Physicians' Occupational Licensing and the Quantity-Quality Trade-Off*

Juan Pablo Atal Tomás Larroucau Pablo Muñoz Cristóbal Otero

July 28, 2025

Abstract: Occupational licensing is a widespread quality regulation that increases the quality of labor but reduces its quantity. We provide a framework to empirically quantify this trade-off and apply it to physician licensing, where both quality and access to care are critical concerns. Using quasi-exogenous variation driven mostly by a recent and unprecedented migration of physicians to Chile, we show that more physicians improve access and patient outcomes in tertiary care, including mortality. We also find that lower quality—as measured by physician performance on the licensing exam—worsens patient outcomes. Building on these findings, we evaluate the implications of locally changing the stringency of the current licensing policy.

JEL Codes: I18, J24, J44, J45

*We thank Francesco Agostinelli, Esteban Aucejo, Hal Cole, Matt Grennan, Ashley Swanson, Matt Wiswall, Evan Soltas, and seminar participants at Chicago BFI Health Economics Conference, Fordham University, Johns Hopkins University, RIDGE Labor Forum, TSE Health Economics Conference, UCL, Princeton, Queen Mary University, Toulouse Health Economics Conference, FGV-EESP, Insper, PUC-Rio, Universidad de los Andes-Chile, PUC-Chile, Universidad Chile, and Universidad de Santiago for valuable comments and suggestions. Pablo Muñoz thanks FONDECYT Iniciación (ANID-FONDECYT-11230049) and the Millenium Nucleus MIGRA (ANID-MILENIO-NCS2022051) for financial support. Simon Andrade, Nicolas Bozzo, Wagner F. Oliveira, and Sofía Pontigo provided superb research assistance. We thank the Health Ministry of Chile and EUNACOM for data access. Atal: University of Pennsylvania and NBER, `ataljp@econ.upenn.edu`. Larroucau: Arizona State University, `tomas.larroucau@asu.edu`. Muñoz: Universidad de Chile, `pablomh@uchile.cl`. Otero: Columbia Business School, `c.otero@columbia.edu`.

1 Introduction

The shortage of physicians is a long-standing and increasingly urgent concern. Providing adequate healthcare access is estimated to require at least an additional 7 million physicians globally, and a deficit of 125 thousand physicians is projected by 2034 for the United States alone (Haakenstad et al., 2022; Association of American Medical Colleges, 2024).

Licensing requirements in the medical profession have often been identified as a significant contributor to physician shortages, with unclear benefits terms of quality improvements.[1] Indeed, assessing the quality-quantity trade-off embedded in licensing requirements has proven difficult, as it requires credible estimates of its effects on the quantity and quality of physicians and on the resulting impacts on access and quality of care. The increasing global migration of physicians has further heightened the impact of licensing policies, potentially altering both the nature and the magnitude of these trade-offs.[2]

In this paper, we provide a simple framework to highlight the key economic fundamentals that govern the quantity-quality trade-off when deciding the stringency of licensing requirements. From this framework, we derive a set of sufficient statistics that allow us to evaluate the optimality of the licensing requirements and to quantify the net benefits of locally changing them. Using rich administrative data from Chile and a quasi-experimental research design, we estimate those sufficient statistics and evaluate the optimality of the licensing policy that determines physicians' eligibility to practice.

Chile offers an ideal setting to study physician licensing and how its design may depend on changing labor market fundamentals, such as increased immigration. The number of physicians taking the licensing exam increased fivefold between 2009 and 2018, driven largely by an unprecedented rise in foreign-trained test-takers. In tandem with this surge, passing rates decreased from 92% to 58% during the same period. This migration shock generated rich variation, allowing for a credible estimation of the effects of the quantity and quality of physicians on patient outcomes. Also, revisiting Chile's licensing policy is crucial from a policy perspective, for three reasons. First, the migration wave has made licensing requirements increasingly binding, highlighting the need to

[1] A influential early critique of licensing in the medical profession appears in Friedman (1962), with a more recent review offered by Svorny (2004). While early discussions centered on health care, occupational licensing regulations extend far beyond the medical field: approximately 25% of all US workers are employed in occupations that require an occupational license (Dillender et al., 2023).

[2] The number of physicians born and educated outside the US and Canada represented around 20% of the total count of active physicians in the United States in 2021 (Association of American Medical Colleges, 2023). Similarly, in OECD countries, the share of foreign-trained physicians has increased over the past decade, with more than 18% of doctors in 2017/18 having obtained their first medical degree abroad (Socha-Dietrich and Dumont, 2021).

reassess their design. Second, the healthcare system remains strained, with 25% of the country's annual mortality attributed to individuals dying while on waiting lists.[3] Finally, Chile's licensing exam shares characteristics and objectives that align with those used in other countries (Mena, 2021).[4]

Our theoretical framework highlights the quantity-quality trade-off embedded in the design of licensing policies. We model hospital care with a production function whose primary inputs are the number and quality of physicians. The planner faces an exogenous distribution of physicians and has a licensing technology that provides a noisy signal of their quality. The planner grants licenses to those with a signal above a threshold. Changing the licensing threshold affects hospital outcomes depending on two sets of sufficient statistics: the output elasticities with respect to the quantity and quality of labor and the elasticity of those inputs with respect to the licensing threshold. The input elasticities, in turn, depend on the fraction of test-takers at the threshold, their score relative to the score of supra-marginal physicians, their probability of matching with the focal hospital, and the precision of the licensing score as a signal of quality (i.e., the signal-to-noise ratio).

Building on our theoretical framework, we estimate the sufficient statistics necessary to gauge how marginal changes in the licensing threshold would affect hospital outcomes in Chilean public hospitals. We first recover output elasticities with respect to both the quantity and the quality of physicians, using mortality and access to care as our primary outcomes. To address the endogeneity of inputs—pervasive in production-function estimation—we use shift-share instruments (Altonji and Card, 1989; Autor et al., 2013; Borusyak et al., 2022) in a two-stage least squares (2SLS) approach. The instruments for physicians quantity and quality leverage the change in both the quantity and average performance of licensing-exam takers, driven mainly by the unprecedented migration of physicians into the country and, to a lesser extent, by the domestic expansion of medical schools.

Our 2SLS estimates of the output elasticities reveal that both the quantity and quality of physicians matter for mortality. We find that decreasing the number of physicians by 1% increases the in-hospital death rate by 0.55%, while a one standard deviation decrease in average physicians' quality increases death rates by 21%. Together, these estimates imply that a 10% decrease in the quantity of physicians would increase mortality unless their quality improves by more than 0.26 standard deviations. Reassuringly, this "marginal rate of technical substitution" between quantity

[3] As of 2019, Chile had 17.5 physicians per 10,000 inhabitants—nearly half the average of countries with a comparable burden of diseases, injuries, and risk factors (33.4 physicians per 10,000). This figure also falls below the minimum threshold of 20.7 physicians per 10,000 needed to achieve an effective Universal Healthcare coverage score of 80 out of 100 at the global level proposed by Haakenstad et al. (2022). The shortage of physicians is also reflected in the current long waiting lists, with approximately 3 million individuals—15% of the population—awaiting medical attention. An estimated 40,000 individuals die annually while on waiting lists (MINSAL, 2025).

[4] Similar test are the USMLE in the US, the NCE in France, and the NMLE in Japan.

and quality implied by our 2SLS approach can also be transparently recovered from the reduced-form effects of the migration wave on mortality and on the quantity and quality of physicians. Regarding access to care, we estimate an elasticity of 0.99 for the service rate with respect to the number of physicians, indicating the more physicians rises patient volume nearly proportionally, while physician quality has no impact on patient volume. Complementary evidence from waiting lists further supports this result: exogenous increases in physician supply enable hospitals to clear both surgical and medical queues more rapidly.

We proceed by estimating the input elasticities that govern how both the quantity and quality of physician inputs respond to changes in the licensing threshold. To do so, we specify a dynamic model of score realizations and retake decisions, which maps each physician's observed history of scores into a latent quality measure. By embedding endogenous retaking behavior, the model allows us to recover a hospital's input quality under both realized and counterfactual score distributions, and to quantify the short-run and medium-run effects of changing the threshold. Our estimates reveal a high signal-to-noise ratio in test scores, implying that scores are informative signals of latent physician quality. As a result, adjustments to the licensing stringency can meaningfully shift the distribution of quality across hospitals. To estimate the resulting elasticities of the quantity and quality of labor across different hospitals, we combine the distribution of test takers and scores with a matching function between physicians and hospitals. Guided by theory, we approximate this labor market matching as a flexible hospital-specific function of physicians' latent quality and other observables, as well as of observable proxies for labor-market tightness. As such, our estimated labor elasticities account for the fact that the matching function can vary with counterfactual licensing thresholds, since labor market tightness depends on how many physicians pass the exam. Our estimated input elasticities reveal the quantity-quality trade-off for inputs at the margin. In the average hospital-year in our sample, an increase in the threshold that lowers quantity by 10% would increase average quality by 0.32 standard deviations.

Equipped with estimates for the sufficient statistics, we quantify the net benefits of marginally increasing the licensing threshold in the Chilean healthcare context. We begin by assessing the short-run effects of a counterfactual threshold increase in 2018—the final year of our sample and 5 years after the onset of the large migration wave. Our results imply that increasing the threshold would have decreased access, although it would have had positive effects on treatment quality measured by patient mortality. Using standard VSL estimates, we find that the reform would be cost-neutral if a hospital admission per year is valued at 1.6 times the yearly minimum wage.

We also investigate how the short-run effects differ across the different phases of the migration wave. Overall, we find qualitatively similar short-run effects of the policy throughout our sample period, despite the large changes in labor market fundamentals produced by the migration wave. Still, our results help us shed light on how licensing policies interact with labor market conditions. The magnitudes of policy effects are non-monotonic over time; the largest impacts would have occurred in the middle of the migration wave when the number of marginal physicians was large and the stock of physicians already hired was still relatively low. Finally, we use our model of scores and retaking behavior to assess the policy's dynamic ("medium-run") effects, taking into account that examinees can retake the exam. Because retaking allows scores to improve over successive attempts, it attenuates the impact of raising the licensing threshold. In fact, more than 80% of the candidates who failed their first attempt in 2013 passed by 2018. As a result, the effect of increasing the threshold wanes over time, although a permanent increase in the threshold in 2013 still yields measurable impacts in 2018.

Our paper contributes to several strands of the literature. We add to a long-standing debate about the desirability of occupational licensing in healthcare (Friedman and Kuznets, 1945; Friedman, 1962; Svorny, 2004; Kleiner, 2014; Kleiner et al., 2016; Xia, 2021; Dillender et al., 2023) and other settings, particularly in the public sector (Kleiner and Wang, 2023; Angrist and Guryan, 2008; Kleiner, 2011; Larsen et al., 2020). We make two contributions relative to this literature. First, while most of the occupational licensing literature in healthcare studies the impacts on workers, we focus on patient outcomes. Notable exceptions include studies on the introduction of licensing requirements for midwives in the early twentieth-century United States, which document significant reductions in maternal mortality and improvements in the long-term health of newborns (Anderson et al., 2020; Noghanibehambari and Fletcher, 2023). Also, Taek Oh and Kleiner (2025) show that the universal license recognition for out-of-state physicians improved access to care. Second, we provide a framework for understanding the quantity-quality trade-off embedded in licensing and empirically evaluate the stringency of licensing policies in relevant health outcomes. Close to our work, Kleiner and Soltas (2023) provide a general equilibrium framework to estimate the welfare effect of licensing policies in the US under competitive product and labor markets.[5] In their framework, labor quality is inferred from wages and equilibrium conditions in the labor market. In contrast, we use direct measures of key patient outcomes—namely, access and mortality—and develop a framework to assess the impact of licensing in a setting with rigid wages.

We also contribute to the literature by providing quasi-experimental evidence on the impact of physician quantity and quality on health outcomes. Previous studies have examined the rela-

[5]See also Atal et al. (2022) for a general-equilibrium analysis of the welfare effects of minimum quality standards in product markets.

tionship between physician availability and health outcomes, including the link between location-specific mortality and the number of physicians (Finkelstein et al., 2021), as well as the effect of primary care physician supply on infant health (Carrillo and Feres, 2019). In this paper, we provide direct evidence that physician quantity influences health outcomes in a tertiary care setting, leveraging quasi-exogenous variation for identification. In addition, we examine physician quality—proxied by licensing-exam performance—and its impact on patient health outcomes. Prior work has documented associations between patient outcomes and various quality proxies: elite medical training (Doyle et al., 2010), physician value-added measures (Fletcher et al., 2014; Ginja et al., 2024), medical school exit exams (Guarin et al., 2021), community-level experiments that replaced mid-level providers with physicians (Okeke, 2023), endowment-funded health care modernization (Hollingsworth et al., 2024), and medical school closures (Clay et al., 2025). However, to our knowledge, no study has causally linked licensing-exam performance to patient health outcomes.

Finally, our work also contributes to recent research on the impact of high-skill migration on receiving countries (Beerli et al., 2021; Brinatti, 2024; Dodini et al., 2025) and to the scarce literature estimating production functions and value added in healthcare (Gaynor et al., 2015; Grieco and McDevitt, 2017; Einav et al., 2025).

The remainder of the paper is organized as follows. Section 2 describes the setting and the data. In Section 3, we provide descriptive evidence to illustrate the main variation leveraged to assess the quantity-quality trade-off. Section 4 presents a stylized model of the licensing problem. In Sections 5 and 6, we present the model and estimates of the output and input elasticities. We present our main counterfactual exercises in Section 7 and conclude in Section 8.

2 Setting and Data

2.1 Institutional Setting

The Chilean Public Healthcare System The healthcare system in Chile is divided into public and private insurers and providers. Our focus is on public providers, which mainly serve patients with public insurance. Public insurance covers approximately 80% of the population and is financed through monthly contributions deducted from labor income, cost-sharing mechanisms, and resources from the general government.[6]

[6]Individuals also have the option to redirect their health contributions toward purchasing private insurance, which covers 15% of the population. The remainder 5% is covered through schemes exclusively for the police and armed forces. In practice, public insurance primarily serves the relatively more disadvantaged population, while wealthier, healthier, and younger individuals tend to opt for private insurance (Pardo, 2019).

Public health insurance provides coverage within the network of public providers, with varying levels of copayment determined by income and family size. Individuals who cannot afford to pay are granted free access to the public system, which ensures nearly universal health coverage in public hospitals. Beneficiaries of the public insurer may also opt for private providers, although the co-payments for such "out-of-network" services are significantly higher than those in the public network.

In our sample, the network of public hospitals comprises 176 hospitals that belong to one of 29 referral regions called Health Services (*Servicios de Salud*).[7] The referral and counter-referral system also operates at this level. Individuals must register with their local primary healthcare provider, and those requiring specialized care are referred to one of the public hospitals within their region. Referrals follow strict and predetermined guidelines based on diagnosis, location, and other patient demographics.[8] Hospitals are categorized into three levels of complexity—low, medium, and high—depending on their size and the range of medical services they offer.

Physician Licensing To work in the healthcare sector, physicians are required to pass the EUNACOM exam (*Examen Único Nacional de Conocimientos de Medicina*).[9] The exam was established in 2009 and is aligned with the characteristics and purposes of medical licensing exams in other countries.[10] It comprises a theoretical section and a practical section. The theoretical section consists of a multiple-choice test with 180 questions that cover various areas of medical knowledge. The score reflects the candidate's absolute performance and is not standardized relative to other test-takers.[11] To pass, candidates must achieve a score of 51 or higher out of 100 points.

Upon passing the theoretical section, candidates must complete a pass-or-fail practical section, which involves an examination in a real or simulated clinical environment. However, this requirement is waived for candidates with medical degrees from local universities and physicians with a

[7] The 176 hospitals corresponds to those that are present every year throughout our study period and report deaths at least in one year. We exclude psychiatric hospitals.

[8] Patients can also be admitted directly to hospitals through the Emergency Room in cases of emergency.

[9] Although the exam is not technically mandatory for physicians in private healthcare organizations, it is effectively binding across all healthcare institutions because passing it is a legal requirement in order to treat patients covered by public health insurance, regardless of the treatment location. This explains why, de facto, most private healthcare institutions only consider applications admissible if candidates include EUNACOM approval. Performance on the EUNACOM is also determinant for accessing training opportunities, as it plays a key role in admissions to residency programs in the country.

[10] Comparable exams include the Medical Licensing Examination (USMLE) in the US, the Medical Council of Canada Qualifying Examination (MCCQE) in Canada, the National Competency Exam (NCE) in France, the National Medical Licensing Examination (NMLE) in Japan, and the Korean Medical Licensing Examination (KMLE) in Korea. For an in-depth discussion of the exam and evidence of its validity and reliability, see Mena (2021).

[11] A key objective of EUNACOM designers is to make the difficulty of the exam comparable across years. Empirically, we observe that although the scores are not standardized, the distribution of scores at local universities is very similar across years.

medical degree from a select group of countries with which the Ministry of Health has established bilateral agreements.[12] The practical portion of the exam is largely not binding (Kunakov et al., 2018).

Importantly, passing the EUNACOM automatically validates the medical degrees of foreign-trained physicians, and grants them the same job and training opportunities as locally trained physicians.

Physicians' Wages and Hiring in the Public Sector: Physicians can work in both the private and public sectors. In the private sector, employment operates under standard market dynamics, whereby wages, benefits, and working conditions are negotiated directly between employers and employees. In contrast, in public hospitals, wages are legally regulated and follow a public-sector wage schedule, with annual adjustments based on sector-wide wage revisions.[13]

Hiring in public hospitals is decentralized and managed by the regional Health Services within the budgetary constraints established by the Ministry of Health. Health Services determine staffing needs for hospitals and hospital directors oversee the recruitment process (Muñoz and Otero, 2025). Although passing the EUNACOM exam is sufficient for eligibility, hospitals reportedly consider the candidate's EUNACOM score an important factor in hiring decisions (Equipo EUNAMed, 2021).

2.2 Data

Data Sources Our main analysis combines several data sources: matched employer-employee data for all public hospitals, the national registry of public health care providers, data on licensing scores, and individual-level discharge and death records. We combine these data with records on waitlists in secondary empirical exercises.

Employer-Employee Data for Public Hospitals: We use matched employer-employee administrative records managed by the Ministry of Health, which cover the universe of physicians employed in all public hospitals in Chile between 2011 and 2019 (SIRH, 2019). These data consolidate information from various public health organizations into a unified registry and provide comprehensive records for payroll processing and workforce management, including detailed wage data and employment information. We complement these data with hospital-level characteristics, including size, location, and referral area, among others (DEIS, 2020).

[12] Argentina, Brazil, Colombia, Ecuador, Spain, the United Kingdom, and Uruguay.

[13] This is similar to the pay scales for physicians in the NHS, which are determined nationally through negotiations between government bodies and medical unions. In Appendix Figure A.1 we plot the average hourly wages for physicians in public hospitals from 2011 to 2018. Despite the significant growth in the number of physicians, hourly wages remained flat after adjusting for sector-wide remuneration adjustments.

National Registry of Public Healthcare Providers: We use a registry of all physicians (and other healthcare professionals) legally authorized to practice in Chile (RNPI, 2024). The registry is managed by Chile's Superintendency of Health (*Superintendencia de Salud*), and provides detailed information on registered healthcare professionals, including their nationality and details of their professional degree including title, date of issuance, and the name and country of the granting institution. For degrees obtained abroad, the registry includes the date of revalidation in Chile.

Licensing Exam Scores: We use confidential score records for all physicians who have taken the national licensing exam (EUNACOM). These data were provided directly by ASOFAMECH (*Asociación de Facultades de Medicina de Chile*), the organization responsible for administering and overseeing the exam (ASOFAMECH, 2019). The dataset includes the date and scores of the theoretical portion for the EUNACOM for every attempt made by the universe of physicians who have taken the exam.

Individual-level Discharge and Death Records: We measure outcomes using administrative records of individual-level inpatient events in all public hospitals in Chile from 2011 to 2019 (DEIS, 2019). The data include diagnoses (ICD-10 codes), discharge or death dates, and patient characteristics such as birth date, gender, residence, and health insurance type. We link these data at the individual level with the universe of death records processed by the Vital Records Office, which includes deaths outside public hospitals but are only available until 2018.

Waiting Lists: We access individual-level administrative records of waiting lists for surgical procedures and specialist consultations for non-prioritized conditions (SIGTE, 2019),[14] obtained through a Freedom of Information Act (FOIA) request. In Chile, primary care physicians refer patients to hospitals for specialist consultations or surgical procedures, which place them on a waitlist. We observe each patient's entry and exit dates, along with the assigned hospital.

Data Aggregation and Outcomes We construct a hospital-by-year panel dataset using all of the sources of information described above. We assess hospital performance in terms of access and quality. Our primary measure of access is the yearly hospital service rate, defined as the number of admissions in a given year divided by the eligible population in the hospital's referral region in the corresponding year. We also use exits from the waitlist as additional access measures.

Our primary measure of quality of care is the hospital's yearly average death rate, defined as the ratio between deaths and admissions in a given year. We complement this in-hospital death rate

[14] Waitlists are divided into prioritized and non-prioritized conditions. The "Explicit Healthcare Guarantees" program, established by law, prioritizes specific diseases with evidence-based procedures and timelines for diagnosis and treatment. See Menares and Muñoz (2025) for details.

with the death rate derived by counting deaths within 28 days of a patient's admission—regardless of the place of death. In robustness exercises, we also replace death rates with complications rates. We calculate complications rates as the number of patients discharged and later readmitted for inpatient care (within 3 months) due to an ICD-10 code related to infections, hemorrhage, or other complications, divided by the total number of admissions. Panel A in Appendix Table A.1 summarizes key patient and hospital characteristics. Per hospital-year, there are on average 5,600 inpatient admissions and 78 physicians. The average in-hospital death rate is 3.3%.

Since EUNACOM was introduced in 2009, we impute scores for physicians who did not take the exam.[15] Our imputation procedure assigns scores to physicians from the region of origin r working at hospital h by combining the average score of all physicians at hospital h (the grand mean) with the score differential of physicians from region r working at hospital h. To estimate the latter, we run several hospital-level regressions of licensing scores on region-of-origin fixed effects, imposing sum-to-zero constraints on those fixed effects. We then adjust these estimates using empirical Bayes, such that region differentials are shrunk towards zero as the estimates' precision decreases (Efron and Morris, 1973; Walters, 2024). In Appendix A, we present additional details on our imputation procedure. Panel B in Appendix Table A.1 describes the evolution of test scores across multiple years, and we discuss them in further detail in the next section.

3 Descriptive Evidence

In line with similar trends in other OECD countries, the number of physicians per capita in Chile has increased, driven by a growing number of domestic graduates and greater reliance on foreign-trained physicians (OECD, 2019). Panel A in Figure I shows that the number of new physicians enrolled in the National Registry of Healthcare Providers rose from around 1,500 in 2011 to almost 3,500 in 2018, with most of the increase explained by an unprecedented inflow of foreign-trained physicians. According to the OECD (2019), the share of physicians in Chile who are foreign-trained was 21% in 2017, higher than the OECD average of 16%, but similar to the levels observed in Canada (24%)

[15] Before the introduction of EUNACOM, Chile had a voluntary National Medical Examination (EMN) administered in Chilean medical schools from 2003 to 2008. Prior to the EMN, licensing requirements varied: local medical graduates were required to pass the Medical Surgeon Degree Examination, while foreign-trained physicians had to complete a Foreign Medical Qualification Revalidation Examination.

and the United States (25%).[16][17] Interestingly, there is no evidence of an increase in the share of foreign-trained healthcare workers beyond physicians. See Appendix B for additional details.

The increasing number of physicians allowed to work in the country is also reflected in the number of physicians taking the licensing exam and their scores. In 2013, relatively few foreign-trained physicians took and passed the exam (i.e., had a score ≥ 51), as shown in Panel B of Figure I. By 2018, the number of foreign-trained physicians taking the licensing exam had increased nearly tenfold, and they outnumbered local graduates by 2.5 to 1, as illustrated in Panel C.

Importantly, not only did the number of foreign-trained physicians increase, but the whole distribution of scores shifted to the left, and by 2018 the distribution of foreign-trained physicians was centered around the licensing threshold. This indicates that while many successfully validated their medical degree, many were unable to do so because the licensing threshold was a binding constraint for them. Panel B of Appendix Table A.1 shows that the proportion of all test takers who achieved a passing score declined from 92% in 2009 to 58% in 2018. Meanwhile, the number of individuals scoring within the [51–56) range increased substantially, from 70 in 2009 to 856 in 2018, which reflects the growing mass of physicians who qualify with a score very close to the threshold.

