0.105 | 0.118 | 0.084 |
| Orthopedics | 0.073 | 0.064 | 0.093 | 0.062 |
| Cardiology | 0.020 | 0.025 | 0.033 | 0.031 |
| Urology | 0.016 | 0.016 | 0.026 | 0.017 |
| Number of physicians | 1,021 | 7,177 | 117,839 | 189,882 |

*Notes:* This table compares observable characteristics of deceased and surviving physicians. Columns (1)–(2) report averages for deceased physicians, distinguishing those included in the final estimation sample from other deceased physicians. Columns (3)–(4) report analogous statistics for surviving physicians. Experience is measured as years since medical school graduation. "Top specialties" report the share of physicians in each specialty.
* $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

**Table A.2:** Summary Statistics:
Hospital-Level

| | Treated hospitals | Matched hospitals | Other untreated |
|---|---|---|---|
| | (1) | (2) | (3) |
| Number of FTE physicians | 98.07 | 67.80 | 23.57 |
| Number of FTE nurses | 65.06 | 53.06 | 19.05 |
| Inpatient mortality rate | 0.049 | 0.043 | 0.033 |
| Volume | 1649.80 | 1502.57 | 590.08 |
| Specialized hospital | 0.113 | 0.148 | 0.112 |
| Teaching hospital | 0.286 | 0.266 | 0.107 |
| Emergency department | 0.919 | 0.923 | 0.872 |
| Infraestructure indexes | | | |
| Imaging devices | 1.91 | 1.42 | 0.668 |
| Non-clinical infrastructure | 0.530 | 0.400 | 0.270 |
| Optical diagnostic equipment | 1.29 | 0.852 | 0.413 |
| Electrodiagnostic devices | 1.05 | 0.626 | 0.391 |
| Vital support devices | 4.08 | 3.26 | 1.23 |
| Other | 0.582 | 0.494 | 0.160 |
| Ownership | | | |
| Not-for-profit | 0.403 | 0.406 | 0.328 |
| For-profit | 0.080 | 0.085 | 0.155 |
| Public | 0.516 | 0.450 | 0.502 |
| Beds | 194.40 | 167.47 | 80.61 |
| Hospital-event pairs | 1,159 | 1,159 | 6,575 |

*Notes:* This table reports hospital-level summary statistics for treated hospitals, matched control hospitals, and other untreated hospitals. Treated hospitals are those experiencing the unexpected death of a physician during the sample period. Matched hospitals are selected using a two-step matching procedure as described in Sectio 2.

**Table A.3:** Effects on Quality Composition of Physicians: Further Sample Restrictions

| | Baseline | Exclude COVID | Exclude multiple simultaneous deaths | Exclude overlapping treatment periods |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Quality shock $\times$ $\mathbb{1}$\{post event\} | -0.381 | -0.394 | -0.393 | -0.394 |
| | [0.037]*** | [0.050]*** | [0.036]*** | [0.039]*** |
| Basic controls | ✓ | ✓ | ✓ | ✓ |
| Pre-period mean | -0.001 | 0 | -0.001 | -0.001 |
| Observations | 2,469,143 | 1,888,781 | 2,166,667 | 1,544,430 |

*Notes:* This table explores the robustness of the baseline estimates to further sample restrictions. Column (1) replicates the baseline specification. Column (2) excludes observations from the COVID-19 period. Column (3) excludes hospital–event cases with multiple physician deaths occurring in the same quarter. Column (4) excludes hospital–event cases with overlapping pre- and post-event windows arising from multiple deaths in close succession. Standard errors are clustered at the hospital level.

