Organizations and the Production Side of Gender Inequality: Evidence from Surgeons*

Danyelle Branco Bladimir Carrillo Gustavo Cordeiro Laisa Rachter

October 24, 2025

Abstract

This paper provides the first systematic evidence that organizations contribute to gender inequality by designing or sustaining work environments with unequal opportunities to perform productive tasks. We explore this question in the context of surgical specialists, a setting where task output is readily measurable, determines career success, and is comparable across firms. Using rich administrative data from Brazil, we document a gender gap in surgical output of about 27 percent that emerges early in careers, cannot be explained by surgeons' background characteristics or work schedules, and accounts for most of the gap in payments directly tied to medical procedures. Exploiting variation generated by surgeons moving across hospitals in a triple-difference design, we show that workplace environments causally account for a large share of the gap. Estimates of hospital-specific work environment effects indicate that more male-centric hospitals retain fewer women and exhibit lower scores on other revealed-preference measures of firm value, suggesting that women are more dissatisfied at these organizations. Female managers mitigate these disparities, but only when baseline gaps are especially pronounced. These findings recast part of gender inequality as a problem of production, not only compensation, and highlight the importance of organizational design for inequality in high-skill labor markets.

JEL codes: J16, J44, I18, L20

Keywords: Gender inequality, workplace environment, tasks, career outcomes, healthcare

*Contact information: Branco: Sao Paulo School of Economics-FGV, Sao Paulo, SP, Brazil, 01332-000 (e-mail: danyelle.branco@fgv.br). Carrillo: Sao Paulo School of Economics-FGV, Sao Paulo, SP, Brazil, 01332-000 (e-mail: bladimir.carrillo@fgv.br). Cordeiro: Instituto de Estudos para Politicas de Saude, SP, Brazil 01332-000 (e-mail: gustavo.cordeiro@ieps.org.br). Rachter: Inter-American Development Bank, Brasilia, Brazil, 70800-400 (e-mail: laisarachter@gmail.com). Financial support for this research from the Brazilian National Council for Scientific and Technological Development (CNPq —# 04/2021) is gratefully acknowledged. We thank David Chan, Carlos Charris, Fernanda Estavan, Luiz Fontes, Philip Oreopoulos, Fabio Miessi, Pablo Muñoz, Esteban Rossi-Hansberg, and David Seim as well as participants at various seminar participants for helpful conversations and comments on this project. We thank Thiago Costa for excellent research assistance. The views expressed in this article are those of the authors and do not necessarily represent those of the Inter-American Development Bank. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

1 Introduction

Despite decades of progress in equal pay and anti-discrimination policies, women still earn substantially less than men.[1] Much of the scholarly and policy focus on the causes of this persistent disparity has rightly been on women's time away from work, often due to the gender division of domestic labor (Bertrand et al., 2010; Antecol et al., 2018; Kleven et al., 2019). Yet an important but underexplored part of the gap may stem from how tasks are allocated and executed in the workplace. Across a wide spectrum of professions, from financial consultants to physicians, immediate compensation, and promotions are tied to the volume of work delivered. If women are displaced from valuable tasks or complete such tasks less effectively, they may earn less than men even under equal pay rates and time at work.

While variation in output could in principle come from individual factors, such as motivation, preferences, or ability, firms also differ in ways that shape both access to valuable tasks and individual performance within them. In many settings, the organization of productive work is mediated by competition, managerial discretion, and informal networks. When these channels operate within cultures and norms that reflect or reinforce male-dominated social structures, they could activate heuristics, in-group favoritism, and implicit biases that favor men over women. Existing research suggests that such biases exist and may be important. Experimental evidence and field studies indicate that women are underrepresented in influential networks within organizations (Cullen and Perez-Truglia, 2023), face evaluation biases (Biernat and Fuegen, 2001; Wynn and Correll, 2018), engage less in competitive behaviors (Niederle and Vesterlund, 2007; Flory et al., 2015), are subject to negative stereotypes about their competence (Reuben et al., 2014), and are disproportionately asked to complete non-promotable work that drains time away from valuable tasks (Babcock et al., 2017). What is missing is a unified framework quantifying the extent to which firms generate or perpetuate work environments that shape gender inequality in task output and associated financial rewards.

Understanding the relative role of organizations versus individuals in driving gender gaps in task output has first-order implications for the theory of the firm and policy. If differences in output are largely shaped by organizational environments rather than individual preferences or abilities, then firms are not merely neutral platforms where pre-existing inequalities play out but active producers of them. This perspective contrasts with canonical models of wage-setting in which inequality within firms arises solely from gender differences in the willingness or ability to bargain over the division of an existing surplus (Card et al., 2016). If biases in the workplace play a central role, policies that equalize wage rates ex post may do little to reduce inequality without interventions targeting the environments in which work is organized and performed.