Figure II shows aggregate trends in public hospitals' inputs and outcomes, and previews some of the key results of our empirical analysis. There was a large increase in the number of physicians working in public hospitals in Chile during our period of analysis. Panel A shows the number of full-time equivalent physicians working in public hospitals. By 2018, the influx of newly registered physicians is reflected in around a 50% increase in full-time physicians in public hospitals compared with 2011.

Concurrently with the increase in the quantity of physicians, the average scores of physicians working in hospitals decreased during the period. Panel B of Figure II, shows the distribution of the average hospital scores across all public hospitals in the country. We observe that the whole distribution shifted to the left. On average, the scores of physicians in public hospitals decreased by 3 points between 2013 and 2018.

[16] While immigrants represented 2% of the population in 2011, this figure rose to 8% by 2022 (INE, 2024), largely driven by the mass migration of Venezuelans to Chile and other countries in the region, which began in late 2015 following the collapse of the Venezuelan economy. Venezuelan migration to Chile and other countries in the region has been leveraged as a shock to explore the impacts of migration on various outcomes, including crime (Ajzenman et al., 2023); labor markets (Lebow, 2022; Olivieri et al., 2022; Bahar et al., 2024); and discrimination (Groeger et al., 2024), among others.

[17] Differences in remuneration and non-wage amenities are well-documented "push" and "pull" factors in physician migration (OECD, 2019). Various pieces of evidence highlight Chile as an attractive destination for medical professionals in Latin America, driven by significant wage differentials and better working conditions. Physicians in Chile can earn up to eight times more than their counterparts in Argentina, which has fueled temporary migration among Argentine physicians (Castro, 2024). Similarly, professional insecurity, low salaries, and limited social recognition have prompted physicians to emigrate from Venezuela (Hernández and Ortiz Gómez, 2011).

The aggregate trends in mortality provide suggestive evidence for a quantity-quality tradeoff at play. Panel C of Figure II shows the trend in aggregate death rates across public hospitals over time as the number of physicians increased but their scores decreased. At face value, there is no visible change in average mortality over time. While the increase in the number of physicians could have, in principle, decreased hospital mortality, this effect seems to have been counterbalanced by the decrease in the average quality of the hospitals they work in.[18] This is exactly the trade-off embedded in physician licensing that we study in detail the remainder of the paper.

4 A Simple Licensing Problem

We begin with a simple theoretical framework to study the economic fundamentals that govern the effects of licensing in our context. We focus on a particular margin: How the stringency of the requirement, as determined by the passing score of the licensing exam, changes quantity and quality of labor and ultimately patients' outcomes.

Consider a public hospital endowed with a production function F. The planner faces a distribution of physicians with latent quality θ and has access to a licensing technology based on a noisy signal of quality, s. The planner grants licenses to those with $s > \underline{s}$, where $\underline{s}$ is the policy variable determining the licensing threshold (passing score). The output of interest is an increasing function of labor $L(\underline{s})$ and a quality index $Q(\underline{s})$,

$$Y\left(\underline{s}\right) = F\left(L(\underline{s}), Q(\underline{s})\right),$$

where $F_L > 0$ and $F_Q > 0$. The outcome Y can represent population survival (the opposite of mortality), which increases with L through, for example, shorter waiting times (and less deterioration of health capital) and also increases with Q through better quality of care.

This production function generates a quality-quantity trade-off in the licensing policy, as long as there is positive mass in around $\underline{s}$ and the quality input is positively related to the cutoff: $Q(\underline{s})' > 0$. Specifically, the elasticity of the outcome with respect to the licensing threshold, $\eta^Y_{\underline{s}} \equiv \frac{\partial Y}{\partial \underline{s}} \frac{\underline{s}}{Y}$, equals

$$\eta^Y_{\underline{s}} = \underbrace{\eta^Y_L \eta^L_{\underline{s}}}_{\text{Licensing Quantity Effect}} + \underbrace{\eta^Y_Q \eta^Q_{\underline{s}}}_{\text{Licensing Quality Effect}}, \tag{1}$$

where η^Y_L and η^Y_Q are the output elasticities with respect to the quantity and quality inputs, respectively, and $\eta^L_{\underline{s}}$ and $\eta^Q_{\underline{s}}$ denote the elasticities of each input with respect to the licensing threshold.

[18] Section 5 is devoted to quantify this trade-off using quasi-experimental variation.

The impact of the policy variable $\underline{s}$ on the outcome $Y(\underline{s})$ is fully characterized by these four sufficient statistics.[19] On the one hand, increasing the threshold decreases output by reducing the quantity of labor. This *Licensing Quantity Effect* depends on the output elasticity with respect to quantity, and the elasticity of labor with respect to the threshold. On the other hand, increasing the threshold increases output by improving quality. This *Licensing Quality Effect* depends on the output elasticity with respect to quality and the elasticity of quality with respect to the threshold.

The ratios of input and output elasticities are sufficient to determine the sign of the effect of changing the licensing threshold on the outcome as

$$\eta_{\underline{s}}^{Y} > 0 \iff \eta_{L}^{Y}/\eta_{Q}^{Y} < -\eta_{\underline{s}}^{Q}/\eta_{\underline{s}}^{L}. \tag{2}$$

Equation (2) states, simply, that it is beneficial to increase the licensing threshold if and only if the ratio of output elasticities between quantity and quality is smaller than the ratio of the associated changes in those inputs. In other words, it is more beneficial to increase the threshold the higher the return to quality relative to the return to quantity, and the higher the gains in quality relative to the losses in quantity.

Parameterizing the model allows us to gain further insights into the microfoundations for the elasticities that govern the quality-quantity trade-off in licensing. Let the signal s have a distribution with density $h(s)$ and total mass m, and further assume that $s = \theta + \epsilon$, with $\epsilon \sim N(0, \sigma_\epsilon^2)$ and $\theta \sim N(\mu_\theta, \sigma_\theta^2)$; defining the signal-to-noise ratio as $\text{SNR} \equiv \frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\epsilon^2}$. Physicians who pass the exam can either work at the hospital or go to an outside option. We let $p(s|\underline{s})$ denote the share of physicians with score s who match with the hospital; this may depend on the licensing threshold $\underline{s}$ through equilibrium effects in the labor market. The total labor function is therefore $L(\underline{s}) = m\int_{\underline{s}}^{\infty} h(s)p(s|\underline{s})ds$. Let the quality index be equal to average quality, $Q(\underline{s}) = \bar{\theta}(\underline{s})$, with $\bar{\theta}(\underline{s}) = 1/L(\underline{s})\int_{\underline{s}}^{\infty} \theta(s)h(s)p(s|\underline{s})ds$. Finally, the production function has the form $F\left(L, \bar{\theta}\right) = \exp(\alpha_{\bar{\theta}}\bar{\theta}) \cdot L^{\alpha_L}$; where the term $\exp(\alpha_{\bar{\theta}}\bar{\theta})$ represents a quality-adjusted productivity, and α_L represents the returns to the scale of labor.

Under this parameterization, the elasticity of output with respect to the licensing threshold $\underline{s}$ can be written as

$$\eta_{\underline{s}}^{Y} = \alpha_L \cdot \eta_{\underline{s}}^{L} + \alpha_{\bar{\theta}} \cdot \tilde{\eta}_{\underline{s}}^{\bar{\theta}},$$

[19] Chetty (2009) provides a general framework for welfare analysis from a set of sufficient statistics derived using envelope conditions under the assumption that agents make optimal decisions. However, Chetty (2009) also notes that the assumption of optimizing behavior is not necessary as long as researchers can estimate the terms included in the derivative of the planner's objective function with respect to the policy variable. That is indeed our case. For the interested reader, in Appendix C we recast the licensing problem as a maximization problem and derive Equation (1) in that framework using envelope conditions.

where $\eta_{\underline{s}}^{L}$ is the elasticity of labor with respect to the licensing threshold and $\tilde{\eta}_{\underline{s}}^{\bar{\theta}}$ is the semi-elasticity of quality with respect to the threshold. Moreover, in the absence of equilibrium effects in the labor market—i.e., when $\partial p(s|\underline{s})/\partial \underline{s} = 0$—the elasticities can be expressed as simple functions of the underlying model parameters.[20] The elasticity of labor with respect to the threshold is

$$\eta_{\underline{s}}^{L} = \frac{-m \cdot h(\underline{s}) \cdot p(\underline{s}|\underline{s}) \cdot \underline{s}}{L},$$

which depends on the mass of workers at the threshold who would match with the hospital, as a fraction of workers above the threshold. In turn, the semi-elasticity of quality with respect to the licensing threshold can be written as

$$\tilde{\eta}_{\underline{s}}^{\bar{\theta}} = -\eta_{\underline{s}}^{L} \cdot \mathrm{SNR} \cdot \left(\mathbb{E}[s|s > \underline{s}] - \underline{s}\right),$$

which depends on the share of marginal workers,$-\eta_{\underline{s}}^{L}$; the effect of marginal workers on average *scores*, $(\mathbb{E}[s|s > \underline{s}] - \underline{s})$; and the precision of the signal, SNR.

On net, the elasticity of outcome with respect to the threshold is therefore

$$\eta_{\underline{s}}^{Y} = \eta_{\underline{s}}^{L} \cdot \left(\alpha_L - \alpha_{\bar{\theta}} \cdot \mathrm{SNR} \cdot \left(\mathbb{E}[s|s > \underline{s}] - \underline{s}\right)\right),$$

where the first term is the share of marginal workers, and the second term is the net effect of a marginal worker on production. The quantity effect of an extra worker is α_L. However, the marginal worker decreases average quality by $\mathrm{SNR} \cdot (\mathbb{E}[s|s > \underline{s}] - \underline{s}))$. This term, multiplied by the return on quality, gives the quality effect of the marginal worker.

Under this parameterization, it is beneficial to increase the licensing threshold as long as

$$\frac{\alpha_L}{\alpha_{\bar{\theta}}} < \mathrm{SNR} \cdot \left(\mathbb{E}[s|s > \underline{s}] - \underline{s}\right) \equiv R(\underline{s}), \tag{3}$$

where $R(\underline{s}) = -\tilde{\eta}_{\bar{\theta}}^{L} / \eta_{\bar{s}}^{L}$ is the marginal rate of transformation between quantity and quality, and thus represents the trade-off embedded in the licensing problem: By decreasing the quantity of labor by 1% the planner can increase its quality by $R(\underline{s})$. Under this parametrization, this increase in quality is the product of the distance between the average score and the marginal score and the precision of the signal. Increasing the licensing threshold increases output if and only if the ratio between the output elasticity with respect to quantity and the output elasticity with respect to

[20] In our empirical model, we allow for equilibrium effects to affect the matching probabilities.

quality is lower than $R(\underline{s})$.

Figure III provides a graphical depiction of the licensing problem and the net benefits of increasing the licensing threshold. Different licensing thresholds yield different combinations of available quantity and quality, and trace a decreasing relationship between the two as depicted by the blue line. The slope of this relationship at a given licensing threshold $\underline{s}$ corresponds to $R(\underline{s})^{-1}$. The production function is depicted by the solid black line, with a slope that reflects the marginal rate of technical substitution between quantity and quality, $\alpha_{\bar{\theta}}/\alpha_L$. The net effect of changing the licensing threshold at the margin depends on the relative magnitude of the slopes at the current threshold. In the example depicted in Figure III, the increase in quality achieved by a 1% decrease in quantity is higher than the increase in quality needed to leave output unchanged. In this case, increasing the licensing threshold would improve outcomes. A key goal of the empirical analysis is thus to quantify these two slopes, which are the sufficient statistics needed to identify the effects of changing the threshold on outcomes.[21]

Outcomes of Interest For the empirical counterpart of the model, we focus on mortality and access as the primary outcomes of interest.

Multiple channels may underlie the relationship between physicians' quantity and quality and patient mortality. On the one hand, having a larger pool of doctors may generate economies of scale that allow for more intensive monitoring and faster intervention when complications arise, thereby improving outcomes and lowering death rates (Hoot and Aronsky, 2008; Hoe, 2022). A greater presence of physicians might also reduce treatment delays for inpatients—a factor that is especially crucial in life-threatening circumstances.[22] On the other hand, licensing policies are justified on the grounds that the performance of physicians on licensing exams is a good predictor of patients' outcomes. Clinically stronger doctors may diagnose and manage complex conditions more accurately, make timely decisions, and consistently follow evidence-based practices. Such expertise may also lower the risk of complications, curtail treatment errors and misdiagnoses, and

[21] The optimal licensing threshold $\underline{s}^*$ can be implicitly characterized with a simple expression:

$$\underline{s}^* = \mathbb{E}[s|s > \underline{s}^*] - \frac{\alpha_L}{\alpha_{\bar{\theta}}} \cdot \frac{1}{\text{SNR}}.$$

The optimal threshold increases with the mean score. A higher mean allows the planner to obtain more quantity for any given level of quality. Also, the optimal threshold decreases with the output elasticity with respect to labor and increases with the output semielasticity with respect to quality. It is also lower when the licensing technology is imprecise (i.e., low signal-to-noise ratio), as in such cases a higher threshold can only weakly differentiate between high- and low-quality physicians.

[22] Earlier studies show that quicker access to care is a principal route through which physician quantity influence mortality (see, e.g., Prentice and Pizer 2007; Pizer and Prentice 2011; Nikolova et al. 2016); Gruber et al. (2023) documents a similar pattern in emergency settings.

ultimately improve patient survival. We therefore hypothesize that licensing requirements create a quantity-quality trade-off with direct consequences for mortality.[23] To measure this trade-off empirically, we model patient mortality as a function of both the number of practicing physicians and their quality as inferred from the licensing exam.

Our second outcome of interest is access, as measured by the service rate. Beyond its impact on mortality, timely access to care is an important outcome in itself, because individuals may incur costs of waiting (as in, e.g. Russo, 2023) or deteriorate their health capital even without effects on mortality. While access to care in capacity-constrained systems is most likely dependent on the quantity, licensing exams could, in principle, capture physicians' ability to reduce length of stay and ultimately improve access as well. Therefore, our empirical model also tests whether the service rate depends on physicians' quality.

Many hospitals and dynamic considerations In the empirical part of the paper, we consider a set of $|\mathcal{J}|$ hospitals and thus estimate hospital-specific input elasticities that we then aggregate. We also quantify both the static (short-run) and dynamic (medium-run) implications of changing the threshold on hospital outcomes. Consider a planner setting the licensing threshold in period 0. We can restate Equation (1) and express the elasticity of the aggregate health outcome across hospitals with respect to the licensing threshold in a period $R \geq 0$ as

$$\eta_{\underline{s}}^{Y,R} = \underbrace{\alpha_L \cdot \bar{\eta}_{\underline{s}}^{L,R}}_{\text{Licensing Quantity Effect}} + \underbrace{\alpha_\theta \cdot \bar{\bar{\eta}}_{\underline{s}}^{\bar{\theta},R}}_{\text{Licensing Quality Effect}},$$

where $\bar{\eta}_{\underline{s}}^{L,R} = \sum_{j \in \mathcal{J}} p_j \cdot \eta_{\underline{s}}^{L_j,R}$ is the average elasticity of labor in period in period R with respect to $\underline{s}$, $\bar{\bar{\eta}}_{\underline{s}}^{\bar{\theta},R} = \sum_{j \in \mathcal{J}} p_j \cdot \eta_{\underline{s}}^{\bar{\theta}_j,R}$ is the average semi-elasticity of quality in period R with respect to $\underline{s}$, and p_j are hospital-specific Pareto weights.[24]

[23] For instance, stricter licensing standards can raise treatment quality but also lengthen waiting times, potentially worsening health outcomes. Appendix D formalizes this intuition.

[24] These average elasticities are consistent with a planner considering aggregate health outcomes of the form:

$$Y_R \equiv \prod_{j \in \mathcal{J}} y_{jR}^{p_j}.$$

The main appeal of this aggregation is that, when $p_j > 0\ \forall j$, it penalizes solutions where the outcome is very low for some hospitals.

5 Output Elasticities

5.1 Health Outcomes Model

In period t, each hospital j produces an outcome y^k_{jt} with the production function

$$y^k_{jt} = \exp\left(\boldsymbol{O'}_{jt}\boldsymbol{\beta^k_O}\right)\exp\left(\alpha^k_\theta Q\left(H_{\theta_{jt}}\right)\right) L_{jt}^{\alpha^k_L}\exp(\nu_{jt}), \tag{4}$$

where L_{jt} is the quantity of physicians, $\boldsymbol{O}_{jt}$ includes other determinants of hospital output such as other labor, capital, patient case-mix, and other unobserved drivers of productivity. There is also a productivity term, $\exp(\alpha^k_\theta Q(H_{\theta_{jt}}))$, which depends on an index of physician quality, $Q(H_{\theta_{jt}})$. This quality index captures how hospital outcomes are affected by the distribution of the quality of physicians at hospital j in period t, $H_{\theta_{jt}}$. Finally, ν_{jt} is a random shock. The parameters of interest for the quantity-quality trade-off are α^k_L and α^k_θ, which correspond to the output-specific elasticities with respect to the quantity and quality of physicians, respectively.

Empirical implementation of this model requires a stance on the functional form for the quality index. In our main specification, we assume that $Q(H_{\theta_{jt}}) = \bar{\theta}_{jt}$; i.e., the quality index is the mean of the quality distribution in hospital j in period t.[25] We also specify $\boldsymbol{O'}_{jt}\boldsymbol{\beta^k_O}$ to be a linear combination of hospital fixed effects ρ^k_j, year fixed effects that vary with hospital complexity, $\gamma^k_{f(j)t}$, the log number of beds, B_{jt}, patients' demographics D_{jt},[26] and an unobservable ω_{jt}:

$$\boldsymbol{O'_{jt}\beta^k_O} = \rho^k_j + \beta^k_D D_{jt} + \beta^k_B B_{jt} + \gamma^k_{f(j)t} + \omega^k_{jt}$$

Using the fact that $\theta_i = s_i + \epsilon_i$, and denoting $\bar{\epsilon}_{jt} \equiv \sum_{i\in I_{jt}} \epsilon_{it}/L_{jt}$ and $\bar{s}_{jt} \equiv \sum_{i\in I_{jt}} s_{it}/L_{jt}$ as the average measurement error and score in hospital j in period t respectively, the empirical analog of Equation (4) becomes

$$\ln(y^k_{jt}) = \alpha^k_L \ln(L_{jt}) + \alpha^k_\theta \bar{s}_{jt} + \gamma^k_{f(j)t} + \rho^k_j + \beta^k_D D_{jt} + \beta^k_B B_{jt} + \omega^k_{jt} + \alpha^k_\theta \bar{\epsilon}_{jt} + \nu^k_{jt}. \tag{5}$$

Equation (5) is our estimating equation to recover α^k_L and α^k_θ—the output-specific elasticities with respect to the quantity and quality of physicians, respectively.

[25] In Appendix Table A.4, we present the estimates from an alternative functional form where we define the quality index as the share of physicians in hospital j during period t with quality below the median of the overall distribution. The results remain qualitatively similar: quality does not affect access, but lower quality is associated with higher mortality.

[26] In our empirical implementation, the vector D_{jt} includes the shares of female and foreign patients, eight age bands (0–29, then 10-year increments to 90+), and insurance types grouped by copayment levels.

5.2 Estimation and Results

The error term in Equation (5) consists of three components: an unobserved productivity shock potentially known before input choices ω^k_{jt}; measurement error in quality, $\alpha^k_\theta \bar{\epsilon}_{jt}$; and an unobserved productivity shock that occurs after input decisions, ν^k_{jt}. To address the identification challenges posed by unobserved productivity shocks and the measurement error in latent quality, we implement an instrumental-variable approach (Agostinelli and Wiswall, 2025). Specifically, we use two shift-share (Bartik) instruments (Altonji and Card, 1989; Autor et al., 2013) to recover the causal effects of physician quantity and quality on hospital-level outcomes.

Our instruments leverage exogenous variation in both the number and average performance of licensing-exam takers, driven mainly by the unprecedented migration of physicians into the country and, to a lesser extent, by the domestic expansion of medical schools. The instrument for physician quantity, Z^L_{jt}, is constructed in the following way: For the shift component, we calculate the percentage change in the number of test-takers who clear the cutoff of the licensing exam from each region of training c, M^c. Formally, and following the notation in Acemoglu and Restrepo (2020), the shift component is defined as $\%\Delta\text{Test-takers}^c_{(t_0,t_1)} \equiv (M^c_{t_1} - M^c_{t_0})/M^c_{t_0}$, where $(t_0, t_1) = (2011, 2015)$ if $t < 2015$ and $(t_0, t_1) = (2015, 2019)$ if $t \geq 2015$. For the share component, we use the percentage of physicians from region c who worked at hospital j the year before. Formally, the share component is defined as $\text{Share Physicians}_{cjt-1} \equiv \frac{L_{cjt-1}}{\sum_c L_{jct-1}}$.[27] The instrument for physician quantity is thus:

$$Z^L_{jt} = \sum_c \%\Delta\text{Test-takers}^c_{(t_0,t_1)} \times \text{Share Physicians}_{cjt-1}.$$

The instrument for physicians' quality uses the same share component but replaces the shift component with the change in the average scores of test takers who clear the cutoff from each region of training c, S^c. Formally, the quality shift component is defined as $\Delta\text{Avg. Quality}^c_{(t_0,t_1)} \equiv \bar{S}^c_{t_1} - \bar{S}^c_{t_0}$, where $(t_0, t_1) = (2011, 2015)$ if $t < 2015$ and $(t_0, t_1) = (2015, 2019)$ if $t \geq 2015$. The instrument for physician quality is thus:

$$Z^\theta_{jt} = \sum_c \Delta\text{Avg. Quality}^c_{(t_0,t_1)} \times \text{Share Physicians}_{cjt-1}.$$

In our setting, identification of the returns to quantity and quality using shift-share instruments is

[27] The share component relies on the idea that past immigrant settlement/job patterns predict where new immigrants will locate/work (Altonji and Card, 1989). We use the lagged share to maximize the predictive power of the instrument (Borusyak et al., 2025). Given that before the migration wave, very few foreign doctors were licensed in Chile and they were concentrated in a few nationalities, using the pre-migration share would result in shares close to zero and consequently, very weak instruments.

predicated on the assumption that the shift component is as good as random and not correlated with factors that would affect the outcomes of interest. In other words, our instruments leverage a shift-level natural experiment mainly created by a migration shock.[28] To support this assumption, we show in Appendix E that the quantity and quality shocks do not predict predetermined variables related to hospitals' workforce and patients' demographics.

Two-Stage Least Squares Results Table I presents the results from the 2SLS estimation. We present point estimates and exposure-robust p-values obtained after clustering standard errors at the shock (region-of-origin) level, following Adao et al. (2019) and Borusyak et al. (2022).[29] We also complement our analysis by reporting the Anderson and Rubin (1949) p-values (Lee et al., 2022).

Panel A focuses on access. Column (1) reports results for the hospital service rate, which is our main measure of access. We find an elasticity of service rate with respect to the number of physicians of 0.99, suggesting that the number of patients admitted increase proportionally with the number of physicians. We also see no impact of physicians' quality on the service rate. Columns (2)-(4) analyze alternative measures of access. Column (2) restricts the sample to hospitals that perform surgeries and analyzes the volume of inpatient surgical procedures. For this outcome, the elasticity with respect to the number of physicians is 4.7, and the effect of physician quality remains statistically insignificant (p-value > 0.2). Column (3) shows that increasing the number of physicians significantly raises the number of exits from the surgical waiting list, while changes in physician quality have no discernible impact. Similarly, Column (4) finds that a higher number of physicians increases exits from the consultation waiting list, with physician quality again showing no effect.

Overall, our findings suggest that the quantity of physicians significantly influences access, consistent with the notion that physician scarcity is the primary binding constraint in expanding hospital capacity and addressing patient demand. Our results also show that the quality of physicians does not impact a hospital's ability to admit more patients, suggesting that the availability of physicians, rather than their ability, is the primary determinant of hospital utilization rates.

Panel B of Table I focuses on treatment quality, where our main measure is the in-hospital death rate. Column (1) shows the main results. We find that (i) increasing the number of physicians

[28] In general, identification using shift-share instruments relies on the assumption that either the "shifts" or the "shares" components are as good as random and not correlated with factors that would affect the outcomes of interest (Goldsmith-Pinkham et al., 2020; Borusyak et al., 2022, 2025).

[29] To obtain these p-values—which are shown to be asymptotically valid (Adao et al., 2019)—we estimate the transformed regression proposed by Borusyak et al. (2022).

by 1% decreases the in-hospital death rate by 0.55% and (ii) a one standard deviation increase in average physicians' quality decreases death rates by 21%. One potential explanation for these results is that the marginal patients—those who were previously not admitted—may be healthier, so a shift in patient composition could be driving the results (Kelly and Stoye, 2020). To explore this hypothesis, Column (2) examines the impact on the predicted hospital death rate, which is estimated based on patients' characteristics.[30] Neither physician quantity nor quality significantly affects predicted mortality rates, which suggests that changes in patient composition do not explain our findings.