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

**Table A.4:** Effects on Quality Composition of Physicians: Robust Inference

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Quality shock $\times$ $\mathbb{1}$\{post event\} | -0.381 | -0.381 | -0.381 | -0.381 |
| | [0.037]*** | [0.039]*** | [0.041]*** | [0.039]*** |
| Clustering on | | | | |
| Hospital | ✓ | | | |
| Pair matching | | ✓ | | |
| Municipality | | | ✓ | |
| Microregion | | | | ✓ |
| Basic controls | ✓ | ✓ | ✓ | ✓ |
| Pre-period mean | -0.001 | -0.001 | -0.001 | -0.001 |
| Observations | 2,469,143 | 2,469,143 | 2,469,143 | 2,469,143 |

*Notes:* This table assesses the robustness of statistical inference to alternative clustering levels. All columns report estimates from the baseline specification. only the level at which standard errors are clustered varies across columns. Column (1) clusters standard errors at the hospital level. Columns (2)–(4) cluster at the matched hospital-pair, municipality, and microregion levels, respectively. The dependent variable is the patient-weighted average observational survival of physicians in a hospital–area–quarter. All specifications include the full set of baseline controls and fixed effects. Standard errors are reported in brackets. * $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

**Table A.5:** Effects on Quality Composition of Physicians: Alternative Constructions of Observational Value-Added

| | Baseline (1) | No Bayes (2) | No case-mix (3) | No Bayes, no case-mix (4) |
|---|---|---|---|---|
| Quality shock $\times$ $\mathbb{1}$\{post event\} | -0.381 | -0.356 | -0.358 | -0.359 |
| | [0.037]*** | [0.038]*** | [0.038]*** | [0.038]*** |
| Basic controls | ✓ | ✓ | ✓ | ✓ |
| Pre-period mean | -0.001 | -0.023 | -0.006 | -0.006 |
| Observations | 2,469,143 | 2,470,122 | 2,469,143 | 2,470,122 |

*Notes:* This table assesses the robustness of the baseline estimates to alternative constructions of the observational value-added measure. Column (1) reports the baseline specification. Column (2) removes the empirical Bayes shrinkage and uses raw risk-adjusted survival differences across physicians. Column (3) constructs observational value added without risk adjustment, and Column (4) removes both empirical Bayes shrinkage and risk adjustment. All specifications include the same set of basic controls and stacked fixed effects as in the baseline analysis. Standard errors are reported in brackets.

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

**Table A.6:** IV Estimates of Observational Value-Added on Clinical Inputs

| | Clinical deviation score (1) | Length of stay (2) | ICU admission (3) | Total spending (4) |
|---|---|---|---|---|
| Observational survival | 0.262 | -2.852 | -0.374 | -1.459 |
| | [0.120]** | [3.070] | [0.143]*** | [0.451]*** |
| Kleibergen-Paap *F*-statistics | 107.49 | 107.49 | 107.49 | 107.59 |
| Risk adjustment | ✓ | ✓ | ✓ | ✓ |
| Basic controls | ✓ | ✓ | ✓ | ✓ |
| Pre-period mean | 0.451 | 5.852 | 0.087 | 6.983 |
| Observations | 2,415,174 | 2,415,174 | 2,415,174 | 2,412,924 |

*Notes:* This table reports instrumental-variable estimates of the effect of physicians' observational value added on clinical inputs. The endogenous regressor is observational survival, instrumented using the turnover-induced quality shock. The dependent variables capture different margins of clinical practice: a case-level clinical deviation score measuring departures from typical procedure patterns for the same three-digit ICD condition, length of stay, an indicator for ICU admission, and total inpatient spending. All specifications include the full set of basic controls and risk adjustment. Kleibergen–Paap F-statistics indicate a strong first stage. Standard errors are clustered at the hospital level. * $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

**Table A.7:** Effects on Incumbents' and New Hires Contributions: (heterogeneity by Labor Market Thickness)

| | Internal reallocation | | External replacements | |
|---|---|---|---|---|
| | Thin market | Thick market | Thin market | Thick market |
| | (1) | (2) | (3) | (4) |
| Quality shock $\times$ $\mathbb{1}$\{post event\} | 0.613 | 0.657 | -0.005 | 0.006 |
| | [0.122]*** | [0.047]*** | [0.044] | [0.022] |
| Basic controls | ✓ | ✓ | ✓ | ✓ |
| Pre-period mean | -0.001 | -0.002 | 0 | 0 |
| Observations | 929,044 | 1,532,124 | 929,044 | 1,532,124 |