This paper provides the first systematic analysis of the interplay between firms, tasks, and gender inequality by examining the context of surgical specialists. Performing surgeries is a welfare-relevant task that determines immediate income, long-term career success, and unlike many other high-skilled tasks, it is relatively homogeneous and comparable across firms. As a historically male-dominated field, it is marked by profound inequalities, and anecdotal accounts describe a hostile environment for female surgeons, characterized by

[1] From a historical perspective, the convergence in male and female earnings since the 1970s has been substantial, yet progress has stalled in recent decades. In OECD countries, the median gender wage gap narrowed from about 30 percent in 1975 to 13.9 percent in 2016, but showed little change thereafter, remaining almost flat by 2023 (OECD, 2023). Distributional patterns further reveal that disparities are especially pronounced among higher-skilled workers, a phenomenon commonly referred to as the "glass ceiling." In fact, convergence has been particularly slow for individuals at the top of the income distribution (Goldin, 2014).

limited support, stereotyping of surgical capabilities, and instances of harassment (Lim et al., 2021). A recent survey of the Fellows of the American College of Surgeons ranked the relegation of female surgeons to time-consuming "secretarial" tasks, as opposed to leadership surgical roles, among the top five barriers to career advancement (Zogg et al., 2023). Despite this rich qualitative evidence on gender biases in the surgical workplace, there is limited empirical evidence quantifying their prevalence and impact, making it difficult to assess their general importance. Using unique administrative registers from the entire public health system in Brazil, we track surgical specialists across their careers to quantify the role of individuals versus organizations in driving gender inequality in surgical output. Brazil has a universal health system that constitute the primary source of care in the country. A key institutional feature of this setting is that patients have essentially no ability to choose their surgeon, even for elective procedures, which largely avoids confounding firm-environment effects with patient-driven demand factors.[2]

The first part of our paper starts by documenting a stark and pervasive disparity in surgical output: female surgeons perform as much as 27 percent fewer surgeries than their male counterparts. This imbalance is remarkably consistent across specialties, from general surgery to gynecology/obstetrics, and emerges immediately upon entry in the surgical career. The magnitude of the gap is more than twice as large for planned than for unplanned procedures (53 versus 23 percent), which is noteworthy given that the former are precisely the instances where hospitals exercise greater discretion in case assignment. The evidence also indicates that the lower surgical volume of female surgeons coincides with their greater involvement in administrative responsibilities and in collaborative roles on surgeries led by others, consistent with a diversion of effort and time away from high-value tasks. These differences in task output are not only quantitatively significant but economically consequential, accounting for approximately three-quarters of the 35-percent gender gap in payments directly tied to medical procedures.

As a basic step toward evaluating whether differences in individual versus environmental factors explain the gender gap in surgical output, we re-estimate the volume gap but condition on a basic set of surgeons' characteristics. Controlling for differences in labor supply, experience, training institution, contractual agreements, and specialty choice reduces the volume gap from 27 to 22 percent, leaving the vast majority unexplained. Moreover, a simple decomposition reveals that almost all of the estimated gap comes from procedures performed within the same surgical sub-area of practice, reinforcing that the observed differences are not driven by the comparison of similar surgeons practicing in different but related fields. We also show that the volume gap cannot be explained by women handling more complex cases that typically require longer operating times and more intensive peri-operative management, since such cases are in fact disproportionately undertaken by male surgeons.

The limited role of observable surgeons characteristics in explaining the volume gap does not automatically imply that the work environment matters, as unobserved surgeon attributes could still explain the residual gap. Even if female and male surgeons do not differ in unobservables, the remaining gap may also be explained by

[2]Brazil provides universal health coverage, with providers funded primarily through public budgets and regulated reimbursements. Surgeries are scheduled according to a centralized allocation system, operated via regulation at municipal or state level, which prioritizes timely access to care based on clinical urgency and available capacity. Typically, patients do not know who their surgeon will be until the day of the operation, and declining the assigned physician is not a realistic option. Allowing patients to reject an assigned surgeon could substantially delay treatment or require the patient to repeat preoperative exams, which themselves are often subject to long waiting lists. For emergency cases, the scope for choice is naturally even more constrained, as immediate medical attention takes precedence over patient preferences. We provide a more detailed discussion of these institutional features in the Background section.

differences in the types of hospitals where they practice, such as those with higher or lower overall surgical demand.[3] In the second part of our paper, we turn to more direct tests of the work environment effect. We first establish that the volume gap varies substantially across hospitals, with an interquartile range exceeding more than twice the median gender gap. We then show that the gender gap among specialists entering practice within the same hospital can be predicted from the specific tendencies observed in that hospital. Hospitals with larger historical gender gaps produce larger gender gaps among their newly affiliated specialists. This relationship is both precise and sizeable, with differences between the 25 and 75th percentiles of the predicted gap implying differences of about 40 percent in the observed gaps.