Columns (3) and (4) present results for additional quality measures. In Column (3), we use the 28-day mortality rate as the outcome and find effects consistent with those reported in Column (1), showing that the impacts of physicians' quantity and quality remain robust even when accounting for out-of-hospital deaths among discharged patients. Finally, Column (4) focuses on the readmission rate due to complications. We define this rate as the number of patients discharged from the hospital who are later (within 3 months) readmitted for inpatient care, at any hospital, due to an ICD-10 code related to infections, hemorrhage, or other complications, divided by the total number of admissions. Consistent with the impacts on mortality, we find that physicians' quantity and quality help to reduce the complication rate.

Heterogeneity The mechanism through which quantity and quality affect mortality discussed in Section 4 may be especially relevant for the most severe conditions, in which delays or errors are likely to be more consequential. We explore this heterogeneity in Appendix G, where we examine whether the effects of physician quantity and quality vary with patient severity. We proxy severity using the predicted mortality risk associated with each patient's condition. The results reveal substantial heterogeneity: both physician quantity and quality significantly reduce mortality among patients with high-risk diagnoses (top 20% of predicted mortality), but have limited effects for lower-risk cases. Nonetheless, both dimensions of physician input affect hospitals' complication rates even among low-risk patients, indicating that staffing decisions broadly influence the quality of care, with especially pronounced effects on mortality among high-risk cases.

Other Inputs A potential concern with our empirical strategy is that our instruments may also capture changes in other healthcare inputs—for instance, if the migration wave increased the

[30] We calculate the expected death rate by fitting a logit model for death outcomes at the inpatient level, controlling for patient demographics and diagnoses group, as per the enhanced Elixhauser comorbidity index (Elixhauser et al., 1998; Quan et al., 2005). The total number of *predicted deaths* at each hospital and year is then divided by the number of admissions.

number of other foreign healthcare workers beyond physicians. However, as discussed earlier and shown in Appendix B, the migration shock did not alter the share of other foreign healthcare workers, based on data from the Registry of Healthcare Providers and public hospitals' employer-employee records. In addition, Appendix F shows that the ratio between physician and other health care workers remained remarkably stable over time. An additional concern is that the migration wave may have altered the share of specialist physicians in public hospitals. However, we also do not find any evidence of this. Figure A.6 shows that the share of specialist physicians remained stable throughout our sample period.

Reduced-form Analysis We conclude this section illustrating how the quasi-experimental variation in our setting allows us to identify the production function. Consider a reduced-form version of Equation (5) that only considers the quantity instrument Z^L_{jt}:

$$\ln(y^k_{jt}) = \beta^k_Z Z^L_{jt} + \gamma^k_{f(j)t} + \rho^k_j + \beta^k_D D_{jt} + \beta^k_B B_{jt} + \omega^k_{jt} + \alpha^k_\theta \bar{\epsilon}_{jt} + \nu^k_{jt},$$

where the coefficient of interest, β^k_Z, quantifies the net effects of the exogenous change in labor supply—largely driven by the immigration of physicians to Chile—on $\ln(y^k_{jt})$.

Columns (1) and (2) of Table II show the results for physicians' quantity and quality, respectively. On average, an increase of one standard deviation of the instrument leads to a 2.5% increase in the number of physicians and a 0.057 standard deviation decrease in average quality. Column (3) shows the results when using logged mortality as an outcome. On average, the changes in labor supply captured in Z^L_{jt}—largely driven by the immigration of physicians to Chile—did not affect mortality.[31]

In sum, Table II reveals that changing the (logged) quantity and quality at a ratio of 2.3 (0.057/0.025) leaves mortality unchanged. This ratio is of particular interest as it corresponds to the slope of the "iso-mortality" curve, which, together with the ratio of input elasticities $R(\underline{\mathrm{s}})$, are the sufficient statistics that sign the net effect of increasing the licensing threshold (see Equation 3 and Figure III).[32]

In our counterfactual analysis, however, we are interested in quantifying the magnitude of the effects of changing quantity and quality by a different magnitude than that induced by the exogenous shock

[31] Column (4) of Table II shows that an increase in the quantity instrument increases the service rate, which is consistent with the findings of our 2SLS approach.

[32] More precisely, this is the case as long as both output elasticities are different from zero. In other words, based on the observation that mortality did not change after the labor supply shock, we cannot rule out the possibility that neither quantity nor quality affects mortality. We are able to rule out this possibility by adding a second instrument.

to labor supply. Therefore, it is crucial to isolate the effects of quantity and quality separately by adding the second instrument, $Z_{jt}^{\bar{\theta}}$, which, together with Z_{jt}^{L}, provides independent identifying variation in quality and quantity.

To illustrate the role that these two instruments play for identification, consider the predictions of the first-stage regressions embedded in the model of Equation (5). Specifically, for each input $I \in \{lnL, \bar{\theta}\}$ we use the first-stage coefficients $\hat{\gamma}_{\theta}^{I}$ and $\hat{\gamma}_{L}^{I}$ to predict $\hat{I}_{jt} = \hat{\gamma}_{\theta}^{I} Z_{jt}^{\bar{\theta}} + \hat{\gamma}_{L}^{I} Z_{jt}^{L}$. This prediction isolates the exogenous variation in inputs leveraged by the instruments. Similarly, to evaluate the effects on health outcomes, we estimate the reduced-form versions of Equation (5), that is, we use the instruments instead of the inputs as independent variables. Again, we can construct a prediction of health outcome $ln(y^k) \in \{ln(\text{mortality})), ln(\text{service rate})\}$ based solely on the reduced-form coefficients on the instruments, i.e., $\widehat{ln(y_{jt}^{k})} = \hat{\gamma}_{\theta}^{k} Z_{jt}^{\bar{\theta}} + \hat{\gamma}_{L}^{k} Z_{jt}^{L}$.

Panels A and B of Figure IV present scatter plots of the average (over time) predicted quantity and average quality of physicians for each hospital, with marker sizes proportional to hospital size at the beginning of our sample. In Panel A, we color each marker based on whether the hospital mortality rate—as predicted by our instruments—falls into the top, middle, or bottom tercile. Similarly, in Panel B, we color markers based on the terciles of the service rate.

The instruments generate substantial variation across hospitals in the predicted quantity and quality of physicians. The gray dots in Panel A show the set of hospitals $j \in \mathcal{J}'$ with a predicted mortality in the middle tercile. For them, we also overlay the linear fit between quantity and quality. For these hospitals, the differences with respect to the overall prediction are approximately zero—i.e., $\Delta \log \hat{y}_j \equiv \log \hat{y}_j - |\mathcal{J}'|^{-1} \sum_{j \in \mathcal{J}'} \log \hat{y}_j \approx 0$. Since the total difference in the outcome $\Delta \log \hat{y}_j$ can be expressed as the contribution of the differences in quantity and quality (i.e., $\Delta \log \hat{y}_j = \alpha_L \Delta \log \hat{L}_j + \alpha_\theta \Delta \hat{\theta}_j$), we can leverage this set of hospitals to identify the slope of the "iso-mortality curve": $-\Delta \hat{\theta}_j / \Delta \log \hat{L}_j \approx \alpha_{\bar{\theta}}^{\text{mortality}} / \alpha_{L}^{\text{mortality}}$. Consistent with the previous finding, we estimate a coefficient of approximately -2.6 when regressing $\Delta \hat{\theta}_j$ on $\Delta \log \hat{L}_j$ among this set of hospitals.

Figure IV illustrates the relationship between predicted inputs and predicted outcomes. The instruments predict lower mortality for hospitals located above the iso-mortality curve—that is, those for which the predicted quantity and quality of physicians are positive. In contrast, hospitals below the iso-mortality curve are predicted to experience higher mortality. Similarly, the instruments predict a higher service rate for hospitals with a greater predicted number of physicians, regardless of changes in average quality. By linking the magnitude of predicted input changes to corresponding changes in outcomes, our 2SLS estimation recovers α_L and α_θ separately.

6 Input Elasticities

6.1 Scores and Labor Market Matching Model

As shown in Section 4, the elasticity of inputs depends on the mass of test-takers at the threshold, the mapping between scores and quality, and the matching of physicians and hospitals. We recover these objects with a dynamic model of scores and a labor-matching model.

Dynamic Model of Scores Modeling the dynamics of scores serves two purposes. First, it allows us to infer physicians' latent quality from their histories of scores. Second, it allows us to predict score histories under counterfactual thresholds, which is a key input for predicting the dynamic effects on the mass of test-takers when evaluating the medium-run effects of changing the threshold.

Each physician i is characterized by her latent quality θ_i. The vector of scores is determined by their realizations in each attempt, as well as retaking decisions after failing. On average, scores improve over attempts, and thus we model the score of the n^{th} attempt as a noisy measure of latent quality and test-taking ability, Γ_{ni}, which can change over attempts, as in Gilraine and Penney (2023).[33] We let the mean and variance of quality, as well as the variance of the noise, to depend on the physician's region of training (or "type") $\tau_i \in \{\text{National, Foreign}\}$, to allow for the possibility that the precision of the test differs across these groups.[34] Finally, as average score gains over attempts are decreasing and convex (see Figure A.4), we assume that test-taking ability—which we normalize to zero in the first attempt—improves with exponential decay.

Formally, the model of scores is given by:

$$
\begin{aligned}
s_{in} = & \quad \theta_i + \Gamma_{in} + \epsilon_{in} \\
\theta_i \sim & \quad N(\mu_{\tau_i}, \sigma^2_{\theta,\tau_i}) \\
\epsilon_{in} \sim & \quad N(0, \sigma^2_{\epsilon,\tau_i}) \\
\Gamma_{in} = & \quad \textstyle\sum_{k=0}^{n-1} \gamma \cdot \exp(-\rho \cdot k),
\end{aligned}
$$

[33] We assume that quality is constant over time, which implies that physicians cannot experience quality gains over attempts. To support this assumption, in Appendix H we use regression discontinuity designs to show that physicians who retake the exam around the passing threshold do not experience gains on quality proxies.

[34] There are several reasons to believe that the test's precision may depend on the physician's region of training. Nationals have been trained in a system that often prepares them to take the test, which could reduce the precision of the signal. On the other hand, foreign-trained physicians normally take the test several years after graduating. Also, although most foreigners speak Spanish as their mother tongue, not all of them do, and even for Spanish-speaking foreigners there could be difficulties due to variations of the language in the region.

where μ_{τ_i} and σ^2_{θ,τ_i} are the type-specific mean and variance of quality respectively, $\sigma^2_{\epsilon,\tau_i}$ is the type-specific variance of the noise, γ captures the average improvement in test-taking ability between the first and second attempt, and ρ governs the average rate at which improvements in test-taking ability decrease over subsequent attempts.[35] Modeling the score process allows us to account for selection in estimating the model parameters.[36]

Physicians who fail the exam on attempt n retake it with a probability that is a function of the distance between their past score and the passing threshold $s_{in} - \underline{s}$; the number of attempts, n; their type, τ_i; and the licensing threshold, $\underline{s}$:

$$P(\text{retake}|s_{in}, n_i, \tau_i) = \frac{e^{\beta_{0,\tau_i} + \beta_{n,\tau_i} n_i + \beta_{s,\tau_i}(s_{in} - \underline{s})}}{1 + e^{\beta_{0,\tau_i} + \beta_{n,\tau_i} n_i + \beta_{s,\tau_i}(s_{in} - \underline{s})}}. \tag{6}$$

By allowing the retaking probability to depend on both the distance between an individual's score and the licensing exam threshold and the number of attempts, this specification captures the net benefits of retaking. These net benefits depend on two factors: (i) the probability of passing the threshold on subsequent attempts (determined by the gap between expected future scores and the threshold), and (ii) the expected value of the match conditional on passing the exam (which depends on physician type and posterior quality). In addition, explicitly incorporating the licensing threshold in the specification allows retaking behavior to change under counterfactual scenarios.

The posterior of quality for each physician given a sequence of scores over attempts $\vec{s}_i \equiv \{s_{i1}, s_{i2}, ..., s_{in}\}$ and type τ_i is equal to

$$\hat{\theta}(\vec{s}_i|\tau_i) \equiv \mathbb{E}[\theta_i \mid s_{i0}, s_{i1}, \ldots, s_{in}, \tau_i] = \mu_{\theta,\tau_i} + \frac{\sigma^2_{\theta,\tau_i}}{\sigma^2_{\varepsilon,\tau_i} + (n+1)\sigma^2_{\theta,\tau_i}} \left(\sum_{t=0}^{n} (s_{it} - \Gamma_{it} - \mu_{\theta,\tau_i}) \right). \tag{7}$$

Equation (7) is the basis for inferring a physician's quality from their vector of scores and the estimated parameters of the dynamic model of scores.[37]

Labor Market Matching Our goal is to understand how changes in the licensing threshold affect the allocation of physicians across hospitals. A central challenge in this task is that equilibrium

[35] In principle, γ and ρ could be type-specific. However, since most nationals pass the test on the first attempt, we lack the power to identify parameters for nationals and foreigners separately. This limitation is largely inconsequential for our counterfactual analysis, as with a small increase in the threshold still few nationals would fail.

[36] As discussed by Gilraine and Penney (2023), selection occurs because ϵ_{in} is not mean zero among individuals who failed the exam after attempt n, since they are more likely to have received a negative shock to their score. As a consequence, score gains between two consecutive attempts are higher than the causal effect of the retaking.

[37] Although changes in the licensing-exam threshold alter retaking probabilities, conditional on the observed score sequence $\{s_{it}\}_{t=0}^{n}$, neither the threshold nor the retaking process provides any additional information about the posterior inference of θ_i.

matching probabilities may change under a counterfactual threshold. For example, raising the threshold reduces the mass of lower-quality physicians seeking jobs, potentially altering the degree of competition in the labor market.

We adopt a reduced-form modeling approach that flexibly captures the determinants of physician-hospital matches, and restrict our counterfactuals to *local* changes in order to reduce the sensitivity of the results to our statistical approximations (Chetty, 2009). Consider a physician i with posterior quality $\hat{\theta}_i$ belonging to a quality range $r(\hat{\theta}_i)$. We define the matching probability with location j in period t by:

$$CMP_{ijt} \equiv CMP\Big([x_{i1t}, \ldots, x_{i|\mathcal{J}|t}],\ \vec{s}_i,\ \tau_i,\ \mathbf{M}_{it}(\underline{s}),\ \kappa_t,\ j\Big) = \frac{e^{v\left(x_{ijt}, \hat{\theta}(\vec{s}_i|\tau_i)\right) + g(\mathbf{M}_{it}(\underline{s}), \kappa_t|j)}}{1 + \sum_{j'} e^{v\left(x_{ij't}, \hat{\theta}(\vec{s}_i|\tau_i)\right) + g(\mathbf{M}_{it}(\underline{s}), \kappa_t|j')}}. \tag{8}$$

The function $v(\cdot)$ captures the determinants of matching probabilities based on physicians' and hospitals' preferences, which depends on the posterior quality and other match-specific factors such as distance and amenities included in $x_{i1t}, \ldots x_{i|\mathcal{J}t}$. The function $g(\cdot)$ captures market-level forces—specifically, the extent of competition physician i faces when seeking a position at hospital j, which is determined by the mass of similarly or more qualified job seekers, $\mathbf{M}_{it}(\underline{s})$, and the vector of vacancies k_t. Through this term, the model allows matching probabilities to respond endogenously to changes in the licensing threshold.

Although the matching probability is specified in reduced form, it is grounded in a microfounded model developed in Appendix I. In the model, hospitals have fixed capacities and vertical preferences over physicians' expected qualities, and the equilibrium assignment satisfies pairwise stability.[38] Under those assumptions, the labor market clears with a set of hospital-specific quality cutoffs that determine the minimum quality of physicians that can match with each hospital (Azevedo and Leshno, 2016).[39] A key implication is that the choice set of a physician with posterior quality θ_i depends on the mass of physicians with similar or higher quality. That is, the equilibrium allocation is such that only competition from equally or better-ranked peers can affect a physician's match.

[38] Agarwal (2015) proposes and estimates an empirical model for the US centralized medical residency program under a similar set of assumptions.

[39] This setup is motivated by two facts of our setting. First, wages are rigid: they are set by the public-sector wage schedule and have remained unchanged despite large supply shocks (see Figure I). Second, observed sorting patterns are consistent with vertical preferences: over 70% of the variance in physician quality is explained by differences between hospitals rather than within.

6.2 Estimation and Results

Latent Quality Table III, Panel A, presents maximum likelihood estimation results for the retaking model specified in Equation (6). As expected, the probability of retaking decreases with the number of attempts and the distance of the score from the cutoff.

We estimate the model of score gains via simulated method of moments (SMM), using the estimated retaking probabilities to account for selection behavior in the retaking process. We match the following empirical moments separately for nationally trained and foreign-trained physicians: the means and variances over the first two attempts and the covariance between attempts for retakers. We also match the mean of gains over attempts for the first four attempts in the pooled sample. Panel B of Table III shows the estimated coefficients. Foreign-trained physicians have a lower prior quality than nationals (46.0 versus 65.4), while their quality distribution and noise distribution have a higher variance.[40] On net, the signal-to-noise ratio is higher for foreign-trained than for nationally-trained. The estimated first gain in test-taking ability ($\hat{\gamma}$) is 9.6, which is almost 5 points below the average score increase between attempts in the raw data. This reflects the fact that the difference in scores is an overestimate of the gains (Gilraine and Penney, 2023). We also find that score gains decay at an exponential rate of $\hat{\rho} = 0.285$—i.e., gains decrease by approximately 25% between two consecutive attempts.

With the estimated parameters and given the history of score realizations for each physician, we use Equation (7) to estimate each individual's posterior quality.

Matching Probabilities We estimate the conditional matching probabilities by maximum likelihood following Equation (8). Locations denoted by j include all 29 Health Services and the outside option. We specify $v(x_{ijt})$ to include a rich set of alternative-specific time fixed effects, and alternative-specific shifters based on the physician's posterior quality, $\hat{\theta}_i$, an indicator for whether the physician is national, National_i, and an indicator for whether the physician is a specialist, Specialist_i. We also include the distance between the university of training and the centroid of location j for nationals, $\text{Distance}_{ij} \times \text{National}_i$. Finally, and consistent with our shift-share IV strategy, we include the percentage of physicians from i' region of origin $c(i)$ who worked at loca-

[40]The higher variance of quality is expected, since foreign-trained physicians represent a more heterogeneous group in terms of experience, region of training, etc. The higher noise among foreign-trained physicians is likely explained by the fact that they have had fewer opportunities to practice for the test than nationals.

tion j in $t-1$; Share Physicians$_{c(i)jt-1}$. Thus, v_{ijt} is defined as follows:

$$
\begin{aligned}
v_{ijt} = \alpha^d &\text{Distance}_{ij} \times \text{National}_i + \alpha^h \text{Share Physicians}_{c(i)jt-1} \\
&+ \alpha_{jt} + \alpha_j^n \text{National}_i + \alpha_j^q \hat{\theta}(\vec{s}_i|\tau_i) + \alpha_j^{nq} \hat{\theta}(\vec{s}_i|\tau_i) \times \text{National}_i + \alpha_j^s \text{Specialist}_i .
\end{aligned} \tag{9}
$$

The coefficients α_{jt} represent mean alternative-specific year fixed effects, and α_j^n, α_j^q, α_j^{nq}, and α_j^s allow those mean effects to vary with physicians' characteristics. We normalize $v_{ijt} = 0$ for the outside option.

As specified in Equation (8), the matching probabilities depend on the mass of physicians who approve the licensing exam across different quality ranges and the vacancies in each location. For our empirical application, we capture those general equilibrium effects as a simple linear function of the ratio between job seekers and vacancies (i.e., the inverse of "labor market tightness"):

$$
g(M_{it}(\underline{s}), \kappa_t | j) = \boldsymbol{\beta}_j' \frac{\mathbf{M}_{it}(\underline{s})}{\kappa_{jt}}, \tag{10}
$$

where $\boldsymbol{\beta}_j$ is a 2×1 vector of location-specific coefficients and $\mathbf{M}_{it}(\underline{s}) = [M_{it}^0(\underline{s}), M_{it}^+]'$ is a 2×1 vector composed of the mass of physicians in i's quality range, $M_{it}^0(\underline{s})$, and the mass of physicians above i's quality range, M_{it}^+.[41] We use four (posterior) quality quantiles to construct $\mathbf{M}_{it}$. We proxy the number of vacancies κ_{jt} with the ratio of beds and the stock of physicians in the previous period; that is, $\tilde{\kappa}_{jt} = \frac{\text{Beds}_{jt}}{\text{Stock of Physicians}_{j,t-1}}$.

Results are presented in Table IV. Column (1) shows the estimates from a specification in which $g(\cdot)$ depends only on other physicians within the same quality range (M_{it}^0). Column (2) allows $g(\cdot)$ to also depend on the mass of physicians with a higher quality range (M_{it}^+). Column (3) allows the effects of these masses to vary across three different hospital-quality tiers, which we construct based on the average quality of their physicians. We find not significant heterogeneity along that dimension. As a placebo check, in Column (4) we show that the mass of physicians in a lower quality range (which we denote M_{it}^-) does not affect the matching probabilities. Across specifications, the matching probabilities depend negatively with distance. Also, and in line with our shift-share IV

[41] Let m_t be the number of test takers in period t and $h_t(s, X)$ the joint distribution of scores and characteristics in period t. With a slight abuse of notation, we define the mass of physicians who pass the exam in year t and have expected quality within a range $r \in \mathcal{R}$:

$$
M_{r,t}(\underline{s}) \equiv m_t \int_X \int_{s > \underline{s}:\ \hat{\theta}(s) \in r} h_t(s, X) ds dX. \tag{11}
$$

The mass of physicians in i's quality range is $M_{it}^0(\underline{s}) = M_{r(\hat{\theta}_i),t}(\underline{s})$ and the mass of physicians above i's quality range, $M_{it}^+ = \sum_{r' \succ r(\hat{\theta}_i)} M_{r',t}$. Notice that only M_{it}^0 depends on the licensing score, as the licensing score only affects the mass at the bottom of the distribution.

strategy, the matching probabilities depend positively on Share Physicians$_{c(i)jt-1}$. As predicted by the theory, we find that matching probabilities depend negatively on the (inverse of) labor market tightness. The effects are mostly concentrated within the same quality range (M^0_{it}), with little effects from quality ranges above (M^+_{it}).

The matching probabilities allow us to compute the elasticity of the quantity and quality of physicians in each hospital with respect to the licensing threshold. We compute these elasticities numerically, by simulating the hospital-physician matches under the baseline licensing threshold $\underline{s}$ and under a counterfactual threshold $\underline{s}' = \underline{s} + \Delta\underline{s}$, and compute the resulting qualities and quantities and their elasticities.[42]

These matching probabilities imply that the average quantity elasticity, pooling across years and hospitals, is -0.061, and the average quality semi-elasticity is 0.194 standard deviations. The ratio of these equals -3.2 which implies that, on average, a 10% decrease in labor due to a higher threshold would increase quality by 0.32 standard deviations.

Correlates of the Input Elasticities The flexible specification for the matching probabilities allows for rich heterogeneity in the elasticities across hospitals. Appendix Table A.5 presents the results from estimating a series of OLS regressions in which we project the labor elasticity (Columns 1-4) and the quality semi-elasticity (Columns 5-9) on hospital observables. Consistent with our assumptions on vertical preferences, our estimates predict that increasing the licensing threshold has a larger effect on hospitals with lower-quality doctors (below-median average scores). We also find that increasing the licensing threshold would have higher impacts on hospitals with more scarcity (below-median ratio of physicians per patient) and in the north (above-median latitude).

7 The Impact of Increasing the Licensing Threshold

Taken together, the input and output elasticities estimated in the previous sections enable us to measure the elasticities of hospital-specific health outcomes to changes in the licensing threshold.

The option to retake the exam creates a distinction between the immediate and future impacts of increasing the licensing threshold, because test-takers in any period t' include those retaking the exam after failing in an earlier period $t < t'$. As such, increasing the threshold in a period t_0 not

[42] We use $\Delta\underline{s} = 5$. Due to sparsity, we estimate these matching probabilities at the Health-Service level. We transform the matching probabilities into hospital-specific elasticities by assigning physicians to hospitals in proportion to their relative size within the health service. We provide more details on how we perform this calculation in Appendix J.

only impacts who passes in period t_0 but also the set of test takers in any period $t' > t_0$.

We evaluate the effects of increasing the licensing threshold over two time horizons. First, we evaluate the "short-run" effects ($R = 0$, according to the notation introduced at the end of Section 4) corresponding to the effects in year t of changing the increasing the threshold in year t. Second, we evaluate the "medium-run" effects ($R > 0$); reflecting the effects in year $t' > t$ of permanently increasing the threshold in period t.