*Notes:* This table reports difference-in-differences estimates that decompose the adjustment to turnover-induced quality shocks into internal reallocation among incumbent physicians and external replacement through new hires, separately for thin and thick physician labor markets. Labor markets are classified as thin or thick based on whether the number of physicians per capita in the microrregion is below or above the median. The dependent variables correspond to the incumbent and new-hire components of observational value added, constructed as patient-weighted averages as described in Section 5. All specifications include the baseline set of controls and the same fixed effects structure as in the main decomposition analysis. Standard errors are clustered at the hospital level. * $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

**Table A.8:** Effects on Quality Composition of Physicians:
Observational Value after Removing Specialty and Experience Effects

| | Baseline (1) | Remove specialty FE (2) | Remove experience FE (3) | Remove both specialty experience FE (4) |
|---|---|---|---|---|
| Quality shock $\times$ $\mathbb{1}$\{post event\} | -0.381 | -0.427 | -0.385 | -0.427 |
| | [0.037]*** | [0.040]*** | [0.039]*** | [0.042]*** |
| Basic controls | ✓ | ✓ | ✓ | ✓ |
| Pre-period mean | -0.001 | 0.002 | 0.005 | 0.001 |
| Observations | 2,469,143 | 2,469,143 | 2,456,998 | 2,456,998 |

*Notes:* This table re-estimates the baseline specification using alternative measures of observational value added that remove variation attributable to physicians' observable credentials. In column (2), observational value added is purged of specialty-by-area-of-care fixed effects; in column (3), it is purged of experience fixed effects; and in column (4), both components are removed simultaneously. The residualized measures therefore isolate within-specialty and within-experience variation in physicians' performance. All specifications follow the baseline difference-in-differences design and include the same set of basic controls. Standard errors are clustered at the hospital level. * $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

# B 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 chose 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 contain 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 in Section 2.3.

**Administrative Payroll Records.** Our analysis of physicians' labor market outcomes relies on Brazil's *Relação Anual de Informações Sociais* (RAIS) from 2003 to 2018, an annual employer–employee matched payroll registry maintained by the federal government. The RAIS covers the universe of formal employment relationships in Brazil, including both public and private sector jobs, and is based on mandatory reports that all registered employers must file each year. The registry is primarily used for administering and monitoring labor legislation and social insurance programs, but it has also become a central source of administrative data for labor market research.

The RAIS contains information on workers, including basic demographics (such as year of birth and gender), occupation codes, dates of hiring and separation, type of contract, weekly contracted hours, and wage payments. These variables allow us to construct measures of total labor earnings, job tenure, and work intensity for each physician across all formal jobs. We match the RAIS to our main physician-level dataset using the unique individual taxpayer identifier. It is important to note that while the RAIS provides detailed information on wages and other contract attributes at the establishment–worker level, the definition of an "employer" in the context of physicians is somewhat diffuse. Many healthcare establishments, including hospitals located

in different regions of the country, are grouped under a single employer entity in the RAIS, such as the Ministry of Health or a state-level health department. This administrative structure prevents us from precisely identifying the specific hospital establishment where a physician works, precluding a more granular analysis at the physician–hospital level.