This pattern of transmission could reflect that the work environment differs across hospitals and that these differences are carried forward to successive cohorts of surgeons. However, it may also reflect sorting of different types of surgeons into hospitals. To disentangle environmental effects from selection, we leverage variation across individual surgical procedures within hospitals, and show that the cross-cohort persistence of gender gaps holds even within the most granular classification of surgical procedures. For example, the gap in knee replacements strongly predicts the gender gap among newcomers in that exact same procedure, but has limited predictive power for the gap in hip replacements or other clinically similar surgical interventions. Assuming that any factor that generates sorting of surgeons into hospitals does not vary across clinically similar procedures within hospitals, the sharp correspondence between the predicted and observed gap at the procedure level implies a causal effect of the work environment.[4] Furthermore, the procedure-specific nature of this picture points to localized organizational mechanisms of transmission, such as institutional conventions about which surgeons are entrusted with particular procedures, specific dynamics across teams, and internal incentives to perform certain surgeries.

We next implement an alternative design that directly exploits variation generated by surgeons transitioning across different hospitals.[5] The basic idea of this approach is to examine how the volume gap between female and male surgeons change as they move between hospitals with varying gender gaps. As an illustration, consider a set of specialists moving from a high-gap hospital to another with a low gender gap in surgical output. If the difference in gender gaps between these hospitals is entirely driven by the work environment, then women moving to lower gender gap hospitals would experience an increase in operative volume while men making the same move would see a relative decline. Alternatively, if the differences in gender gaps across hospitals reflect that female surgeons in higher-gender gaps have preferences for lower-volume practice

[3] Note that hospital selection is conceptually distinct from the work environment effect. Two hospitals may each have a gender-neutral work environment, but differ in factors that influence overall surgical volumes. Some hospitals may be located in markets with larger demand for surgical interventions or have better infrastructure to sustain higher throughput. Others may face tighter operating-room capacity, longer case times due to teaching responsibilities, more complex patient populations, or weaker referral pipelines from primary care and specialty networks. If women disproportionately practice in hospitals with lower baseline opportunities for surgical work, they may end up with lower surgical output relative to comparable male surgeons, even if every hospital is internally gender-neutral.

[4] The identifying assumption is that sorting of surgeons across hospitals occurs at a broader institutional level, based on factors like location, prestige, infrastructure, or even overall correlates of gender biases, but not on procedure-specific dynamics. In other words, it is implausible that surgeons choose hospitals by anticipating gender differences in opportunities for each individual type of surgery (e.g., knee replacements versus hip replacements), given the large number of procedures and the complexity of internal hospital practices.

[5] Given evidence suggesting that the mechanisms behind the work environment effect vary within hospitals, we view this complementary design that exploits moves across hospitals as capturing the average effect of distinct local environments operating within each institution.

styles, for example, then we should not observe changes in the volume gap between female and male movers after the transition. We develop and estimate an intuitive triple difference specification that transparently summarizes the relative importance of environmental versus individual factors into a single parameter. The key identification condition is that in the absence of a move, the difference in operative volume between male and female surgeons would have evolved similarly across origin-destination pairs with varying gender gaps.

Implementing this mover design within a flexible event-study setup, we find a discrete jump in the volume gap immediately following the move that is proportional to the origin-destination difference in gender gaps. This effect is not preceded by pre-existing differential trends and exhibits no subsequent upward or downward adjustment in the quarters after the move. The precise alignment of this discontinuous jump with the timing of the move makes it difficult to construct stories of selection, such as slow-moving shifts in reputation, skill acquisition, area of specialization, and practice styles. Our estimates suggests that about 50 percent of the difference in gender gaps between origin and destination is due to workplace factors. When we look at the estimates of the gender-specific changes underlying the triple difference estimates separately, we find that the effects for male and female surgeons are similar in magnitude but opposite in sign. This approximate symmetry suggests that the workplace effects are driven by a reallocation channel in which male gains in surgical volume come at the expense of female peers, rather than a gender-specific change in a surgeon's capacity or efficiency.