7.1 Short-run Impacts of Increasing the Licensing Threshold

Figure V plots, for each hospital, the short-run elasticity of quantity and quality with respect to the licensing threshold in 2018. The size of the marker is proportional to the number of patients in each hospital, which we use as Pareto weights to compute the average elasticities shown in the solid diamond. On average, increasing the threshold by 1% would decrease labor by $\bar{\eta}_{\underline{s}}^{L_j} = -0.048$ % and increase average quality by $\bar{\eta}_{\underline{s}}^{\bar{\theta}_j} = 0.16$ standard deviation. The resulting average marginal rate of transformation is $\bar{R}(\underline{s}) \equiv -\bar{\eta}_{\underline{s}}^{\bar{\theta}_j} / \bar{\eta}_{\underline{s}}^{L_j} = 3.3$.

The dashed black line in Figure V corresponds to the "iso-mortality curve," which depicts the quantity elasticity needed to keep mortality constant, given the quality semi-elasticity. The inverse of the slope of this curve is $\hat{\alpha}_L^{\text{mortality}} / \hat{\alpha}_{\bar{\theta}}^{\text{mortality}} = 2.6$, as estimated in Section 5.2.

For most hospitals, and on average, we find that the marginal rate of substitution is higher than the marginal rate of technical transformation for the case of mortality, i.e., $R(\underline{s}) > \hat{\alpha}_L^{\text{mortality}} / \hat{\alpha}_{\bar{\theta}}^{\text{mortality}}$. In other words, the increase in quality (slightly) outweighs the decrease in quantity so that, on net, the elasticity of mortality with respect to the licensing threshold is negative.

The dashed green line in Figure V corresponds to the "iso-service rate curve," in which we substitute hospital mortality for service rate as the outcome. Since the quality of physicians does not significantly affect the hospital service rate, the iso-service rate curve is close to being a horizontal line at zero. As a result, the elasticity of the service rate with respect to the licensing threshold is negative.

In sum, raising the licensing threshold reduces access but lowers patient mortality. The desirability of this policy hinges on the tradeoff between its countervailing effects. Using estimates of the value of a statistical life (VSL), we find that, for the cost of reduced access to outweigh the mortality benefits, an additional hospitalization per year would need to generate at least US\$12,000 in benefits.[43] To

[43] We use the midpoint estimate of the VSL available for Chile (Mardones and Riquelme, 2018; Parada-Contzen, 2019), and apply the method of Murphy and Topel (2006) to adjust it based on the age profile of the patients in our sample.

put this figure in perspective, it corresponds to about 1.6 times the Chilean yearly minimum wage.

Relationship between Counterfactuals and Identifying Variation The variation in the labor supply that we exploit is well suited for our counterfactual analysis. As shown in our reduced-form exercise of Section 5.2, the variation in the labor supply experienced in Chile led to a decrease in quality and an increase in logged quantity at an average ratio of 2.3. Our policy counterfactual is qualitatively equivalent to "reversing" the migration wave—increasing the threshold decreases quantity but increases quality. It is also quantitatively similar, in that it increases quality and decreases logged quantity at an average ratio of 3.3. To visually assess the overlap between the changes generated in the counterfactual exercise and the identifying variation caused by the migration wave, Appendix Figure A.2 overlays the simulated changes in physician quantity and quality from our counterfactual with the instrument-predicted changes in quantity and quality for each hospital. Reassuringly, the changes induced in the counterfactual analysis are mostly within the convex hull of the identifying variation observed in the data, strengthening our confidence in the robustness of our results to potential misspecification of the production function.

Time-series Evolution of Short-run Impacts We turn to analyze how the effect of the licensing policy depends on the market-level fundamentals affected by the migration wave; the mass and quality distribution physicians— both of test-takers and of the stock of physicians in hospitals—, as well the level of market tightness in the labor market. Specifically, we turn to analyzing whether the policy effects found for 2018—5 years after the onset of the migration wave—differ from those in earlier years.

Panel A in Figure VI shows the evolution of short-run input elasticities over time. For the quantity elasticity (in red), we find an inverse U pattern. The quantity elasticity is the smallest in 2013 when there are few physicians at the margin of passing (see Table A.1). In the following years, the quantity elasticity increases as the migration increases the number of physicians close to the threshold. However, after 2016, the elasticity decreases. Even if the number of marginal physicians increases throughout the sample period, more physicians had already been hired when later cohorts arrived. This decreases the elasticity mechanically since the baseline is higher, but also through the lower capacity in the public sector which increases the share of physicians who match with the outside option as captured by the labor matching function. Conversely, the quality semi-elasticity has a U shape.

Panel B in Figure VI shows the short-run elasticities of mortality and service rate in for all years in our sample. Generally, mortality would decrease throughout the period, because the quality gains of increasing the threshold outweigh the quantity reductions. The service rate would also decrease regardless of the year in which the policy is enacted. For 2018, a 10% increase in the licensing threshold would decrease mortality by approximately 0.1% and would decrease the service rate by 0.5%. The evolution of the magnitudes of these effects mirrors the elasticities of inputs described in Panel A. The effects are maximal in 2016, when a 10% increase in the licensing threshold would decrease mortality by approximately 0.17% and would decrease the service rate by 1%.[44]

7.2 Medium-run Impacts of Increasing the Licensing Threshold

We now turn to analyzing the dynamic effects of permanently changing the licensing threshold. The critical mechanism we stress in this subsection is that retaking might mitigate the relevance of the licensing threshold as test-takers improve their scores over attempts. For instance, 83% of the test-takers who failed the licensing exam on their first attempt in 2013 had passed it by 2018.

To quantify the differences between the short-run and medium-run effects of changing the threshold, we leverage the estimated model of test scores and retaking behavior described in Section 6. We compute the medium-rum impacts by forward-simulating the model under the status quo ($\underline{s} = 51$), and compare it to a counterfactual scenario in which we set the threshold *permanently* higher, starting in 2013.[45] Comparing outcomes across these scenarios for any year $t \geq 2013$ within our sample provides the "medium-run" elasticities.[46]

Table V shows the simulated passing rates for the 2013 cohort, under the current threshold ($\underline{s} = 51$) and under a counterfactual higher threshold ($\underline{s}' = 56$). In the status quo, 79.2% of test takers pass on their first attempt (in 2013), compared with 69.4% under the counterfactual policy. The 9.8 percentage points gap in passing rates determines the short-run elasticities. However, this gap narrows over time as individuals retake the exam. By 2018, 94.2% of physicians from the 2013 cohort pass under the status quo, compared with 90.1% under the counterfactual, reducing the gap to 4.1 percentage points. Appendix Figure A.3 illustrates the divergence between the short-run and the medium-run elasticity over time: while (by construction) short- and medium-run elasticities are identical at the time of implementation (2013), medium-run elasticities fall to roughly 70%

[44] As noted above, the effects of the policy are mediated by the fraction of physicians that match with the inside option. Therefore, to broaden the external validity of our results, we consider a counterfactual scenario where all physicians match with the inside option. In that case, the elasticities would be substantially higher: for 2018, a 10% increase in the threshold would reduce mortality by 1.1% and lower the service rate by 5.9%.

[45] As in the previous section, we estimate the numerical elasticities with $\overline{s} = 56$, and divide the results by 5.

[46] In all these simulations we account for the fact that matching probabilities change endogenously with the threshold. Matching probabilities in any year t depend on the mass of test takers who pass the exam through the term $\mathbf{M}_{it}(\underline{s})$ as well as the shares of physicians from each country through the term Share Physicians$_{c(i)j,t-1}$. Our simulations endogeneize these variables.

of their short-run values by 2018, as some marginal test takers who initially fail eventually pass in subsequent years. Importantly, despite this attenuation, increasing the threshold in 2013 still generates meaningful changes in average physician quality and quality in the following years and, on net, still generates positive impacts on mortality in the medium run.

8 Conclusion

Occupational licensing is a widely used regulation in labor markets to ensure a minimum quality standard. However, it comes at the cost of reducing the labor supply. This quantity-quality trade-off is particularly relevant in health care, where licensing is ubiquitous and its design may have first-order implications. In this paper, we first characterize this trade-off with a simple theoretical framework. Guided by the theory, we develop and estimate an empirical model to quantify the quantity-quality trade-off in physician licensing in Chile, where licensing is decided based on a minimum passing score on an exam.

Our empirical strategy is built around estimating two sets of sufficient statistics to evaluate the quantity-quality trade-off: the elasticities of inputs—the quantity and quality of labor—with respect to the licensing threshold, and the elasticities of hospital outputs—the service rate and mortality rate—with respect to those inputs. We begin by estimating production functions to provide novel results for the elasticity of hospital outcomes with respect to the quantity and quality of physicians. Key to our analysis, we leverage quasi-experimental variation in these inputs arising primarily from an unprecedentedly large immigration of physicians to Chile. We find that both the quantity and the quality of physicians affect access and quality of health care. As a consequence, changing the licensing threshold generally involves a quantity-quality trade-off.

To estimate the elasticities of inputs, we infer the latent quality of physicians from their individual exam score histories and introduce a microfounded labor matching function to model the sorting of physicians with different quality levels across hospitals under alternative licensing scenarios. Through a counterfactual analysis that raises the licensing threshold, we find that the reduction in mortality from improving physician quality slightly outweighs the increase in mortality due to the reduced physician quantity. In contrast, for the service rate, we estimate a negative effect from the reduction in physician quantity, with no detectable effect from the increase in quality. A simple back-of-the-envelope calculation suggests that, for the reform to be cost-neutral, a hospital admission per year should be valued at 1.6 times a yearly minimum wage.

These findings suggest that physician licensing reduces the number of available doctors but increases the quality of care. Comparing adjustments to licensing thresholds with alternative strategies—such as expanding access to high-quality medical education, which may increase the physician supply without compromising quality—offers a promising avenue for future research.

Although our quantitative results are specific to our context, our model and empirical strategy offers novel estimates for the role of quantity and quality on health care outcomes, and provide a rich framework for evaluating the stringency of licensing policies in other settings.

References

Abdulkadiroğlu, A. and T. Sönmez (1998). Random Serial Dictatorship and the Core from Random Endowments in House Allocation Problems. *Econometrica 66*(3), 689–701.

Acemoglu, D. and P. Restrepo (2020). Robots and jobs: Evidence from us labor markets. *Journal of Political Economy 128*(6), 2188–2244.

Adao, R., M. Kolesár, and E. Morales (2019). Shift-Share Designs: Theory and Inference. *The Quarterly Journal of Economics 134*(4), 1949–2010.

Agarwal, N. (2015). An empirical model of the medical match. *American Economic Review 105*(7), 1939–1978.

Agostinelli, F. and M. Wiswall (2025). Estimating the Technology of Children's Skill Formation. *Journal of Political Economy 133*(3), 846–887.

Ajzenman, N., P. Dominguez, and R. Undurraga (2023). Immigration, Crime, and Crime (Mis)Perceptions. *American Economic Journal: Applied Economics 15*(4), 142–176.

Altonji, J. G. and D. Card (1989, September). The Effects of Immigration on the Labor Market Outcomes of Natives. Working Paper 3123, National Bureau of Economic Research.

Anderson, D. M., R. Brown, K. K. Charles, and D. I. Rees (2020). Occupational Licensing and Maternal Health: Evidence from Early Midwifery Laws. *Journal of Political Economy 128*(11), 4337–4383.

Anderson, T. W. and H. Rubin (1949). Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations. *The Annals of Mathematical Statistics 20*(1), 46–63.

Angrist, J. D. and J. Guryan (2008). Does Teacher Testing Raise Teacher Quality? Evidence from State Certification Requirements. *Economics of Education Review 27*(5), 483–503.

ASOFAMECH (2009–2019). EUNACOM. https://www.eunacom.cl. Asociación de Facultades de Medicina de Chile (ASOFAMECH).

Association of American Medical Colleges (2023). 1 in 5 U.S. Physicians Was Born and Educated Abroad: Who Are They and What Do They Contribute? Accessed: 2024-07-30.

Association of American Medical Colleges (2024). The Complexities of Physician Supply and Demand: Projections From 2021 to 2036. Technical report, AAMC, Washington, DC.

Atal, J. P., J. I. Cuesta, and M. Sæthre (2022). Quality regulation and competition: Evidence from pharmaceutical markets. Technical report, National Bureau of Economic Research.

Autor, D. H., D. Dorn, and G. H. Hanson (2013). The China Syndrome: Local Labor Market Effects of Import Competition in the United States. *American Economic Review 103*(6), 2121–2168.

Azevedo, E. M. and J. D. Leshno (2016). A Supply and Demand Framework for Two-Sided Matching Markets. *Journal of Political Economy 124*(5), 1235–1268.

Bahar, D., I. Di Tella, and A. Gulek (2024). Formal Effects of Informal Labor and Work Permits: Evidence from Venezuelan Refugees in Colombia. Working paper.

Beerli, A., J. Ruffner, M. Siegenthaler, and G. Peri (2021). The abolition of immigration restrictions and the performance of firms and workers: evidence from switzerland. *American Economic Review 111*(3), 976–1012.

Borusyak, K., P. Hull, and X. Jaravel (2022). Quasi-Experimental Shift-Share Research Designs. *The Review of Economic Studies 89*(1), 181–213.

Borusyak, K., P. Hull, and X. Jaravel (2025). A Practical Guide to Shift-Share Instruments. *Journal of Economic Perspectives 39*(1), 181–204.

Borusyak, K., X. Jaravel, and J. Spiess (2024, 02). Revisiting Event-Study Designs: Robust and Efficient Estimation. *The Review of Economic Studies 91*(6), 3253–3285.

Brinatti, A. (2024). Third-country effects of us immigration policy. *Available at SSRN 4892498*.

Carrillo, B. and J. Feres (2019). Provider Supply, Utilization, and Infant Health: Evidence from a Physician Distribution Policy. *American Economic Journal: Economic Policy 11*(3), 156–96.

Castro, M. (2024). Salarios Hasta Ocho Veces Más Altos: El Nuevo Fenómeno de los "Médicos Gaviota" Argentinos en Chile. Accessed: 2024-04-11.

Chetty, R. (2009). Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods. *Annual Review of Economics 1*(1), 451–488.

Clay, K., G. Miller, N. M. Portnykh, and E. Schmick (2025). Medical school closures, market adjustment, and mortality in the flexner report era. *NBER*.

DEIS (2019). Egresos Hospitalarios. Available at http://www.deis.cl/estadisticas-egresoshospitalarios/. Accessed: 2020-02-09.

DEIS (2020). Listado de Establecimientos de Salud. Available at https://deis.minsal.cl/#datosabiertos. Accessed: 2020-02-09.

Dillender, M., A. T. Lo Sasso, B. J. Phelan, and M. R. Richards (2023). Occupational Licensing and the Healthcare Labor Market. Working Paper 29665, National Bureau of Economic Research.

Dodini, S., P. Lundborg, K. V. Løken, and A. Willén (2025). The Fatal Consequences of Brain Drain. *NHH Dept. of Economics Discussion Paper* (09).

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

Efron, B. and C. Morris (1973). Stein's Estimation Rule and Its Competitors: An Empirical Bayes Approach. *Journal of the American Statistical Association 68*(341), 117–130.

Einav, L., A. Finkelstein, and N. Mahoney (2025). Producing health: Measuring value added of nursing homes. *Econometrica 93*(4), 1225–1264.

Elixhauser, A., C. Steiner, D. R. Harris, and R. M. Coffey (1998). Comorbidity Measures for Use with Administrative Data. *Medical Care*, 8–27.

Equipo EUNAMed (2021, March). Aprobar el EUNACOM: ¿para qué sirve? https://eunamed.com/blog/para-que-sirve-aprobar-el-eunacom/. Accessed: 2025-07-15.

Fack, G., J. Grenet, and Y. He (2019). Beyond Truth-Telling: Preference Estimation with Centralized School Choice and College Admissions. *American Economic Review 109*(4), 1486–1529.

Finkelstein, A., M. Gentzkow, and H. Williams (2021). Place-Based Drivers of Mortality: Evidence from Migration. *American Economic Review 111*(8), 2697–2735.

Fletcher, J. M., L. I. Horwitz, and E. Bradley (2014). Estimating the Value Added of Attending Physicians on Patient Outcomes. Working Paper 20534, National Bureau of Economic Research.

Friedman, M. (1962). *Capitalism and Freedom*. University of Chicago Press.

Friedman, M. and S. Kuznets (1945). *Income from Independent Professional Practice*. NBER.

Gaynor, M. S., S. A. Kleiner, and W. B. Vogt (2015). Analysis of Hospital Production: An Output Index Approach. *Journal of Applied Econometrics 30*(3), 398–421.

Gilraine, M. and J. Penney (2023). Focused Interventions and Test Score Fade-Out. *The Review of Economics and Statistics*, 1–27.

Ginja, R., J. Riise, B. Willage, and A. Willén (2024). Does Your Doctor Matter? Doctor Quality and Patient Outcomes. *Journal of Political Economy: Microeconomics*.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020). Bartik Instruments: What, When, Why, and How. *American Economic Review 110*(8), 2586–2624.

Grieco, P. L. and R. C. McDevitt (2017). Productivity and Quality in Health Care: Evidence from the Dialysis Industry. *The Review of Economic Studies 84*(3), 1071–1105.

Groeger, A., G. León-Ciliotta, and S. Stillman (2024). Immigration, Labor Markets and Discrimination: Evidence from the Venezuelan Exodus in Perú. *World Development 174*, 106437.

Gruber, J., T. P. Hoe, and G. Stoye (2023). Saving Lives by Tying Hands: The Unexpected Effects of Constraining Health Care Providers. *The Review of Economics and Statistics 105*(1), 1–19.

Guarin, A., C. Posso, E. Saravia, and J. Tamayo (2021). The Luck of the Draw: The Causal Effect of Physicians on Birth Outcomes. Technical report.

Haakenstad, A., C. M. S. Irvine, M. Knight, C. Bintz, A. Y. Aravkin, P. Zheng, V. Gupta, M. R. M. Abrigo, A. I. Abushouk, O. M. Adebayo, G. . Agarwal, and R. Lozano (2022). Measuring the Availability of Human Resources for Health and Its Relationship to Universal Health Coverage for 204 Countries and Territories from 1990 to 2019: A Systematic Analysis for the Global Burden of Disease Study 2019. *The Lancet 399*(10341), 2129–2154.

Hernández, T. and Y. Ortiz Gómez (2011). La Migración de Médicos en Venezuela. *Revista Panamericana de Salud Pública 30*(2), 177–181.

Hoe, T. P. (2022). Does Hospital Crowding Matter? Evidence from Trauma and Orthopedics in England. *American Economic Journal: Economic Policy 14*(2), 231–62.

Hollingsworth, A., K. Karbownik, M. A. Thomasson, and A. Wray (2024). The gift of a lifetime: The hospital, modern medicine, and mortality. *American Economic Review 114*(7), 2201–2238.

Hoot, N. R. and D. Aronsky (2008). Systematic Review of Emergency Department Crowding: Causes, Effects, and Solutions. *Annals of Emergency Medicine 52*(2), 126–136.e1.

INE (2024). Estimación de la Población Extranjera Residente en Chile 2022: Resultados Instituto Nacional de Estadísticas. Accessed: December 11, 2024.

Kelly, E. and G. Stoye (2020). The Impacts of Private Hospital Entry on the Public Market for Elective Care in England. *Journal of Health Economics 73*, 102353.

Kleiner, M. M. (2011). Enhancing Quality or Restricting Competition: The Case of Licensing Public School Teachers. *Journal of Law and Public Policy 5*(2), 1–15.

Kleiner, M. M. (2014). Occupational Licensing in Health Care. In A. Culyer (Ed.), *Encyclopedia of Health Economics*. Elsevier.

Kleiner, M. M., A. Marier, K. W. Park, and C. Wing (2016, 5). Relaxing Occupational Licensing Requirements: Analyzing Wages and Prices for a Medical Service. *The Journal of Law and Economics 59*(2), 261–291.

Kleiner, M. M. and E. J. Soltas (2023, 02). A Welfare Analysis of Occupational Licensing in U.S. States. *The Review of Economic Studies 90*(5), 2481–2516.

Kleiner, M. M. and W. Wang (2023). The Labor Market Effects of Occupational Licensing in the Public Sector. Working Paper 31213, National Bureau of Economic Research.

Kunakov, N., L. Moraga, and L. Ortiz (2018). Revalidación de Títulos Médicos Extranjeros: Eficacia y Eficiencia de un Examen Colaborativo y Estandarizado. *Revista Médica de Chile 146*(2), 232–240.

Larsen, B., Z. Ju, A. Kapor, and C. Yu (2020). The Effect of Occupational Licensing Stringency on the Teacher Quality Distribution. Working Paper 28158, National Bureau of Economic Research.

Lebow, J. (2022). The Labor Market Effects of Venezuelan Migration to Colombia: Reconciling Conflicting Results. *IZA Journal of Development and Migration 13*(1), 1–49.

Lee, D. S., J. McCrary, M. J. Moreira, and J. Porter (2022). Valid t-ratio Inference for IV. *American Economic Review 112*(10), 3260–3290.

Mardones, C. and M. Riquelme (2018). Estimation of the Value of Statistical Life in Chile and Extrapolation to Other Latin American Countries. *Latin American Research Review 53*(4), 815–830.

Mena, B. (2021). Validity of the Single National Examination of Medical Knowledge (EUNACOM). *Validity of Educational Assessments in Chile and Latin America*, 353–369.

Menares, F. and P. Muñoz (2025). The Impact of Standardized Disease-Specific Healthcare Coverage. *Journal of Public Economics*, 105312.

MINSAL (2025). Glosa N°06 – Lista de Espera No GES y Garantías de Oportunidad GES Retrasadas. Government report, Subsecretaría de Redes Asistenciales, Ministerio de Salud, Santiago, Chile. Informe de ejecución presupuestaria, cuarto trimestre 2024, en cumplimiento de la Ley de Presupuesto N°21.640.

Murphy, K. and R. Topel (2006). The Value of Health and Longevity. *Journal of Political Economy 114*(5), 871–904.

Muñoz, P. and C. Otero (2025). Managers and public hospital performance. Working Paper 2025/02, University of Duisburg-Essen, CINCH – Health Economics Research Center.

Nikolova, S., M. Harrison, and M. Sutton (2016). The Impact of Waiting Time on Health Gains from Surgery: Evidence from a National Patient-Reported Outcome Dataset. *Health Economics 25*(8), 955–968.

Noghanibehambari, H. and J. Fletcher (2023). Long-term health benefits of occupational licensing: evidence from midwifery laws. *Journal of health economics 92*, 102807.

OECD (2019). Recent Trends in International Migration of Doctors, Nurses, and Medical Students. Accessed: 2024-06-17.

Okeke, E. N. (2023). When a doctor falls from the sky: The impact of easing doctor supply constraints on mortality. *American economic review 113*(3), 585–627.

Olivieri, S., F. Ortega, A. Rivadeneira, and E. Carranza (2022). The Labour Market Effects of Venezuelan Migration in Ecuador. *The Journal of Development Studies 58*(4), 713–729.

Parada-Contzen, M. V. (2019). The Value of a Statistical Life for Risk-Averse and Risk-Seeking Individuals. *Risk Analysis 39*(11), 2369–2390.

Pardo, C. (2019). Health Care Reform, Adverse Selection, and Health Insurance Choice. *Journal of Health Economics 67*, 102221.

Pizer, S. D. and J. C. Prentice (2011). What Are the Consequences of Waiting for Health Care in the Veteran Population? *Journal of General Internal Medicine 26*, 676–682.

Prentice, J. C. and S. D. Pizer (2007). Delayed Access to Health Care and Mortality. *Health Services Research 42*(2), 644–662.

Quan, H., V. Sundararajan, P. Halfon, A. Fong, B. Burnand, J.-C. Luthi, L. D. Saunders, C. A. Beck, T. E. Feasby, and W. A. Ghali (2005). Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data. *Medical Care*, 1130–1139.

RNPI (2024). Registro Nacional de Prestadores Individuales. `https://rnpi.superdesalud.gob.cl/`. Superintendencia de Salud de Chile.

Russo, A. (2023). Waiting or Paying for Healthcare: Evidence from the Veterans Health Administration. Technical report, Working Paper.

SIGTE (2015–2019). Sistema de Gestión de Tiempos de Espera. `https://sigte.minsal.cl/`. Ministerio de Salud de Chile.

SIRH (2011–2019). Sistema de Información de Recursos Humanos. `http://sirh.minsal.cl`. Ministerio de Salud de Chile.