# C Empirical Bayes Correction

We estimate differences in observational survival rates across physicians operating within the same area of care, accounting for differences in baseline case mix. We do so by estimating the following patient-level regression of survival outcomes on a full set of physician-by-subchapter fixed effects and rich patient characteristics:

$$\text{Survival}_{it} = \theta_{js} + \mathbf{X}_{it}'\Theta + \varepsilon_{it},$$

where $Survival_i$ is an indicator for whether the patient survived the hospitalization, and $\mathbf{X}_{it}$ is a vector of pre-determined covariates. We include patient demographic and clinical characteristics, including age, gender, race, residential zipcode, history of recent hospital admissions, and a complete set of comorbidity indicators. We also control for hospital and quarter-by-year fixed effects to absorb time-invariant differences across hospitals and common shocks affecting all hospitals in a given period. The parameter of interest is $\theta_{js}$, which represents risk-adjusted physician survival rates specific to each area of care. We normalize estimates $\theta_{js}^*$ to have a mean of zero within each area of care to facilitate interpretation and within area comparability.

One challenge in estimating $\theta_{js}$ is that physicians differ substantially in the number of admissions they handle within a given area of care, which introduces sampling error and heterogeneity in the reliability of the estimated parameters. To account for measurement error, we construct an empirical Bayes estimator that adjusts the estimated physician effects according to the standard shrinkage formula (Morris, 1983; Kane and Staiger, 2008):

$$\left[\frac{\sigma^2_{\theta(s)}}{\sigma^2_{\theta(s)} + \sigma^2_{\epsilon(s)}/n_{js}}\right]\hat{\theta}_{js}$$

where $n_{js}$ denotes the number of patients treated by physician $j$ in area $s$. The term $\sigma^2_{\theta(s)}$ represents the between-physician variance in adjusted survival specific to each area of care, while $\sigma^2_{\varepsilon(s)}$ is the within-physician-area variance of survival outcomes. This procedure shrinks noisy estimates toward the area-level mean, improving precision without altering the interpretation of $\theta_{js}$.

Since the fixed-effect coefficients $\theta_{js}^*$ are measured with error, their conventional variance overstates the true cross-physician dispersion. To correct for this estimation error, we employ a split-sample approach following Kane et al. (2013), Chetty et al. (2014a) and Silver (2021). We first compute the partial residuals $r_{it} = \text{Survival}_{it} - \mathbf{X}_{it}'\hat{\Theta}$ and randomly split the sample halves

$A$ and $B$, stratifying by physician-area pair. For each subsample, we average the residuals by physician and service area to obtain $\bar{r}_{js}^A$ and $\bar{r}_{js}^B$. Let $e_{js}^A$ and $e_{js}^B$ denote the respective the respective estimation errors. Assuming these errors are independent across splits, the covariance between $\bar{r}_{js}^A$ and $\bar{r}_{js}^B$ provides a consistent estimate of the variance of $\theta_{js}$ within a service area:

$$\begin{aligned}
\text{Cov}(\bar{r}_{js}^A, \bar{r}_{js}^B) &= \text{Cov}(\theta_{js} + e_{js}^A,\ \theta_{js} + e_{js}^B) \\
&= \underbrace{\text{Cov}(\theta_{js}, \theta_{js})}_{=\sigma^2_{\theta(s)}} + \underbrace{\text{Cov}(e_{js}^A, \theta_{js})}_{=0} + \underbrace{\text{Cov}(\theta_{js}, e_{js}^B)}_{=0} + \underbrace{\text{Cov}(e_{js}^A, e_{js}^B)}_{=0} \\
&= \sigma^2_{\theta(s)}
\end{aligned}$$

To estimate $\sigma^2_{\varepsilon(s)}$, we use estimates of the residuals $\hat{\varepsilon}_{it}$ and compute the within-physician-area variance of the patient-level residuals, weighting by the number of observations per physician. Specifically, we calculate:

$$\hat{\sigma}^2_{\varepsilon(s)} = \frac{\sum_{j \in s} \sum_{i \in j} (\hat{\varepsilon}_{it} - \bar{\hat{\varepsilon}}_{js})^2}{\sum_{j \in s} (n_{js} - 1)},$$

where $\bar{\hat{\varepsilon}}js$ denotes the mean residual for physician $j$ in area $s$. This estimate captures the idiosyncratic component of survival outcomes that remains after controlling for patient characteristics and physician effects.