We exploit this rich variation to estimate the causal effect of each hospital on surgeons' surgical output separately by gender and perform a series of counterfactual decompositions. With these estimates, we compute the work environment effect specific to each hospital as the difference between the estimated gender-specific hospital effects. Our key decomposition indicates that approximately 58 percent of the *average* gender gap in operative volume is explained by differences in the practice environment faced by female and male surgeons within the same hospital.[6] The hospital-specific estimates also allow us to evaluate whether part of the gap arises because women and men sort into different types of hospitals. We find that if female surgeons were distributed across hospitals as male surgeons are, the average gender gap in surgical output would decline by 23 percent. The remaining 18 percent is attributable to observable and unobservable individual-level factors.

Our estimates of the work environment effects exhibit intuitive correlations with observable features of hospitals. Hospitals with more male-centric environments are characterized by higher surgical volumes and a lean surgical workforce, consistent with potential mechanisms where performance pressures and limited organizational slack amplify gender biases in the workplace. We also find that hospitals with a higher share of female surgeons tend to exhibit less male-biased environments, potentially reflecting a dilution of influential networks dominated by males. These same hospitals retain fewer female surgeons and exhibit lower scores on other revealed-preference measures of firm value based on female surgeons' mobility across hospitals (Sorkin, 2018; Bagger and Lentz, 2019). The opposite pattern holds for male surgeons. These findings provide suggestive evidence of limited compensating differentials, implying that women may be worse off, on net, at these organizations.

[6] While similar in magnitude, it is useful to distinguish between the interpretation of the event-study and that of the decomposition exercise. The event-study identifies the share of the origin–destination difference in gender gaps that can be attributed to the work environment, and the reported parameter is a weighted average across many such hospital-pair comparisons. By contrast, the decomposition that attributes roughly 58 percent of the mean gender gap to workplace effects refers to the share of the overall average gap in surgical output between male and female surgeons that can be explained by differences in their work environments. In short, the former is about pairwise hospital-to-hospital comparisons, while the latter is about the aggregate mean gap.

Having established the causal effect of the work environment and its general correlates, we go one step further and examine the role of managers in creating or mitigating these disparities. Managers have the potential to influence how tasks are allocated, how resources are distributed, and the cultural norms that govern interactions within teams. Gender is a natural dimension to consider. Female managers are likely to be more aware of the subtle barriers, exclusionary networks, and biases that women face, creating a natural motivation to intervene. To explore this question, we exploit differences in the timing of female manager appointments across hospitals and estimate the impacts of these events on the gender gap in operative volume in a staked difference-in-differences. We find that the appointment of female managers is associated with a reallocation of surgical cases toward female surgeons, but this effect occurs only in hospitals with particularly large pre-existing gender gaps. This finding is consistent with the idea that in settings with pronounced gender biases, inequities are more visible and harder to tolerate, providing female managers with stronger incentives to intervene.

To interpret the welfare and policy implications of our findings, it is important to understand whether female and male surgeons differ in the quality and quantity of care delivered. If female surgeons achieve better patient outcomes than their male peers using the same inputs, then gender biases in the workplace would be costly from a social welfare perspective. But if instead male surgeons achieve better outcomes, then there could be a trade-off between equity and efficiency. In the last part of the paper, we explore this question systematically. It is empirically challenging precisely because the assignment of surgeons to cases is in general gender biased. To overcome this identification challenge, we focus on patients admitted through the emergency department and who are on the margin of survival, as identified by a Random Forest machine learning algorithm. Patients with severe conditions require immediate care and case assignment is largely determined by physician availability, limiting the role for gender biases.[7]Using this high-stakes setting, we find no evidence that male surgeons achieve better patient outcomes or use resources more efficiently than their female peers. The estimates are precisely zero across a wide range of quality and quantity metrics, including mortality, readmission, complications, and healthcare spending. This suggests that policies intended to address gender biases in the surgical workplace are unlikely to compromise patient outcomes or impose additional costs on the healthcare system.