Socha-Dietrich, K. and J. Dumont (2021). International Migration and Movement of Doctors to and Within OECD Countries - 2000 to 2018: Developments in Countries of Destination and Impact on Countries of Origin. Technical Report 126, OECD Publishing, Paris.

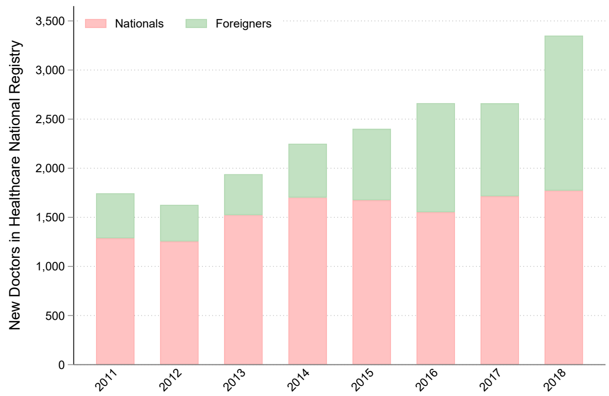
Svorny, S. (2004). Licensing Doctors: Do Economists Agree? *Economics Journal Watch 1*(2).

Taek Oh, Y. and M. M. Kleiner (2025). Does universal licensing recognition improve patient access? evidence from healthcare utilization.

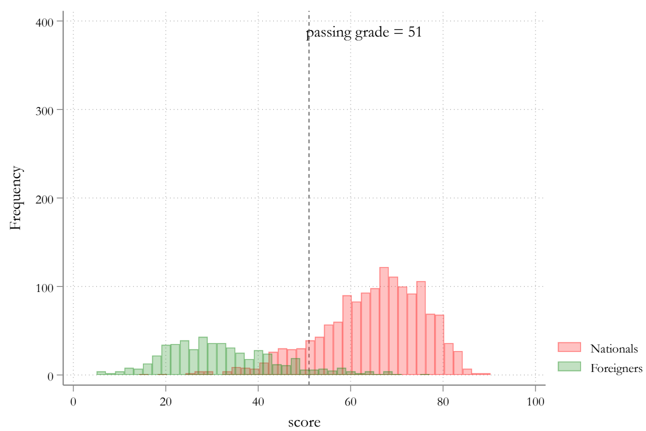
Walters, C. (2024). Empirical Bayes Methods in Labor Economics. In *Handbook of Labor Economics*, Volume 5, pp. 183–260. Elsevier.

Xia, X. (2021). Barrier to entry or signal of quality? the effects of occupational licensing on minority dental assistants. *Labour Economics 71*, 102027.

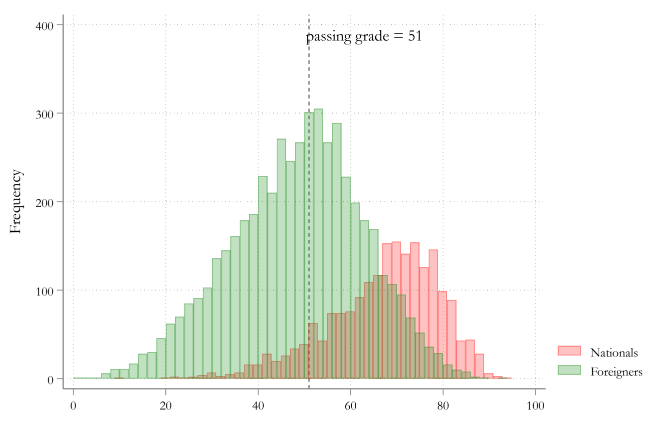
Figure I: Physicians' Labor Market



A. Registered Physicians



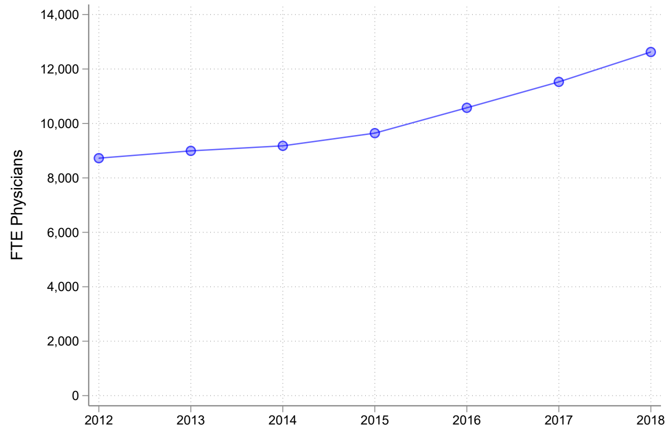
B. Scores in 2013



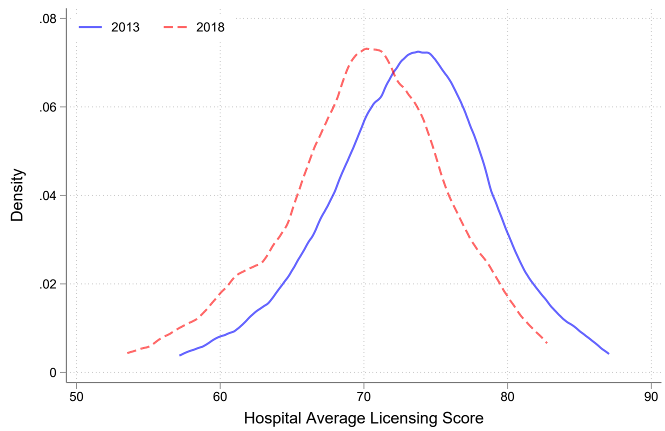
C. Scores in 2018

Notes: Panel A shows the number of newly registered physicians in the National Registry of Healthcare Providers between 2011 and 2018. This registry includes all healthcare workers legally allowed to practice in Chile. The bars are divided into locally trained physicians (red) and foreign-trained physicians (green). Panels B and C display the distribution of test scores for locally trained (red) and foreign-trained (green) physicians in 2013 and 2018, respectively. The dashed line in both graphs represents the minimum passing score.

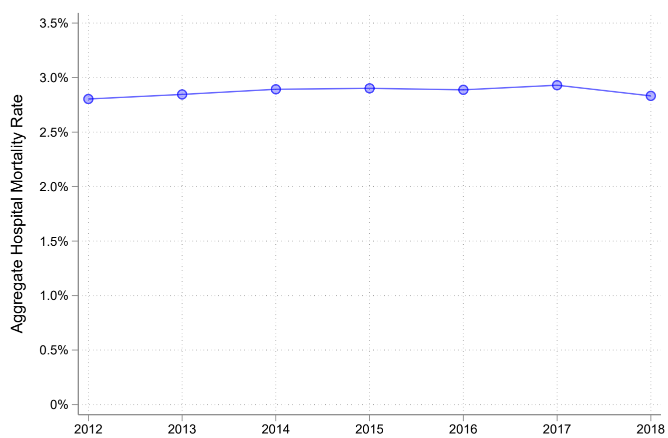
Figure II: Physicians in Public Hospitals



A. Quantity



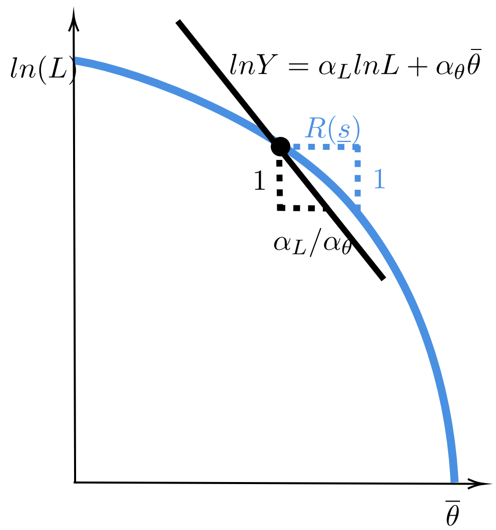
B. Scores



C. Hospital Mortality

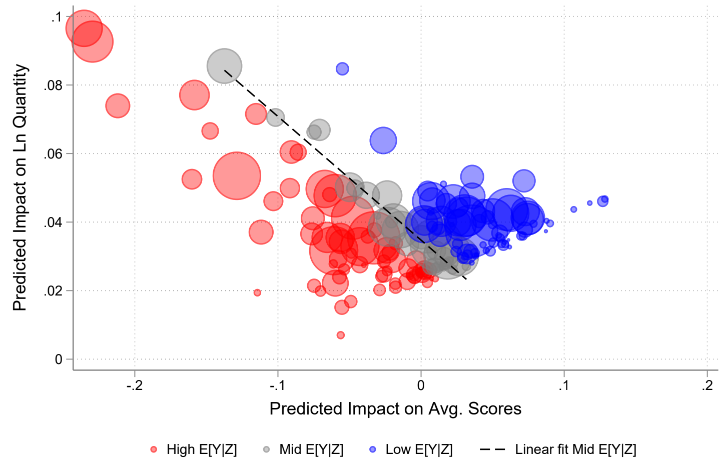
Notes: Panel A plots the number of full-time equivalent physicians working in public hospitals. Panel B presents the distribution of public hospitals' average licensing scores of physicians in 2013 and 2018. Panel C depicts the average aggregate mortality rate in public hospitals.

Figure III: Simple Licensing Problem

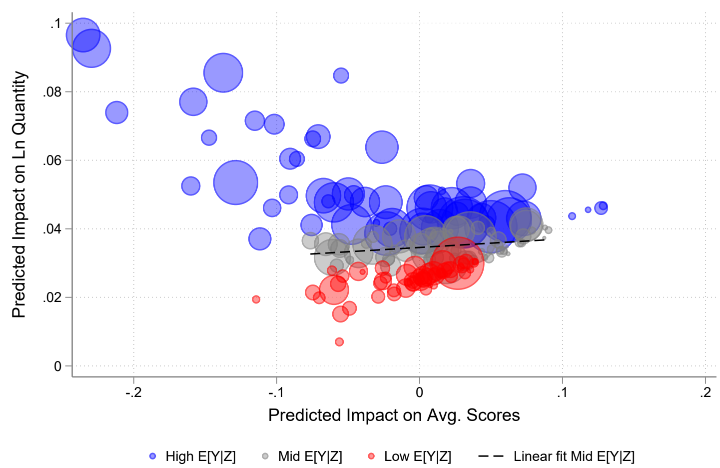


Notes: The figure depicts the licensing problem and the quantity-quality trade-off. The blue line represents the set of possible combinations between labor and quality that are feasible. When the threshold is increased from the point depicted by the dot, decreasing quality by one log point implies an increase in quality by $R(\underline{s})$. The solid black line represents the iso-outcome curve. The slope depends on the marginal product of the inputs. Increasing the licensing threshold would improve outcomes if $R(\underline{s}) > \alpha_L/\alpha_\theta$.

Figure IV: Reduced-form Plots



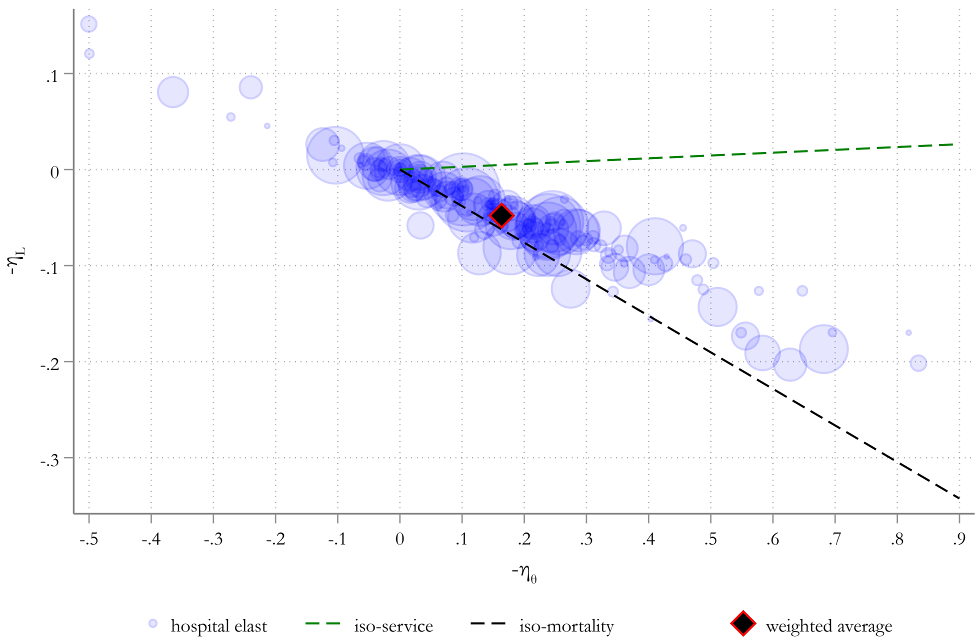
A. Mortality rate



B. Service rate

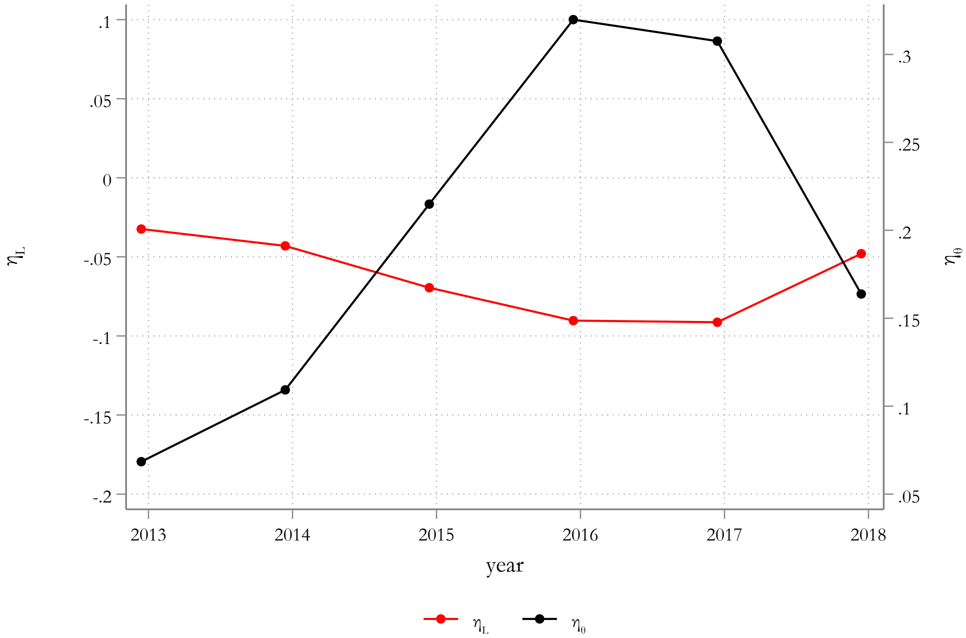
Notes: This figure plots the instrument-predicted quantity and quality of physicians for each hospital. Predictions are constructed using the first-stage coefficients and average instrument values by hospital, isolating only the input variation identified by the instruments. Marker sizes are proportional to hospital size at baseline. Hospitals are color-coded into terciles based on their predicted health outcomes (mortality rate in Panel A and service rate in Panel B), as determined by reduced-form regressions that use only the instruments as predictors. We include the linear fit of quantity and quality among hospitals with predicted outcomes near the sample average.

Figure V: Short-run Elasticity of Quantity and Semi-elasticity of Quality by Hospital in 2018

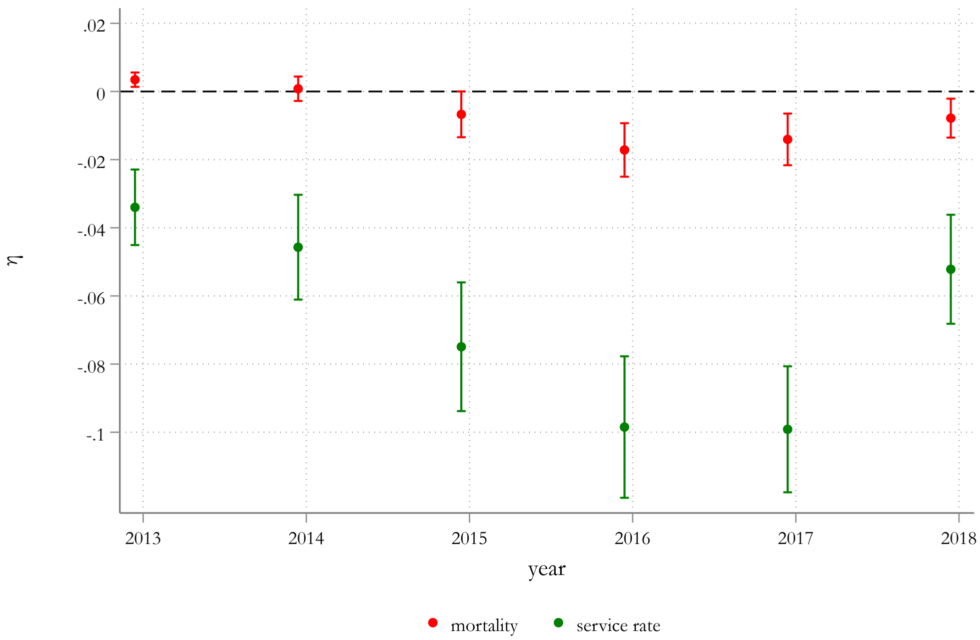


Notes: This figure shows a scatter plot with estimates for the (negative of) the elasticity of quantity and the semi-elasticity of quality with respect to the licensing threshold. Each dot represents a hospital. The size of the marker is proportional to the number of patients. Combinations of elasticities above the dashed lines are such that patient outcomes for the outcomes depicted by that line improve when the threshold is increased.

Figure VI: Time Series of Short-run Elasticities



A. Labor Quantity and Quality



B. Mortality and Access

Notes: Panel A shows the evolution of the average short-run elasticities of the quantity and quality of physicians with respect to the licensing threshold. Panel B shows the resulting average effects on patient mortality and access.

Table I: Impact of Physicians' Quantity and Quality on Access and Quality of Care

Panel A: Access

	Ln Service Rate	Ln Inpatient Surgeries	Ln Exits from Waiting List: Surgical	Ln Exits from Waiting List: Medical
	(1)	(2)	(3)	(4)
Ln Physicians ($\hat{\alpha}_L^{\text{service}}$)	0.99	4.73	3.81	2.99
	[0.00]	[0.00]	[0.00]	[0.00]
Std. Physicians' Scores ($\hat{\alpha}_\theta^{\text{service}}$)	-0.03	0.56	-0.13	0.02
	[0.57]	[0.21]	[0.13]	[0.81]
Observations	1,404	762	746	951
Mean Dep. Var.	0.014	3,741	1,530	8,397
F-stat (First-stage)	29.44	18.52	14.85	21.88
Anderson–Rubin (χ^2) p-value	0.00	0.00	0.00	0.00

Panel B: Quality

	Mortality			Complications
	Ln death rate In-Hospital	Pred. death rate In-Hospital	Ln death rate 28-Days	Ln readmission rate
	(1)	(2)	(3)	(4)
Ln Physicians ($\hat{\alpha}_L^{\text{mortality}}$)	-0.55	0.12	-0.48	-0.34
	[0.00]	[0.14]	[0.00]	[0.01]
Std. Physicians' Score ($\hat{\alpha}_\theta^{\text{mortality}}$)	-0.21	-0.01	-0.22	-0.23
	[0.00]	[0.66]	[0.00]	[0.00]
Observations	1,404	1,404	1,404	1,404
Mean Dep. Var.	3.30	3.53	5.11	3.46
F-stat (First-stage)	29.44	41.96	29.44	29.44
Anderson–Rubin (χ^2) p-value	0.01	0.13	0.00	0.00

Notes: This table presents the impact of the quantity and quality of physicians on public hospital performance. Panel A focuses on utilization, which we proxy through the service rate, inpatient surgeries, and exits from the waiting list. Panel B focuses on patients' mortality and complications. Estimates come from two-stage least squares estimation of Equation (5). The mean dependent variable is presented in levels instead of logs. We present exposure-robust p-values from models clustered at the region-of-origin (i.e., the shock) level in brackets (Adao et al., 2019; Borusyak et al., 2022).

Table II: Effects of Labor Supply Shocks on Inputs and Mortality

	Ln # Physicians	Std. Avg. Scores	Ln Death Rate	Ln Service Rate
	(1)	(2)	(3)	(4)
Z_{jt}^{L}	0.025	-0.057	-0.002	0.026
	(0.003)	(0.008)	(0.007)	(0.004)
Observations	1,404	1,404	1,404	1,404

Notes: The table shows the reduced-form effects of the quantity shift-share IV on the number of physicians, standardized average quality, log of the death date, and log of the service rate. All specifications include the set of controls in Equation (5). Robust standard errors in parentheses.

Table III: Model Estimates

Panel A: Model of Retaking

	Foreign-trained	Nationally trained
n_{it}	-0.231	-0.163
	(0.020)	(0.059)
$\bar{s} - s_{it}$	-0.036	-0.060
	(0.003)	(0.009)
Constant	2.592	1.595
	(0.077)	(0.139)
Observations	8,221	1,340

Panel B: Model of Scores

	Foreign-trained	Nationally trained
$\hat{\mu}_\theta$	46.002	65.474
	(0.864)	(0.107)
$\hat{\sigma}_\theta$	14.723	8.679
	(0.174)	(0.113)
$\hat{\sigma}_\epsilon$	9.227	8.677
	(0.190)	(0.072)
SNR	0.718	0.500
	(0.012)	(0.009)
Observations	9,756	17,335

	Common Parameters
$\hat{\gamma}$	9.599
	(0.541)
$\hat{\rho}$	0.285
	(0.050)

Notes: The table shows the estimated parameters of the dynamic model of score realizations and retaking probabilities. Panel A presents maximum likelihood estimates for the coefficients of the logit model for retaking, as specified in Equation (6), with standard errors in parentheses. Panel B presents the coefficients estimated via simulated method of moments (SMM) with bootstrapped standard errors, computed using the empirical standard deviation of the estimates across 10 simulations, in parentheses. We compute the moments by pooling the cohorts across our sample years and by forward-simulating the outcomes 100 times.

Table IV: Conditional Matching Probabilities Estimates

	Alternative Models			Placebo
	(1)	(2)	(3)	(4)
Distance_{ij}	-0.228	-0.228	-0.228	-0.228
	(0.014)	(0.014)	(0.014)	(0.014)
Share_{ijt-1}	0.651	0.651	0.653	0.652
	(0.146)	(0.146)	(0.146)	(0.146)
$(M_{it}^0)/\kappa_{jt}$	-0.637	-0.676	-0.726	
	(0.152)	(0.159)	(0.186)	
$(M_{it}^+)/\kappa_{jt}$		0.017	0.050	
		(0.021)	(0.038)	
$(M_{it}^0)/\kappa_{jt} \times \mathbb{1}[r_j = 2]$			0.004	
			(0.097)	
$(M_{it}^0)/\kappa_{jt} \times \mathbb{1}[r_j = 3]$			0.115	
			(0.122)	
$(M_{it}^+)/\kappa_{jt} \times \mathbb{1}[r_j = 2]$			-0.037	
			(0.045)	
$(M_{it}^+)/\kappa_{jt} \times \mathbb{1}[r_j = 3]$			-0.067	
			(0.056)	
$(M_{it}^-)/\kappa_{jt}$				0.022
				(0.019)
Log likelihood	-15276.84	-15276.53	-15275.33	-15285.23

Notes: The table shows results from a multinomial logit model for the matching probabilities between physicians and hospital referral regions. The estimating sample has 428,760 observations (30 options for 14,292 individuals). The functions $\mathbb{1}[r_j = k]$ indicate the hospital's quality tier. As shown in Equation (9), all specifications include alternative-specific coefficients for the following variables: Year, National indicator, physician's posterior quality, an interaction between a National indicator and posterior quality, and a Specialist indicator. We display only estimates for the coefficients that do not depend on the alternative.

Table V: Simulated Passing Year for 2013 Cohort

	$\underline{s} = 51$		$\underline{s} = 56$	
Year	Passing rate	Cumulative passing rate	Passing rate	Cumulative passing rate
2013	79.2	79.2	69.4	69.4
2014	10.9	90.1	14.4	83.8
2015	2.8	92.9	4.6	88.4
2016	0.7	93.6	1.1	89.5
2017	0.4	94.0	0.4	89.9
2018	0.2	94.2	0.2	90.1

Notes: The table shows the simulated passing rates over time for the cohort of physicians who take the exam for the first time in 2013, under a (baseline) scenario in which the threshold is kept at 51 points and under a counterfactual scenario in which the threshold is increased to 56 points.

ONLINE APPENDIX

Physicians' Occupational Licensing and the Quantity-Quality Trade-Off

Juan Pablo Atal, Tomás Larroucau, Pablo Muñoz, and Cristóbal Otero

List of Figures

List of Tables

A Score Imputation

The EUNACOM exam was introduced in 2009. Previously, physicians were not required to take a standardized test in order to practice medicine. To account for the quality of physicians licensed before EUNACOM, we impute their hypothetical scores. This appendix outlines the imputation procedure.