# D Clinical Deviation Score

We build a patient-level measure that captures how much the set of procedures performed in a given admission departs from typical practice for the same primary diagnosis. For each primary diagnosis $d$ based on the 3-digit ICD-10, we use the national universe of inpatient admissions to tabulate the frequency of every procedure code observed. Let $r_{kd}$ denote the national share of procedure $k$ among all procedures used for diagnosis $d$, normalized so that $\sum_{k \in \mathcal{K}_d} r_{kd} = 1$, where $\mathcal{K}_d$ is the full support of procedure codes appearing for diagnosis $d$. The vector $r_d$ therefore summarizes the modal procedure distribution for diagnosis $d$, without presuming that the typical bundle is clinically optimal.

For each admission $i$ with primary diagnosis $d$ treated by physician $j$, we define an indicator $x_{ikd} = 1$ if procedure $k \in \mathcal{K}_d$ is performed in that admission and $x_{ikd} = 0$ otherwise. Let $x_{id} = (x_{ikd})_{k \in \mathcal{K}_d}$ denote the resulting procedure-set vector for admission $i$, and let $r_d = (r_{kd})_{k \in \mathcal{K}_d}$ denote the corresponding modal procedure distribution. We measure the closeness between the admission's realized procedure set and modal practice using cosine similarity:

$$\text{Sim}_{id} = \frac{\sum_{k \in \mathcal{K}_d} x_{ikd}\, r_{kd}}{\sqrt{\sum_{k \in \mathcal{K}_d} x_{ikd}^2}\ \sqrt{\sum_{k \in \mathcal{K}_d} r_{kd}^2}}.$$

Because $x_{ikd}$ is binary, $\sum_k x_{ikd}^2 = \sum_k x_{ikd} \equiv m_{id}$ equals the number of distinct procedures performed in admission $i$. Thus the numerator is the modal-weighted overlap between the admission's procedure set and typical practice for diagnosis $d$, while the denominator normalizes by the overall intensity of procedures in the admission and by the concentration of modal practice. By construction, $\text{Sim}_{id}$ lies in $[0, 1]$. A value of $\text{Sim}_{id} = 1$ obtains when the admission's procedure set is perfectly aligned with the modal distribution for diagnosis $d$, whereas $\text{Sim}_{id} = 0$ indicates no overlap with typical practice.

We define the *clinical deviation score* as the complement of this similarity measure:

$$\text{Dev}_{id} = 1 - \text{Sim}_{id},$$

so higher values indicate that the admission's procedure set departs more from modal practice for that diagnosis. It is important to stress that $\text{Sim}_{id}$ is normalized by the overall number of distinct procedures realized in the admission. This ensures that the score does not mechanically changes with the sheer number of procedures recorded for a patient. Instead, conditional on intensity, this measures *composition* of the procedure bundle. Throughout, we interpret this clinical deviation score descriptively: it captures departures from typical practice patterns rather than adherence to clinical guidelines or appropriateness of care.

## References

**Chetty, Raj, John N Friedman, and Jonah E Rockoff**, "Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates," *American economic review*, 2014, *104* (9), 2593–2632.

**Kane, Thomas J. and Douglas O. Staiger**, "Estimating teacher impacts on student achievement: An experimental evaluation," Working Paper w14607, National Bureau of Economic Research 2008.

**Kane, Thomas J, Daniel F McCaffrey, Trey Miller, and Douglas O Staiger**, "Have we identified effective teachers? Validating measures of effective teaching using random assignment," *Research Paper. MET Project. Bill & Melinda Gates Foundation*, 2013.

**Morris, Carl N.**, "Parametric empirical Bayes inference: theory and applications," *Journal of the American Statistical Association*, 1983, *78* (381), 47–55.

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