Formally, let y_{ih} denote the licensing score of physician i in hospital h. Our baseline model for scores is

$$y_{ih} = \mu_h + \alpha^h_{r(i)} + \varepsilon_{ih},$$

where μ_h represents the hospital-level mean, $\alpha^h_{r(i)}$ captures the region-of-origin effect (or "differential score") for physicians from region r working in hospital h, and ε_{ih} is an idiosyncratic error term with a mean of zero.

To improve out-of-sample predictions, which is particularly important when hospitals have very few physicians from a particular region, we apply an empirical Bayes shrinkage procedure (Efron and Morris, 1973; Walters, 2024). Specifically, we regress scores on region-of-origin fixed effects for each hospital h using OLS with sum-to-zero constraints (i.e., $\sum_r \alpha^h_r = 0; \forall h$). We then treat each α^h_r and its standard error s_h as coming from a prior distribution centered at 0—which reflects the sum-to-zero constraint—with limiting variance s^2_h. Given the OLS estimate $\widehat{\alpha}^h_r$ and its sampling variance $\widehat{\mathrm{Var}}(\widehat{\alpha}^h_r)$, the empirical Bayes estimate (posterior mean) is a weighted average of the OLS estimate and the prior mean (zero). Specifically:

$$\widetilde{\alpha}^h_r \;=\; \underbrace{\frac{\widehat{\mathrm{Var}}(\widehat{\alpha}^h_r)}{\widehat{\mathrm{Var}}(\widehat{\alpha}^h_r) + s^2_h}}_{\text{shrinkage factor}} \;\widehat{\alpha}^h_r.$$

Since the shrinkage factor ranges between 0 and 1, each estimated effect is pulled ("shrunk") toward zero, with the degree of shrinkage increasing as the precision of $\widehat{\alpha}^h_r$ decreases.

Imputation of Missing Scores: For a physician i at hospital h from region $r(i)$ who does *not* have an observed licensing score, we impute her score using the empirical Bayes-adjusted model:

$$\widehat{y}_{ih} \;=\; \widehat{\mu}_h \;+\; \widetilde{\alpha}^h_{r(i)},$$

where $\widehat{\mu}_h$ is the average score of physicians working at hospital h. Thus, the imputed score is a combination of the grand mean for hospital h ($\widehat{\mu}_h$), and the "shrunk" region-of-origin differential at hospital h ($\widetilde{\alpha}^h_{r(i)}$).

In very few cases (1.6%), the data do not include a region of origin for the physician. In these cases, we predict the individual score using a LASSO model. This model incorporates hospital indicators as well as controls for physicians' age and gender.

B Setting and Descript: Other Labor Inputs

To work in a public hospital, foreign-trained healthcare professionals must validate their degree in Chile. This process is overseen by specific recognized Chilean universities and involves submitting academic credentials, with additional coursework or exams required in some cases. For physicians, passing the EUNACOM is also necessary, because it qualifies them to practice in public healthcare settings.[47] After validation, they must register with the National Registry of Individual Health Providers by providing their validated degree and other required personal documentation.

We observe the impact of the migration wave only in the introduction of foreign-trained physicians. As described in the main text, there is a significant increase in the share of foreign-trained physicians, both among those newly enrolled in the National Registry of Healthcare Providers and those working in public hospitals. However, we do not find similar effects for other healthcare workers. As shown in Panel A of Figure A.5, the share of foreign-trained newly registered healthcare workers in the National Registry of Individual Health Providers remained fairly stable between 2000 and 2019, except for physicians. Panel B highlights the share of foreign-trained healthcare workers employed in public hospitals, and shows a significant increase for physicians—from 7% in 2011 to 18% in 2019. In contrast, there is no noticeable change in the share of foreign-trained workers in other healthcare categories.

C A Mapping with Chetty (2009)

Chetty (2009) provides a framework to compute the welfare effects of policies from a set of sufficient statics rather than from the entire set of model primitives. Chetty's main framework provides a rubric of steps to derive the sufficient statics in the context of a static single-agent model, in which the agent takes actions (e.g., decides consumption and leisure) to maximize utility subject to constraints that are affected by government policies (e.g., budget constraints affected by taxes and transfers). However, Chetty also notes that assuming optimizing behavior is not needed to derive sufficient statistics, as long as we can directly estimate the objects that determine the derivative of welfare with respect to the policy variable. We follow that approach in the paper.

For completeness, we discuss below how we could apply the optimization framework in our context. We recast our licensing problem by considering the problem of a representative hospital that optimally chooses labor and quality subject to constraints that are affected by the licensing threshold. We follow the same rubric of steps as in Chetty, and show how to use it to derive the sufficient statics formula in the main text (Equation 1).

[47] Passing the EUNACOM simultaneously qualifies physicians to practice and validates their degree, which reduces the incentive to pursue a separate degree validation process before taking the EUNACOM.

Using the notation of Section 4, the structure of the model would be given by

$$\begin{aligned} \max_{L,\bar{\theta}} \quad & F(L,\bar{\theta}) \\ \text{subject to} \quad & G_1(L,\underline{s}) \equiv L - m \int_{\underline{s}}^{\infty} h(s)\, p(s|\underline{s})\, ds = 0, \\ & G_2(\bar{\theta}, L, \underline{s}) \equiv L\,\bar{\theta} - \int_{\underline{s}}^{\infty} \theta(s)\, h(s)\, p(s|\underline{s})\, ds = 0, \end{aligned}$$

where $F(L,\bar{\theta})$ is the hospital's output as a function of the quantity and quality of labor. $G_1(L,\underline{s})$ and $G_2(L,\underline{s})$ are the constraints that link quantity and quality with the licensing threshold. Note that these constraints include the matching functions. The constraints therefore result, in part, from the matching process.

In the paper, we focus on the elasticity of hospital production with respect to the threshold, which is the planner's welfare criterion in Chetty's context. Welfare as a function of the licensing threshold is given by

$$Y(s) \equiv \max_{L,\bar{\theta}} F(L(s),\bar{\theta}(s)) + \lambda G_1(L,s) + \mu G_2(\bar{\theta}, L, s).$$

As in Chetty, we can use the envelope conditions and differentiate Y to get

$$\frac{dY}{ds} = \lambda \frac{\partial G_1}{\partial s} + \mu \frac{\partial G_2}{\partial s}. \tag{A.1}$$

In Chetty's framework, Lagrange multipliers are recovered from marginal utilities. In our framework, Lagrange multipliers are recovered from the output elasticities in the production function. Note that optimization implies that the marginal products of inputs are equated to linear combinations of the Lagrange multipliers:

$$\frac{\partial F}{\partial L} = \lambda \frac{\partial G_1}{\partial L} + \mu \frac{\partial G_2}{\partial L} \tag{A.2}$$

and

$$\frac{\partial F}{\partial \bar{\theta}} = \mu \frac{\partial G_2}{\partial \bar{\theta}}. \tag{A.3}$$

Also, note that differentiating $G_1(L,s)$ with respect to s yields

$$\frac{dL}{ds} = \frac{\partial G_1}{\partial s} \cdot \frac{1}{\frac{\partial G_1}{\partial L}}. \tag{A.4}$$

Similarly, differentiating $G_2(L,s,\bar{\theta})$ with respect to s yields

$$\frac{d\bar{\theta}}{ds} = \frac{\partial G_2}{\partial s} \cdot \frac{1}{\frac{\partial G_2}{\partial \bar{\theta}}} - \frac{\partial G_2}{\partial L} \cdot \frac{1}{\frac{\partial G_2}{\partial \bar{\theta}}} \cdot \frac{\partial G_1}{\partial s} \cdot \frac{1}{\frac{\partial G_1}{\partial L}}. \tag{A.5}$$

Combining equations A.1-A.5 yields the sufficient statistics formula (Equation (1) in the main text).

We also note that Chetty (2009) proposes to use the first order conditions to derive the marginal utilities from observed choices. In our context, this would entail linking the marginal product of inputs to the hospital's input decisions (e.g., using factor shares). Instead, we rely on exogenous variation in inputs to estimate those parameters without imposing optimality.

A crucial feature of the above framework is the assumption that agents maximize utility—although, as Chetyy points out, it is not a necessary feature of the sufficient statistics framework as long as we can identify the relevant elasticities. However, a common challenge in identifying the elasticities without imposing maximization is that marginal utilities are unobserved. Imposing optimizing behavior allows researchers to recover marginal utilities from observed choices. One feature of our application is that we focus on an *observable* outcome (hospital production), and therefore can measure marginal utilities (the marginal products of labor and quality) directly from the data.

D A Simple Microfoundation for the Quantity-Quality Trade-off

Consider a continuum of patients arriving with a rate λ_j to hospital j. Each patient i is treated at a hospital with quality-index Q_j. In the absence of treatment, patient i's "health capital" $\psi_i(w_j)$ deteriorates with waiting time w_j. The hospital's aggregate service rate is $s_j = s(L_j)$, which is a strictly increasing function of the number of physicians L_j. In the steady state–under mild assumptions on the queuing system–the expected waiting time $\bar{w}_j$ satisfies

$$\bar{w}_j = G(\lambda_j, s_j),$$

with

$$\frac{\partial G}{\partial s_j} < 0,$$

so that $\bar{w}_j$ decreases when s_j (and hence L_j) increases. A representative patient's health outcome is given by

$$Y_{ij} = Y(\psi_i(\bar{w}_j), Q_j),$$

with

$$\frac{\partial Y}{\partial Q_j} > 0, \qquad \frac{\partial Y}{\partial \bar{w}_j} < 0.$$

Thus, decreasing L_j worsens outcomes by increasing $\bar{w}_j$ and increasing Q_j improves outcomes via increasing care quality. Therefore, in our licensing setting, changing the licensing stringency entails a fundamental trade-off between quantity and quality in determining aggregate health care outcomes such as overall mortality rates.

E Shift-share Instruments

To assess the robustness of our instrumental variables approach, we build on a recent econometric literature that suggests two distinct paths to identification. One path, developed by Borusyak et al. (2024) and Adao et al. (2019), leverages many exogenous shifts while making no assumption about the exogeneity of the shares. The second path, proposed by Goldsmith-Pinkham et al. (2020), instead focuses on share exogeneity. As pointed out by Borusyak et al. (2025), identification "from the shifts" can be understood as leveraging a shift-level natural experiment, while identification "from the shares" can be viewed as pooling multiple difference-in-differences designs that leverage heterogeneous shock exposure. In this paper, we rely on identification from the shifts.

Shift-based identification stems from the observation that a share-weighted average of random shifts is itself as good as random (Borusyak et al., 2025). This is true even if the shares are econometrically endogenous, in the sense that units with different shares may have systematically different unobservables. Indeed, Borusyak et al. (2022) show that classical shift-share IV regression coefficients are numerically equivalent to those obtained from a regression in which the outcome and treatment variables are first averaged, using exposure shares as weights, and the shocks are directly used as instruments for the aggregated treatment.

The fact that shift-share estimates can be equivalently obtained by a shock-level IV procedure suggests ways to establish their consistency. Following Borusyak et al. (2022), we begin by showing statistics of the shocks. Table A.2 shows the distribution of the quantity and quality shocks. As shown in columns (1) and (3), the distribution of the quantity shocks has an average of -0.2, a standard deviation of 0.8, and an interquartile range of 0.2; the distribution of the quality shocks has an average of -0.4, a standard deviation of 0.7, and an interquartile range of 1.3. Columns (2) and (4) show that there is residual shock variation even conditional on period fixed effects. The inverse HHI of the exposure shares is 28.8 across region-by-period cells. The largest shock weights are 6% across region-by-periods cells.

To corroborate the plausibility of the conditional quasi-random shock assignment, we perform a shock balance test. If the migration shocks are as-good-as-randomly assigned, we expect them not to predict predetermined variables related to hospitals' workforce and patients' demographics. Figure A.8 reports the results of our balance tests on potential confounders. Panel A focuses on the quantity shocks and Panel B on the quality shocks. Reassuringly, we find that there is no statistically significant relationship between most variables and the shocks. There is some evidence of imbalance between the quantity shock and a few patient and hospital characteristics. Nonetheless, these are variables we control for in our main analysis.

F Do Other Inputs Affect the Results?

In this appendix, we first document the existence of complementarities between physicians and other healthcare workers at hospital level and then discuss the implications for identifying physicians' returns on hospital outcomes.

Panel A in Appendix Figure A.9 presents a scatter plot of the number of physicians and other FTE healthcare workers at the hospital. An observation is hospital-by-year. The figure shows a strikingly linear relationship between physicians and other inputs. We further examine whether the mix of other inputs relative to physicians changed with the increase in the number of physicians by differentiating between the pre- and post-migration wave and do not find differences between them. Since both sets of observations overlap, this provides evidence that as hospitals hired more physicians, they hired additional inputs at a similar rate.

A related question is whether the mix of physicians and other healthcare workers varies according to physician quality. In Panel B, we plot the ratio of other healthcare workers to physicians against average physician quality. We do not find that hospitals with higher-quality physicians substitute away from other inputs.

Given this evidence, it is natural to ask about the implications of incorporating other inputs in the production function as a function of the number of physicians. Since the ratio of other inputs to physicians does not change with physician quality, we assume away quality for ease of exposition. The production function takes the following form: $Y_i = AL_i^{\alpha_L} O_i^{\alpha_O}$, with $O_i = e^c L_i^{\gamma}$. Thus,

$$\ln Y_i = \phi + \underbrace{(\alpha_L + \gamma\alpha_O)}_{\tilde{\alpha}_L} \ln L_i.$$

Given the complementarities between physicians and other inputs, the impact of physicians on the outcome of interest, $\tilde{\alpha}_L$, is a bundled effect: a direct effect of an extra doctor, α_L, and an indirect effect from the increase in other inputs, $\gamma\alpha_O$.

We next discuss the implications for identification in our setting, in which the quantity of physicians and the number of other inputs are endogenous. Assume that we specify the following estimating equation:

$$y_i = \phi + \alpha_L l_i + \alpha_O o_i + \epsilon_i,$$

where y_i is a logged outcome (e.g., death rate, service rate, etc.); l_i is the logged quantity of doctors; and o_i is the logged number of other healthcare workers. Assume that l_i and o_i are endogenous (i.e., $\text{Cov}(l_i, \epsilon_i) \neq 0$, and $\text{Cov}(o_i, \epsilon_i) \neq 0$). Further, assume a plausibly exogenous instrument z_i for l_i, such that $\text{Cov}(l_i, z_i) \neq 0$, but which might be correlated with o_i ($\text{Cov}(z_i, o_i) \neq 0$), which could violate the exclusion restriction. The IV estimator of α_L is given by

$$\begin{aligned}\hat{\alpha}_L^{IV} &= \frac{\text{Cov}(y_i, z_i)}{\text{Cov}(l_i, z_i)} \\ &= \alpha_L + \alpha_O \frac{\text{Cov}(o_i, z_i)}{\text{Cov}(l_i, z_i)},\end{aligned}$$

where the second step follows from the assumption that $\text{Cov}(\epsilon_i, z_i) = 0$. Incorporating the complementarity between physicians and other inputs as $o_i = c + \gamma l_i + \nu_i$,

$$\begin{aligned}\hat{\alpha}_L^{IV} &= \alpha_L + \alpha_O \frac{\text{Cov}(c + \gamma l_i + \nu_i, z_i)}{\text{Cov}(l_i, z_i)} \\ &= \underbrace{\alpha_L + \gamma\alpha_O}_{\tilde{\alpha_L}} + \underbrace{\alpha_O \frac{\text{Cov}(\nu_i, z_i)}{\text{Cov}(l_i, z_i)}}_{bias}.\end{aligned}$$

Thus, $\hat{\alpha_L}^{IV}$ is the bundled effect of the direct and indirect effects of more physicians on the outcome, plus *bias* if $\text{Cov}(\nu_i, z_i) \neq 0$ (i.e., if innovations that shift o_i correlate with z_i). We can empirically gauge the magnitude of the bias by estimating the linear projection of the instrument on other inputs:

$$o_i = \alpha + \beta z_i + e_i.$$

Using the fact that $\beta = \frac{Cov(o_i, z_i)}{Var(z_i)}$, it is straightforward to show that

$$\hat{\text{Cov}}(\nu_i, z_i) = \hat{\beta}\hat{\text{Var}}(z_i) - \gamma\hat{\text{Cov}}(l_i, z_i).$$

Empirically, we obtain $\hat{\beta} = 0.01$, $\hat{\text{Var}}(z_i) = 0.97$, $\hat{\gamma} = 0.49$, and $\hat{\text{Cov}}(l_i, z_i) = 0.02$. Adding these numbers, we find that $\hat{\text{Cov}}(\nu_i, z_i) \approx 0$, which implies that the bias is also close to zero. Complementary evidence for these results comes from the trends in labor supply of L_i and O_i over time, as discussed in Appendix B. The evidence therein shows that migration shocks, which are the main quasi-exogenous variation leveraged by our instrument, mostly shifted the supply of physicians, with negligible effects on other inputs.

G Heterogeneous Effects by Mortality Risk of Diagnosis

In this appendix, we explore whether the impacts of physician quantity and quality vary depending on the riskiness of patients' conditions. Certain medical conditions, such as cardiovascular diseases, carry higher mortality risks compared to others, like pregnancy; thus, physician staffing may be more influential for some diagnoses than others. To study this heterogeneity, we estimate a predictive model that identifies which ICD-10 codes correspond to diagnoses associated with higher mortality rates. We then separately estimate the model on high- and low-risk patient samples.

We first fit a logistic regression model to predict the probability of death associated with each diagnosis category, using individual-level data from 2005 to 2011. We then use the estimated coefficients to compute predicted mortality risks (or risk scores) for each patient in the full sample.[48] We classify patients as having a high-mortality condition if their predicted mortality falls within the top 20% of the distribution; all others are classified as having a low-mortality condition. Based on this classification, we construct high- and low-mortality samples for each hospital-year.

To examine heterogeneous effects, we estimate the same two-stage least squares (2SLS) model described in Equation (5) separately for the high- and low-mortality samples. The results are presented in Table A.3. Panel A presents results based on the sample of individuals whose predicted risk scores place them in the top 20% of the risk distribution. Panel B includes the results for the remaining patients. For each sample, we recalculate patient characteristics and outcomes at the hospital level using only the corresponding subset of patients.

Column (1) shows that, as expected, average in-hospital mortality across hospitals is substantially higher for patients in the top 20% of predicted risk (9.6%) compared to those in the bottom 80% (1.2%). In terms of output elasticities, we find that both the quantity and quality of physicians significantly affect mortality only for patients diagnosed with high-mortality conditions—this is true for both in-hospital mortality (Column 1) and 28-day mortality (Column 2). Interestingly, however, both physician quantity and quality matter for hospital complication rates in the low-mortality sample, as shown in Column 3 of Panel B. This suggests that even though reductions in physician staffing levels or quality might not substantially impact mortality for lower-risk conditions, they do affect the quality of care more broadly, as reflected by complication rates. Column (4) presents an additional measure that combines readmission due to complications and deaths at the hospital-year level. This combined outcome is relevant because, for high-risk diagnoses, complication rates are mechanically lower as patients who die cannot be readmitted. This phenomenon may explain why the estimated effects on complications in Column 3 are larger for the low-risk sample (Panel B) than for the high-risk sample (Panel A).

H Quality Gains

In general, our model could allow for scores to improve over attempts due to increased test-taking ability as well as the improvements in quality. With data on scores alone, it is not possible to disentangle both mechanisms. In this section, we show that auxiliary data are consistent with no quality improvements over attempts. This justifies the assumption of no quality gains over

[48]We use diagnoses at the chapter and first-digit level of the ICD-10 codes, resulting in a total of 206 categories. A common approach to predicting in-hospital mortality is to estimate patient risk using the enhanced Elixhauser Index, which classifies diagnoses into 31 comorbidity categories and uses them as predictors in a risk model. However, this method excludes a substantial number of ICD-10 codes, limiting its coverage for our analysis. In results available upon request, we demonstrate that our prediction method outperforms predictions based on the Elixhauser Index.

attempts.

We leverage the discontinuity in retaking probability around the passing threshold to show that retaking does not improve labor-market outcomes that proxy for quality. Figure A.7-A shows the probability of ever retaking the test as a function of the distance from the score on the first attempt to the threshold. It shows that the retaking probability drops substantially around the cutoff. Figure A.7-B shows that the maximum achieved score changes discontinuously around the threshold. Note that score gains in Panel B are a combination of gains in test-taking ability, gains in quality, and selection around cutoff (Gilraine and Penney, 2023).

We investigate how two proxies of quality change around the threshold. Figure A.7-C shows the probability of being selected as *Médico General de Zona*—a merit-based position for physicians in Chile who serve in underserved areas and grants elegibility for state-funded specialization. The likelihood increases substantially with the score on the first attempt. However, there are no discernible increases around the threshold. Similarly Figure A.7-D shows a (more modest) increase in the hours of contract with the score on the first attempt, with no discernible increases around the threshold. Together, we take this as evidence that contradicts the idea that there are quality improvements from retaking.

I Microfoundation of CMPs

We now provide a micro-foundation for the reduced-form CMPs presented in the main text. We consider a static version of the model and fix t. Consider a continuum of physicians indexed by $i \in \mathcal{I}$. Each physician i belongs to a type $\tau_i \in \{N, F\}$, with a total mass m_τ for each type τ. Let $\mathcal{J}$ be a finite set of hospitals, each indexed by j and endowed with $\kappa_j \in \mathcal{K}$ vacancies, and let ϕ be physicians' outside option.

Physician preferences. The physician's indirect utility from matching with hospital $j \in \mathcal{J}$ is given by

$$u_{ij} \;=\; \tilde{v}\left(x_{ij}, \hat{\theta}(\vec{s}_i|\tau_i)\right) \;+\; \varepsilon_{ij}, \tag{A.6}$$

Hospital preferences. Hospitals maximize an objective function that is strictly increasing in the expected quality $\hat{\theta}$ of the physicians they match with and find all to be admissible. Strict monotonicity in $\hat{\theta}$ implies that if a hospital j can replace a physician of lower expected quality with one of higher expected quality, it will strictly prefer to do so.

Equilibrium. An equilibrium allocation satisfies *pairwise stability*:

(i) *Individual rationality*: No matched physician prefers their outside option to their allocation.

(ii) *No blocking pairs*: No hospital-physician pair wants to deviate from their current match and match with each other.

Following Azevedo and Leshno (2016), the equilibrium can be characterized by a system of demand and supply equations that determine a set of—expected quality—cutoffs. In particular, each hospital j admits any physician i with $\hat{\theta}_i$ above their cutoff $\underline{\hat{\theta}}_j$, while each physician i chooses the hospital j that maximizes her indirect utility from among those that would admit her. Mathematically, i matches with j—i.e., $\mu(i) = j$, if and only if

$$\mu(i) = \arg\max_{j \in \mathcal{J}_i} u_{ij},$$

where $\mathcal{J}_i = \{j \in \mathcal{J} : \hat{\theta}_i \geq \underline{\hat{\theta}}_j\} \cup \{\phi\}$.

Given our distributional assumptions on physicians' unobserved preferences, the conditional matching probability for a physician i matching with hospital j can be written as

$$CMP^m\big(\mu(i) = j\big) \ = \ \frac{e^{\tilde{v}\left(x_{ij}, \hat{\theta}(s_i|\tau_i)\right)}\, \mathbf{1}\{\hat{\theta}(s_i|\tau_i) \geq \underline{\hat{\theta}}_j\}}{1 \ + \ \sum_{j'} e^{\tilde{v}\left(x_{ij'}, \hat{\theta}(s_i, \tau_i)\right)}\, \mathbf{1}\{\hat{\theta}(s_i|\tau_i) \geq \underline{\hat{\theta}}_{j'}\}}, \tag{A.7}$$

where the indicator $\mathbf{1}\{\hat{\theta}(s_i|\tau_i) \geq \underline{\hat{\theta}}_j\}$ requires that we only consider hospitals that are ex post feasible for physician i (Fack et al., 2019), and the equilibrium cutoffs are determined by the following system of demand and supply equations:

$$\kappa_j \geq M_j\left(\underline{\hat{\theta}}\right) \forall j \in \mathcal{J}, \tag{A.8}$$

where

$$M_j\left(\underline{\hat{\theta}}\right) = \sum_{\tau} m_\tau \int_{\underline{\hat{\theta}}_j}^{\bar{\hat{\theta}}} \int_{\mathcal{X}} \frac{e^{\tilde{v}(X,y|j)}\, \mathbf{1}\{y \geq \underline{\hat{\theta}}_j\}}{1 \ + \ \sum_{j'} e^{\tilde{v}(X,y|j')}\, \mathbf{1}\{y \geq \underline{\hat{\theta}}_{j'}\}} h\left(X, y|\tau\right) dX dy \tag{A.9}$$

with $\underline{\hat{\theta}}$ denoting the vector of posterior quality cutoffs, $h\left(X, y|\tau\right)$ the joint density over observable characteristics X and posterior quality y given type τ, and $\bar{\hat{\theta}}$ the upper bound in the domain of posterior mean qualities.

The previous equations make explicit the dependency of equilibrium cutoffs on hospitals' vacancies and the mass and joint distribution of physicians' preferences and posterior qualities.

Approximation of CMPs. We now provide conditions under which our reduced-form CMPs can be seen as approximations of the micro-founded CMPs. The reduced-form CMPs for this simplified model, specified in Equation (8), would be given by

$$CMP^r\big(\mu(i)=j\big) \;=\; \frac{e^{v\left(x_{ij},\hat{\theta}(\vec{s}_i|\tau_i)\right)+g(\mathbf{M}_i(\underline{s}),\kappa|j)}}{1+\sum_{j'} e^{v\left(x_{ij'},\hat{\theta}(\vec{s}_i|\tau_i)\right)+g(\mathbf{M}_i(\underline{s}),\kappa|j')}}. \tag{A.10}$$

Therefore, the reduced-form CMPs given in Equation (A.10) can approximate the micro-founded CMPs from Equation (A.7) if:

$$e^{\Delta_v\left(x_{ij},\hat{\theta}(\vec{s}_i|\tau_i)\right)+g(\mathbf{M}_i(\underline{\mathrm{S}}),\kappa|j)} \;\approx\; \mathbf{1}\{\hat{\theta}(s_i,\tau_i)\geq \underline{\hat{\theta}}_j\} \quad \forall j\in\mathcal{J}, \tag{A.11}$$

where $\Delta_v\left(x_{ij},\hat{\theta}(\vec{s}_i|\tau_i)\right) \equiv v\left(x_{ij},\hat{\theta}(\vec{s}_i|\tau_i)\right) - \tilde{v}\left(x_{ij},\hat{\theta}(\vec{s}_i|\tau_i)\right)$.[49]

To establish under which conditions Equation (A.11) might hold, notice that the indicator function can be smoothed out by using the Gompertz function,

$$e^{-e^{-\lambda\left(\hat{\theta}(\vec{s}_i|\tau_i)-\underline{\hat{\theta}}_j\right)}} \approx \mathbf{1}\{\hat{\theta}(\vec{s}_i|\tau_i)\geq \underline{\hat{\theta}}_j\} \quad \forall j\in\mathcal{J},$$

which converges to an indicator as $\lambda\to\infty$. Thus, our reduced-form CMPs can approximate their micro-founded counterparts if

$$\Delta_v\left(x_{ij},\hat{\theta}(\vec{s}_i|\tau_i)\right) + g(\mathbf{M}_i(\underline{\mathrm{s}}),\kappa|j) \approx e^{-\lambda\left(\hat{\theta}(\vec{s}_i|\tau_i)-\underline{\hat{\theta}}_j\right)}. \tag{A.12}$$

Equation (A.12) suggests that for our reduced-form CMPs to be a reasonable approximation of their micro-founded counterparts, the function $g(\cdot)$ must capture the determinants of the equilibrium cutoffs $\underline{\hat{\theta}}$. As shown in Equations (A.8) and (A.9), these equilibrium cutoffs are the solutions to a system of supply and demand equations that depend on the distribution of preferences, the masses of physicians, and hospitals' capacities.

In counterfactual analyses, changes in the mass of physicians with different preferences and qualities might affect the equilibrium cutoffs in complex ways. However, under the assumption of vertical sorting, an important restriction arises: the (ex post feasible) choice set—and thus the matching probabilities—of a physician with posterior quality $\hat{\theta}_i$ can only be affected by changes in the mass of physicians with qualities above $\hat{\theta}_i$.

To see this more concretely, consider a large matching market with a finite set of agents and a unique stable matching. In this case, the equilibrium allocation coincides with the outcome of the serial dictatorship (SD) mechanism (Abdulkadiroğlu and Sönmez, 1998), in which physicians are ordered by descending posterior quality and sequentially choose their most preferred hospital among those with remaining vacancies. Since hospitals have vertical preferences, a physician's assignment

[49] Notice that because $v\left(x_{ij},\hat{\theta}(\vec{s}_i|\tau_i)\right) = \Delta_v\left(x_{ij},\hat{\theta}(\vec{s}_i|\tau_i)\right)+\tilde{v}\left(x_{ij},\hat{\theta}(\vec{s}_i|\tau_i)\right)$, in our reduced-form CMPs we will not be able to separately identify how much of the determinants of matching probabilities (v) are affected by physicians' preferences ($\tilde{v}$), versus the effect of their observable characteristics on their choice sets (Δ_v).

is determined solely by her relative position in the quality ranking. When physician i makes her choice, the set of available hospitals depends only on the prior selections made by physicians with higher quality. Thus, changes in the mass of physicians with quality below $\hat{\theta}_i$ do not alter the set of options available to her, and consequently leave her allocation and conditional matching probability unchanged.

Using this result from our micro-founded model, we specify the function $g(\cdot)$ to depend on programs' vacancies κ and on the mass of physicians with quality above i's posterior quality, denoted by M_i.

J Definition and Computation of Elasticities

Hospital-specific Elasticities: The labor-matching probabilities and the licensing threshold determine the hospital-specific inflows of physicians and their respective quality. For a physician $i \in \mathcal{I}$ with type $\tau \in \mathcal{T}$, exam score s, and a matrix of physician–hospital characteristics $X \in \mathbb{R}^{K \times |\mathcal{J}|}$ (whose jth column is the vector $\boldsymbol{x}_{ijt} \in \mathbb{R}^K$ for hospital $j \in \mathcal{J}$), in a market characterized by (M_t, κ_t) in period t, we define the conditional probability that such a physician matches with hospital $j \in \mathcal{J}$ in period t as

$$CMP_j\big(X, s \,\big|\, \tau;\, M_t;\, \kappa_t;\, \underline{s}\big) \;\equiv\; CMP\Big(\big[\,x_{i1t}, \ldots, x_{i|\mathcal{J}|t}\,\big] = X,\; s_i = s,\; \tau_i = \tau,\; M_{it} = M_t,\; \kappa_t,\; \underline{s},\; j\Big).$$

Letting $h_t^\tau(s, X)$ be the type- and time-specific density of the joint distribution of observables and scores, the labor inflow of type τ in hospital j in period t is given by

$$\Delta L_{jt}^\tau(\underline{s}) \;=\; m_{\tau,t} \int_X \int_{s \geq \underline{s}} CMP_j\big(X, s \mid \tau;\, M_t;\, \kappa_t;\, \underline{s}\big)\, h_t^\tau(s, X)\, ds\, dX.$$

In addition, the average quality of physicians' inflow of type τ in hospital j at time t is given by

$$\Delta \bar{\theta}_{jt}^\tau(\underline{s}) \;=\; \frac{\displaystyle\int_X \int_{s \geq \underline{s}} CMP_j\big(X, s \mid \tau;\, M_t;\, \kappa_t;\, \underline{s}\big)\, \mathbb{E}\Big[\hat{\theta}(\vec{s} \mid \tau) \mid s, X\Big]\; h_t^\tau(s, X)\, ds\, dX}{\displaystyle\int_X \int_{s \geq \underline{s}} CMP_j\big(X, s \mid \tau;\, M_t;\, \kappa_t;\, \underline{s}\big)\, h_t^\tau(s, X)\, ds\, dX,}$$

where the term $\mathbb{E}[\hat{\theta}(\vec{s} \mid \tau) \mid s, X]$ denotes the conditional expectation of posterior quality for a physician of type τ, given current score s and characteristics X, averaging $\hat{\theta}(\vec{s} \mid \tau)$ over the distribution of score histories $\vec{s}$ consistent with s.

The labor in hospital j at time t and its corresponding average quality are given by

$$L_{jt}(\underline{s}) = L_{j,t-1} + \sum_{\tau \in \mathcal{T}} \Delta L_{jt}^{\tau}(\underline{s}) \tag{A.13}$$

$$\bar{\theta}_{jt}(\underline{s}) = \frac{1}{L_{jt}} \left(\bar{\theta}_{j,t-1} L_{j,t-1} + \sum_{\tau \in \mathcal{T}} \Delta L_{jt}^{\tau}(\underline{s}) \cdot \Delta \bar{\theta}_{jt}^{\tau}(\underline{s}) \right) \tag{A.14}$$

The elasticity of quantity and semi-elasticity of quality with respect to the licensing threshold follow directly from the expressions above.

Computation of Numerical Elasticities: To compute the elasticity of labor and quality, we simulate $R = 100$ draws for the EV1 shocks in the estimated CMPs and compute the resulting matches for physicians who take the test in year t and have a score above the corresponding threshold. If $\Delta^r L_{jt}(\underline{\mathrm{s}})$ physicians of average quality $\Delta^r \theta_{jt}(\underline{\mathrm{s}})$, we match with hospital j in simulation r; we compute the resulting quantity and quality of each hospital by adding the marginal physicians who match with hospital j in year t to the stock of physicians of hospital j in year $t-1$. That is, the resulting simulated physicians is $L_{jt}^r(\underline{\mathrm{s}}) = L_{jt-1} + \Delta^r L_{jt}(\underline{\mathrm{s}})$ and the resulting simulated average quality is $\bar{\theta}_{jt}^r(\underline{\mathrm{s}}) = \frac{(\bar{\theta}_{jt-1} L_{jt-1} + \Delta^r L_{jt} \Delta^r \theta_{jt}(\underline{\mathrm{s}}))}{L_{jt}^r(\underline{\mathrm{s}})}$.

We compute the quantity elasticity numerically as

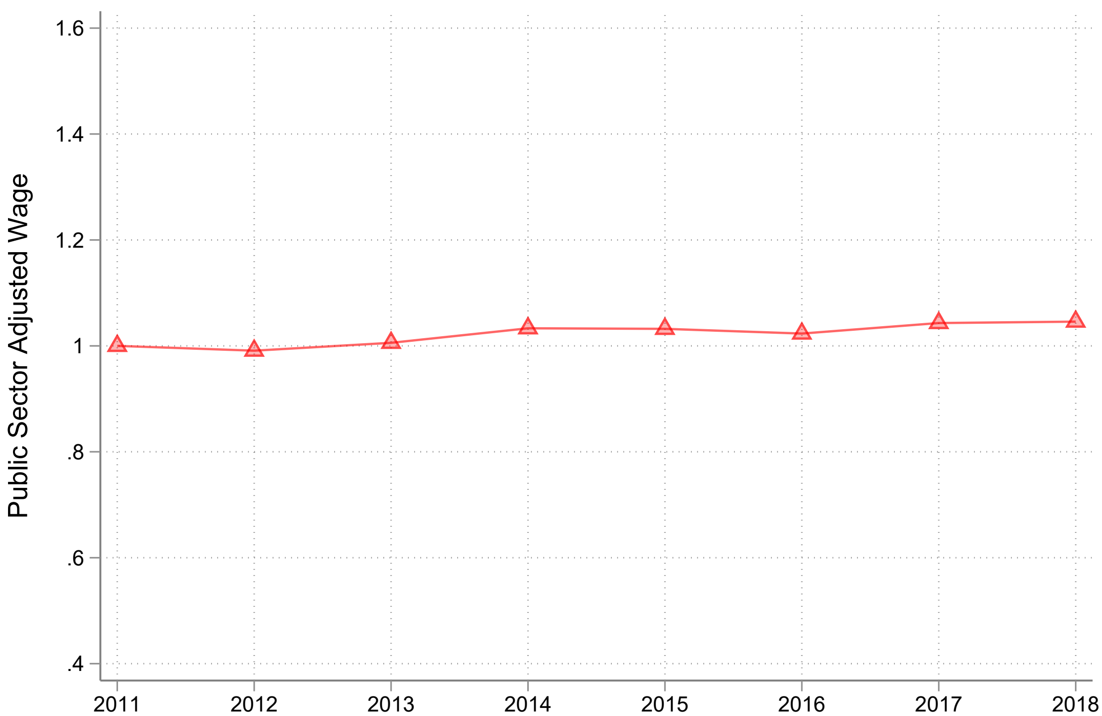
$$\eta_{L_{jt},\underline{s}}^r \approx -\frac{\left(L_{jt}^r(\underline{s} + \Delta \underline{\mathrm{s}}) - L_{jt}^r(\underline{\mathrm{s}}) \right)}{\Delta \underline{\mathrm{s}}} \cdot \frac{\underline{s}}{L_{jt}^r(\underline{\mathrm{s}})}$$

and the quality semi-elasticity as

$$\tilde{\eta}_{\bar{\theta}_{jt},\underline{s}}^r \approx -\frac{\left(\bar{\theta}_{jt}^r(\underline{s} + \Delta \underline{\mathrm{s}}) - \bar{\theta}_{jt}^r(\underline{\mathrm{s}}) \right)}{\Delta \underline{\mathrm{s}}} \cdot \underline{s}.$$

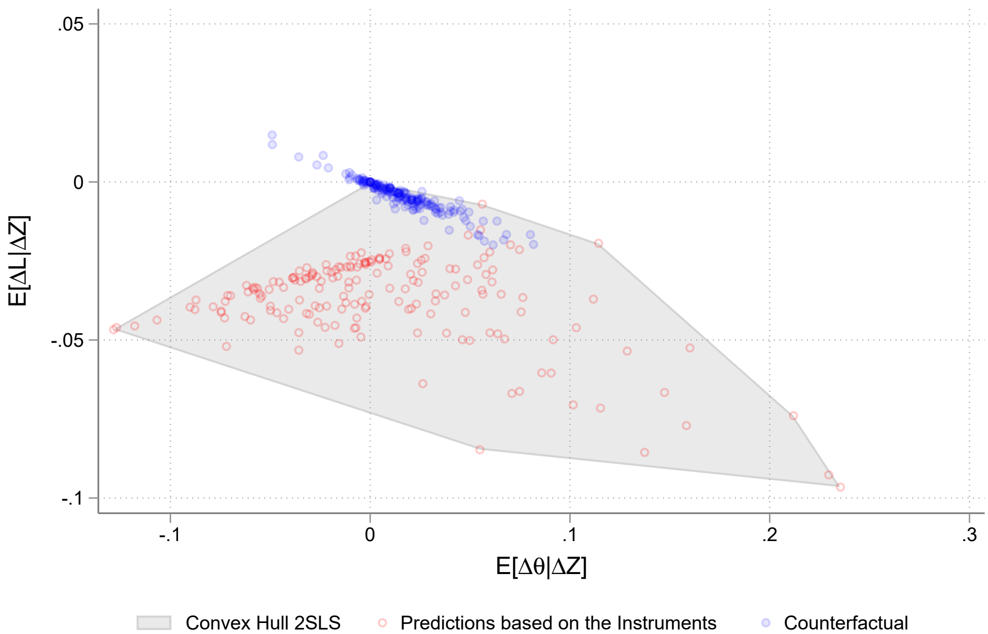
We then compute the estimated elasticities by averaging out across simulations.

Figure A.1: Physicians' Wages in Public Hospitals



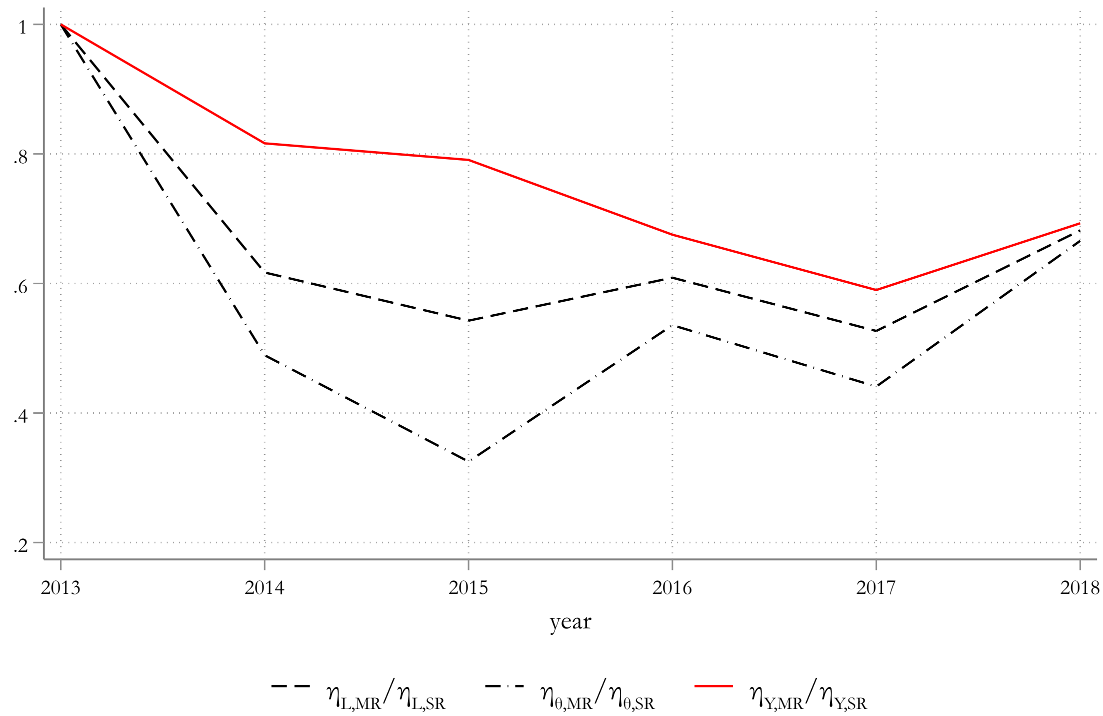
Notes: This figure illustrates the growth of physicians' wages in public hospitals. Wages are adjusted using public sector-wide adjustment and indexed to 2011, which is set to 1. Data are sourced from SIRH (2019).

Figure A.2: Convex-Hull



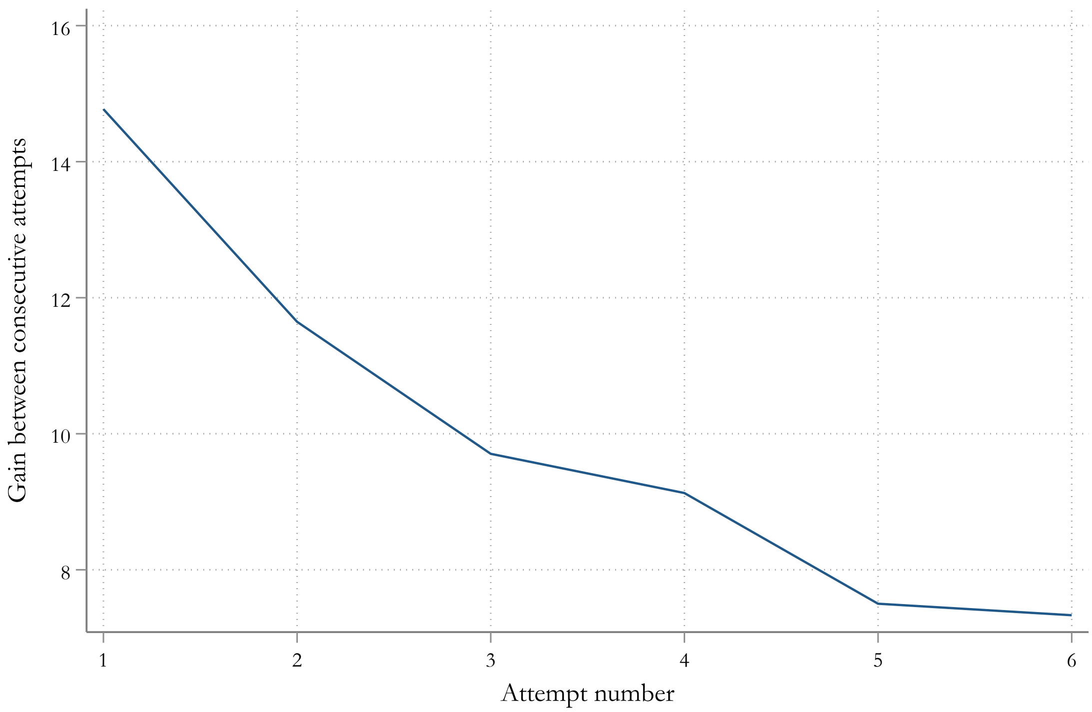
Notes: This figure plots the differences between negative of the instrument-predicted quantity and quality of physicians (in red) and the simulated changes in quantity and quality in our counterfactual (in blue). Predictions are constructed using the first-stage coefficients and average instrument values by hospital. Simulated changes are obtained by rescaling the input elasticities by a factor of 5/51. The convex hull of the instrument's predictions is shown in gray.

Figure A.3: Ratio of Simulated Short- and Medium-run Elasticities



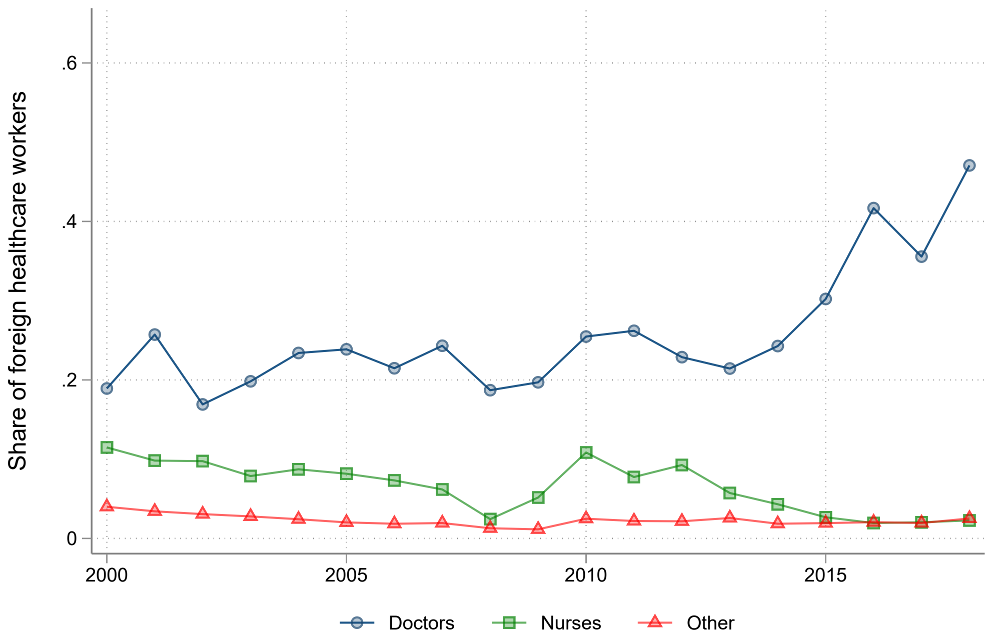
Notes: The figure shows the ratio between the short- and medium-run elasticities of labor, quality, and mortality.

Figure A.4: Score Gains over Attempts

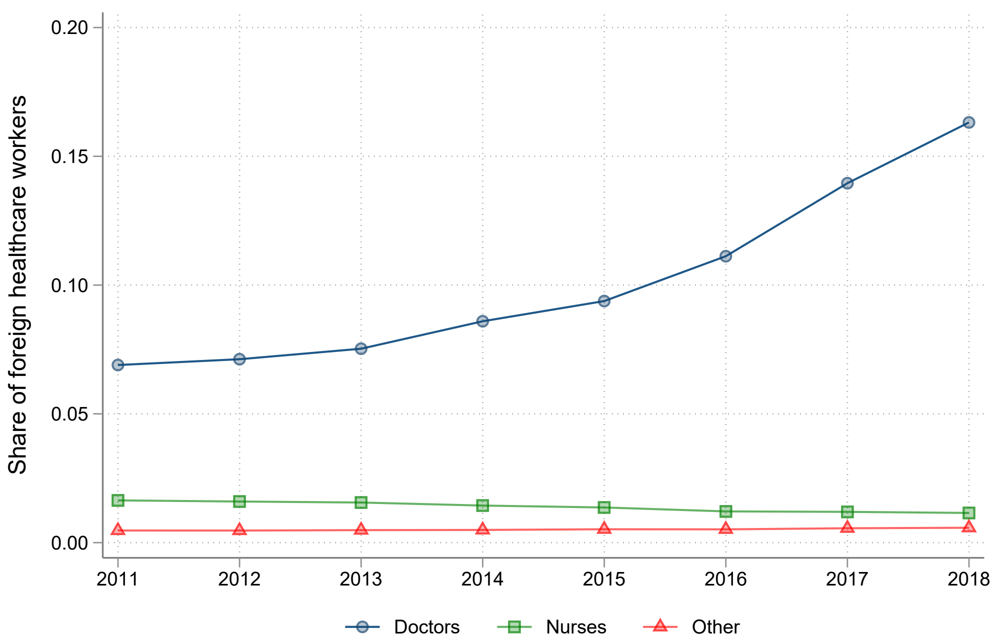


Notes: This figure presents the score gains after consecutive test-taking attempts by the same individual. It shows that on average, score gains diminish with each additional attempt. Data are from ASOFAMECH (2019) and the figure includes the universe of test-takers between 2009 and 2018.

Figure A.5: Migration Wave: Other Labor Inputs



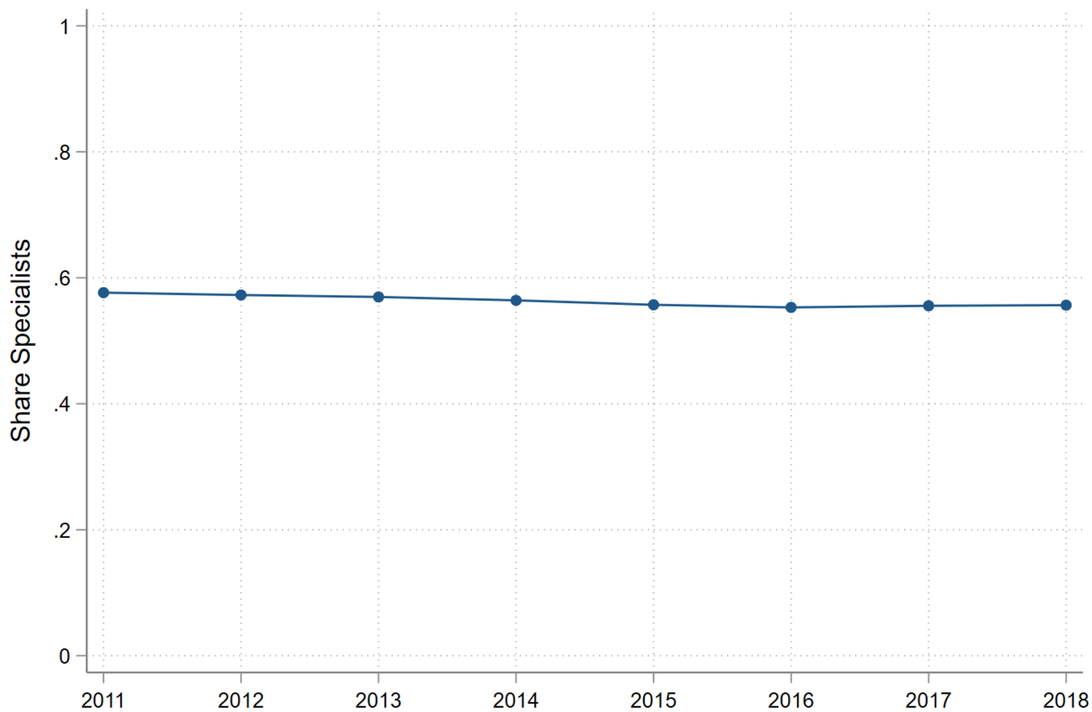
A. Newly Registered Providers



B. Stock in Public Hospitals

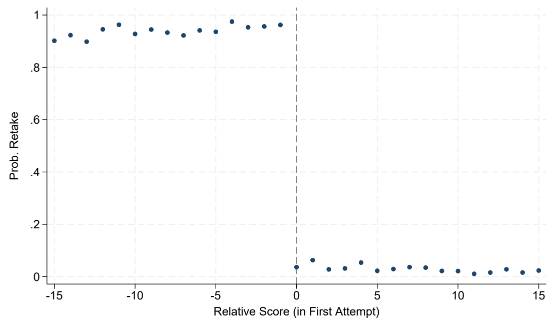
Notes: This figure shows the evolution of the share of foreign-trained healthcare providers in Chile. Panel A examines trends from 2000 to 2018 using data from the National Registry of Healthcare Providers (RNPI, 2024). Panel B focuses on healthcare providers working in public hospitals from 2011 to 2018, based on data from SIRH (2019).

Figure A.6: Specialist Physicians in Public Hospitals Over Time

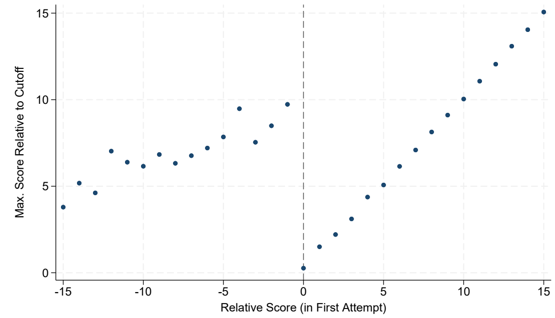


Notes: This figure shows the share of hours worked by specialist physicians relative to all physicians in public hospitals over time. Data are from (SIRH, 2019).

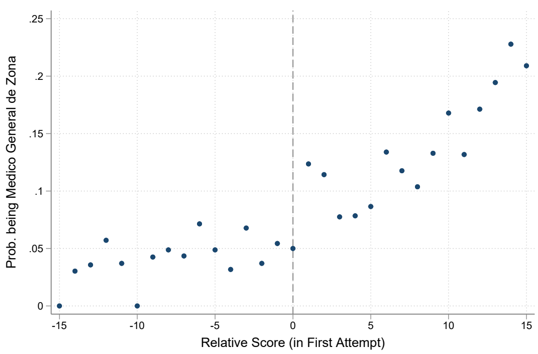
Figure A.7: Testing for Quality Gains



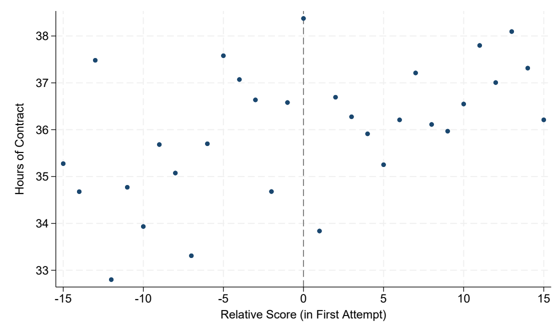
A. Retaking probability



B. Maximum achieved score



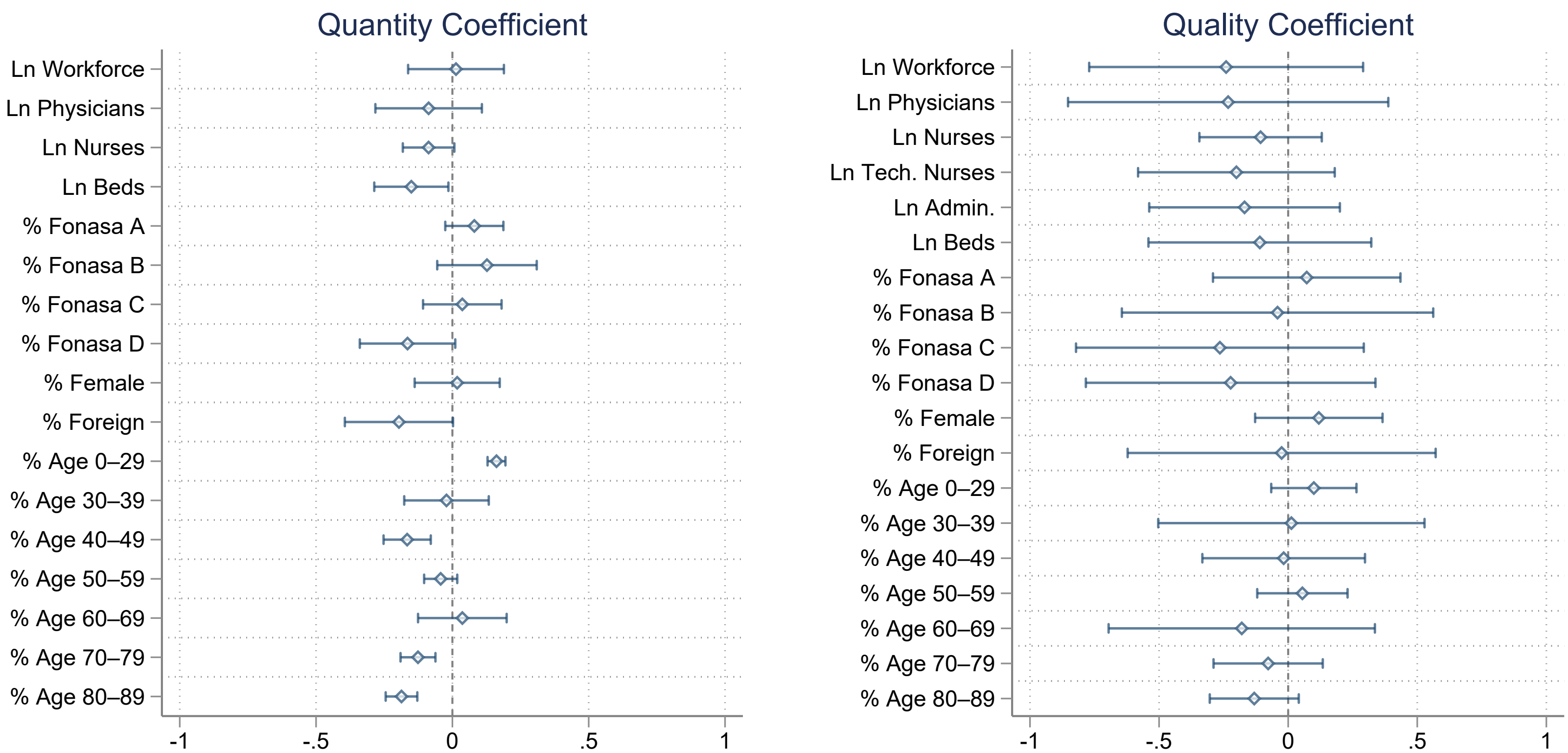
C. Medico General de Zona



D. Hours of contract

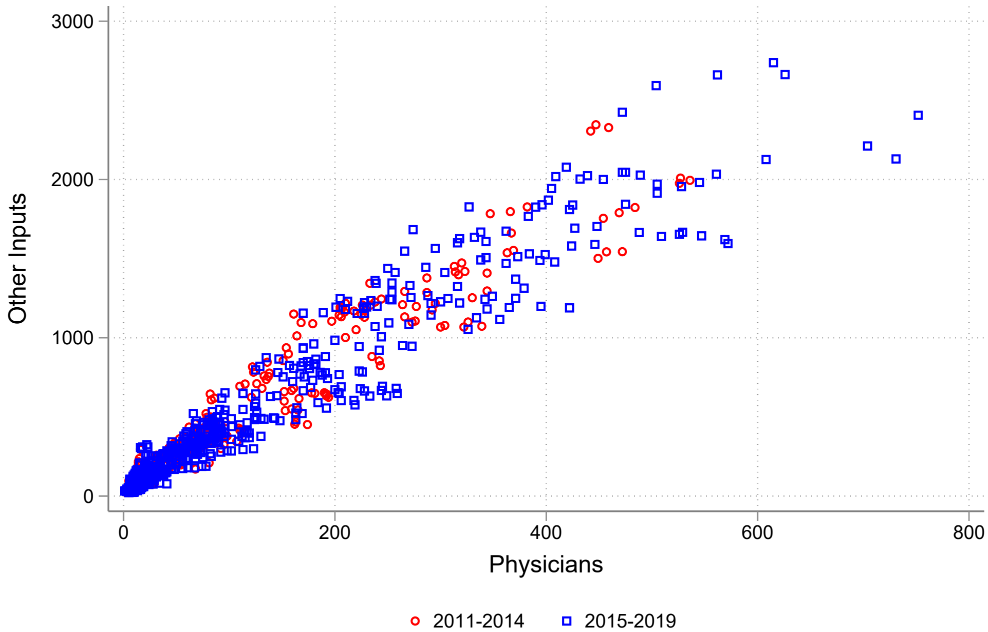
Notes: This figure illustrates the effect of exam retaking on physician-quality improvements. Panel A depicts the probability of individuals retaking the exam based on whether they scored above the passing threshold on their first attempt. Panel B presents the maximum additional points individuals gained across all their retake attempts, relative to the passing threshold. The line on the right closely follows a 45-degree trajectory, consistent with the fact that individuals who pass the exam do not retake it. In contrast, the line on the left indicates that individuals just below the threshold improve their scores by approximately 10 points on average, with smaller improvements observed for those further below the threshold. Panel C examines the probability of becoming *Medico General de Zona* as a proxy for physician quality (see main text for the definition). There is no visible discontinuity at the threshold on their first test-taking attempt, which is consistent with the absence of quality gains from repeated test-taking. Panel D use weekly working hours as an additional quality measure, and similarly, there is no visible effect at the threshold. Data on scores are from ASOFAMECH (2019). Panels A and B use the full sample of test-takers, while Panels C and D incorporate data from SIRH (2019), restricting the sample to physicians observed working in public hospitals.

Figure A.8: Shocks Balance Test

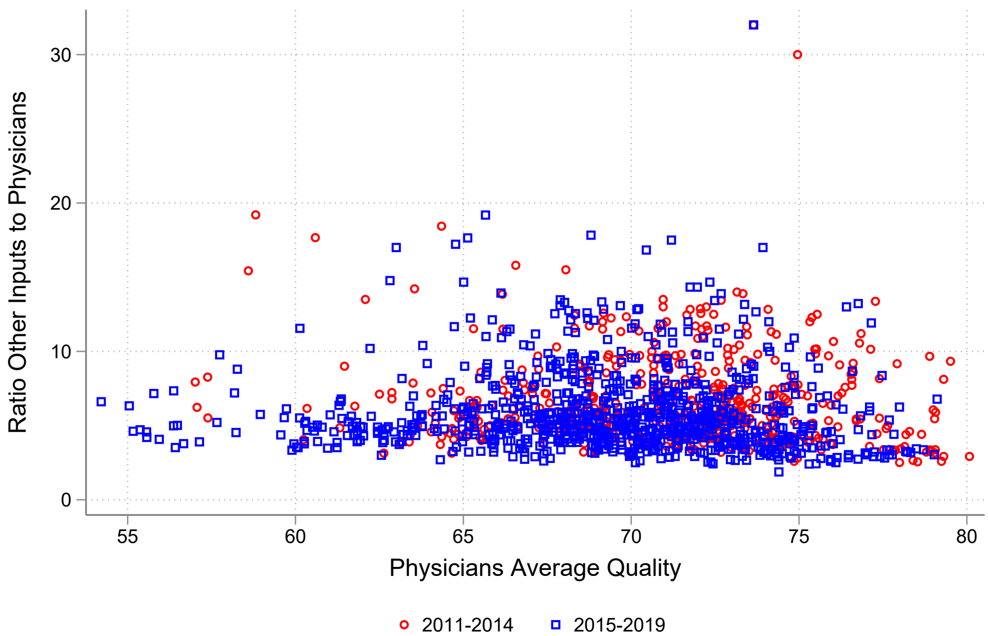


Notes: These figures assess the plausibility of the conditional quasi-random assignment of shocks. Panels A and B show the point estimate and confidence interval obtained from separate regressions of predetermined variables (as of 2012) on quantity and quality shocks, respectively. All variables are standardized to have a mean of zero and a standard deviation of one.

Figure A.9: Other Hospital Healthcare Workers



A. Ratio of other inputs by physicians



B. Mix by physician quality

Notes: This figure examines the relationship between physicians and other full-time equivalent (FTE) healthcare workers in public hospitals. Panel A presents a scatter plot of the number of physicians and other healthcare workers at the hospital level, with each observation representing a hospital-year. Panel B plots the ratio of other healthcare workers to physicians against average physician quality. Pre-migration wave observations are represented by red circle markers, and post-migration wave observations by blue square markers in both panels.

Table A.1: Descriptive Statistics

Panel A: Health Outcomes

	Mean	Std. Dev.	Median (p50)	# of Obs.
	(1)	(2)	(3)	(4)
Patient Characteristics:				
% Female	0.57	0.07	0.58	1,404
% Foreign	0.01	0.03	0.00	1,404
% Age < 29	0.30	0.15	0.31	1,404
% Age $\in$ (30,29)	0.10	0.04	0.10	1,404
% Age $\in$ (40,49)	0.09	0.02	0.09	1,404
% Age $\in$ (50,59)	0.11	0.03	0.11	1,404
% Age $\in$ (60,69)	0.13	0.04	0.12	1,404
% Age $\in$ (70,79)	0.14	0.05	0.13	1,404
% Age $\in$ (80,89)	0.11	0.06	0.10	1,404
% Age > 89	0.03	0.02	0.02	1,404
% Public Insurance	0.97	0.04	0.98	1,404
Hospital Characteristics:				
In-hospital Death Rate	3.30	1.80	2.93	1,404
28-day Death Rate	5.11	2.67	4.53	1,404
Patients (# Admissions)	5,668	7,685	1,931	1,404
Service Rate (# Admissions/Beneficiaries)	0.01	0.02	0.00	1,404
Average Length of Stay	3.51	2.13	3.00	1,404
Readmission Rate	3.46	1.39	3.24	1,404
Number of Physicians	77.55	119.66	20.00	1,404
Average Scores	71.38	5.89	71.76	1,404

Panel B: Test Scores

Year	Number Tests Takers	Average score	% Approved (score ≥ 51)	Average score if score ≥ 51	# Tests $\in [51-56)$
2009	1,389	71.8	92	74.3	70
2010	1,535	65.1	80	72.1	88
2011	1,748	66.6	81	73.3	105
2013	2,003	56.1	66	67.5	129
2014	2,557	55.8	65	67.5	203
2015	3,641	54.7	60	66.5	323
2016	4,999	53.0	54	66.9	466
2017	6,014	52.1	55	64.9	662
2018	7,121	53.9	58	65.0	856

Notes: This table presents descriptive statistics for the data used in the primary analysis. Panel A summarizes patient and hospital characteristics for all public hospitals included in the analysis. Data are derived from individual-level inpatient records reported by DEIS (2019) and restricted administrative records on public hospital employees from SIRH (2019). Panel B provides statistics on EUNACOM scores from 2009 to 2018. For cases in which the exam was taken twice in the same year, all data are pooled. The dataset includes all records of test takers and is from ASOFAMECH (2019).

Table A.2: Shock Summary Statistics

	Quantity Shock		Quality Shock	
	(1)	(2)	(3)	(4)
Mean	-0.2	0.0	-0.4	0.0
Standard deviation	0.8	0.8	0.7	0.4
Interquartile range	0.2	0.1	1.3	0.4
Residualizing on period FE	No	Yes	No	Yes
Effective sample size ($1/HHI$ of s_{nt} weights)	28.86	28.86	28.86	28.86
Largest s_{nt} weight	0.06	0.06	0.06	0.06
No. of regions-period shocks	126	126	126	126
No. of regions	16	16	16	16

Notes: This table summarizes the distribution of migration shocks across regions of origin and periods. Quantity shocks are measured as the percentage change in the number of physicians from each region of origin who clear the cutoff of the licensing exam. Similarly, quality shocks are measured as the change in the average quality of eligible test-takers from different regions of origin. All statistics are weighted by region-of-origin exposure shares. Columns (1) and (2) consider the quantity shocks, and Columns (3) and (4) the quality shocks. Columns (2) and (4) residualize migration shocks on period indicators. As in Borusyak et al. (2022), we also report the effective sample size (the inverse re-normalized Herfindahl index of the weights) as well as the largest shares.

Table A.3: Heterogeneous Impact of Physicians' Quantity and Quality

	Ln in-hospital death rate	Ln 28-days death rate	Ln readmission rate	Ln readmission or death rate
	(1)	(2)	(3)	(4)
Panel A: Top 20% inpatients				
Ln Physicians ($\hat{\alpha}_L^{\text{mortality}}$)	-0.51	-0.45	-0.28	-0.69
	[0.000]	[0.001]	[0.193]	[0.000]
Std. Physicians' Quality ($\hat{\alpha}_L^{\text{mortality}}$)	-0.24	-0.28	-0.12	-0.27
	[0.000]	[0.000]	[0.094]	[0.000]
Observations	1,389	1,389	1,366	1,389
Mean Dep. Var.	9.613	13.58	3.802	14.63
F-stat (First-stage)	24.80	24.80	25.02	24.80
Panel B: Bottom 80% inpatients				
Ln Physicians ($\hat{\alpha}_L^{\text{mortality}}$)	0.12	-0.01	-0.38	-0.36
	[0.470]	[0.960]	[0.132]	[0.084]
Std. Physicians' Quality ($\hat{\alpha}_L^{\text{mortality}}$)	-0.08	-0.06	-0.32	-0.31
	[0.435]	[0.749]	[0.007]	[0.002]
Observations	1,362	1,362	1,357	1,362
Mean Dep. Var.	1.172	2.221	3.431	4.627
F-stat (First-stage)	21.55	21.55	21.52	21.55

Notes: This table presents the impact of the quantity and quality of physicians on public hospital performance. Panel A focuses on patients in the top 20% of predicted mortality based on their diagnosis, and Panel B on the remainder of patients. Estimates are from the two-stage least squares estimation of Equation (5). Column (4) includes counts of patient readmission or death. The mean dependent variable is presented in levels instead of logs. We present exposure-robust p-values from models clustered at the region-of-origin (i.e., the shock) level in brackets (Adao et al., 2019; Borusyak et al., 2022).

Table A.4: Impact of Physicians' Quantity and Quality: Alternative Quality Index

Panel A: Access

	Ln Service Rate	Ln Inpatient Surgeries	Ln Exits from Waiting List	
			Surgical	Medical
	(1)	(2)	(3)	(4)
Ln Physicians ($\hat{\alpha}_L$)	0.96	4.68	3.82	2.96
	(0.46)	(2.55)	(1.41)	(1.29)
% Low Quality Physicians ($\hat{\alpha}_\theta$)	0.07	-1.00	0.25	0.03
	(0.14)	(0.80)	(0.81)	(0.89)
Observations	1,403	758	744	953
Mean Dep. Var.	1.40	3,756	1,533	8,376
F-stat (First-stage)	18.22	10.64	9.29	10.91
Anderson-Rubin (χ^2)	0.05	0.07	0.00	0.00

Panel B: Quality

	Mortality			Complications
	Ln death rate In-Hospital	Pred. death rate In-Hospital	Ln death rate 28-Days	Ln readmission rate
	(1)	(2)	(3)	(4)
Ln Physicians ($\hat{\alpha}_L$)	-0.63	0.15	-0.54	-0.31
	(0.37)	(0.10)	(0.38)	(0.29)
% Low Quality Physicians ($\hat{\alpha}_\theta$)	0.47	0.01	0.45	0.38
	(0.19)	(0.07)	(0.17)	(0.20)
Observations	1,403	1,403	1,403	1,403
Mean Dep. Var.	3.27	3.47	5.05	3.28
F-stat (First-stage)	16.18	17.06	16.18	16.18
Anderson-Rubin (χ^2) p-value	0.01	0.16	0.00	0.11

Notes: This table presents the impact of the quantity and quality of physicians on public hospital performance. Panel A focuses on utilization, which we proxy through the service rate, inpatient surgeries, and exits from the waiting list. Panel B focuses on patients' mortality and complications. Estimates are from the two-stage least squares estimation of Equation (5), where we use the share of physicians in each hospital and time with quality below the median of the entire distribution as the quality index. For this exercise, we only consider physicians with EUNACOM scores. The mean dependent variable is presented in levels instead of logs. Robust standard errors in parentheses.

Table A.5: Correlates of Elasticities

	Labor elasticity $\eta_{\underline{s}}^{L_{jt}}$				Quality Semi-elasticity $\eta_{\underline{s}}^{\bar{\theta}_{jt}}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
High (phys/pat)$_{j,t-1}$	0.008			0.009	-0.012			-0.023
	(0.004)			(0.004)	(0.015)			(0.016)
High average score$_{j,t-1}$		0.023		0.021		-0.024		-0.019
		(0.004)		(0.004)		(0.016)		(0.016)
North$_j$			-0.013	-0.014			0.057	0.060
			(0.005)	(0.005)			(0.017)	(0.018)
Mean Dep. Var.	-0.061	-0.061	-0.061	-0.061	0.194	0.194	0.194	0.194
Observations	1086	1086	1086	1086	1086	1086	1086	1086

Notes: The table shows OLS regression coefficients that relate the estimated quantity and quality elasticities to hospital observables. High (phys/pat)$_{j,t-1}$ is an indicator variable for hospitals above the median in the distribution of physicians per patient in the previous period. High average score$_{j,t-1}$ is an indicator variable for hospitals above the median in the distribution of average scores in the previous period. North$_j$ is an indicator variable for hospitals above the median in the distribution of latitudes. All specifications include year fixed effects.