These findings speak to the production side of inequality and contribute to a large social science literature on the economics of gender gaps. A well-established body of work has shown that women earn less in part because they spend less time at work, often because of childcare responsibilities or other individual constraints (Kleven et al., 2019; Antecol et al., 2018; Bertrand et al., 2010; Cook et al., 2021). Our findings indicate that an important part of the gap also emerges during the hours they are at work, where biases in the workplace can disproportionately reduce women's realized task output and associated rewards. Our findings also contribute to a substantial literature documenting that gender earnings gaps exist not only across occupations and firms but also within the same organization, especially among high-skill workers (Bayard et al., 2003; Goldin et al., 2017). Card et al. (2016) and a series of subsequent studies by Sorkin (2017), Barth et al. (2021), and Card et al. (2025) demonstrate that a key reason for this within-firm gap is that women receive lower firm-specific pay

[7]Consistent with limited scope for discretion in these emergency situations, female and male surgeons treat patients who are statistically indistinguishable across a comprehensive set of health and demographic characteristics.

premiums than their male colleagues.[8] While this literature establishes that firms matter for gender inequality, this paper extends existing work to investigate why and how they matter. Our findings show that, even under equal wage rates, firms can substantially contribute to gender inequality by creating unequal work environment to transform productive tasks into revenue.

This paper also contributes to our understanding of why gender inequality persists in high-skilled workers. Across a wide range of other professions, including CEOs, academics, lawyers, accounting, pharmacy, and physicians in general, women face large and persistent gaps in earnings and promotion (Flynn et al., 1996; Bertrand and Hallock, 2001; Jagsi et al., 2006; Goldin and Katz, 2016). Despite the clear link between task performance and compensation in high-skilled professions, systematic evidence on its role in driving gender gaps remains scarce, perhaps because of the paucity of data and inherent difficulty of comparing heterogeneous tasks across firms in many sectors. Notable exceptions to this pattern are found in studies of lawyers and physicians. Azmat and Ferrer (2017) show that differences in revenue-generating tasks per hour worked are a major source of income inequality among lawyers, though they do not examine the separate contribution of individuals and firms in driving such differences. In the context of health care, Zeltzer (2020) and Sarsons (2017) document that biases in external referral patterns create gaps in patient volumes and earnings among physicians in the United States. We add to this literature by providing causal evidence that the work environment *within* organizations is a prime driver of gender inequality in the medical practice. Given the similarity of our setting to other high-skill professions characterized by competitive career progression, intensive training, and reliance on performance-based incentives, these findings are likely to generalize beyond healthcare.

Finally, our paper connects to the literature on the role of managers in the outcomes of employees. This line of research shows that female leaders can influence workplace dynamics in ways that particularly benefit women lower in the hierarchy. For instance, Kunze and Miller (2017) show that establishments with a greater representation of female managers tend to display smaller gender gaps in promotion rates in Norway. Similar patterns have been documented in a variety of other settings (Matsa and Miller, 2011; Bertrand et al., 2019; Flabbi et al., 2019; Cullen and Perez-Truglia, 2023). Most of this literature focuses on aggregate outcomes like wages or promotions, but cannot pinpoint *how* female managers make a difference. Our analysis highlights that improved access to valuable tasks constitutes a key channel by which managers can reduce gender inequality.

The rest of the paper is organized as follows. Section 2 provides background information on our setting and describes the data. Section 3 presents key basic facts on the surgical practice motivating our analysis on the role of hospitals in driving gender gaps. Section 4 investigates the role of the work environment. Section 5 explores whether female and male surgeons differ in the quantity and quality of care delivered. Section 6 concludes.

[8] Card et al. (2016) extend the standard Abowd et al. (1999)'s two-way fixed effects model by allowing the firm-specific pay premiums to differ flexibly across genders and mapping those premiums to structural bargaining shares and surplus. Other complementary work includes Bruns (2019), Coudin et al. (2018), Cruz and Rau (2022), Palladino et al. (2021), Lentz et al. (2023), Vattuone (2023), and Morchio and Moser (2024).

2 Setting and Data

2.1 Institutional Setting

Brazil has a universal public healthcare system called the Unified Health System (SUS, its acronym in Portuguese), which provides completely free medical care to the population, including surgical procedures.[9] The SUS provides care through both publicly managed institutions and a network of contracted philanthropic and private providers, with the latter group accounting for approximately 50 percent of all surgeries performed between 2008 and 2024. Payments for medical services are made to all hospitals based on nationally standardized rates, which can serve as the basis for compensating the physician who performed the surgical service. Physicians can be employed under various contractual arrangements, including as locum tenents, civil servants, regular employment contracted staff, or independent providers, often holding multiple jobs simultaneously. The most common types include regular employment contracts and agreements with independent professionals, both of which offer hospitals significant flexibility in managing their workforce.

The pathways through which surgical cases enter the hospital network depend on the urgency of the patient's condition and system-level referral mechanisms. Some surgical cases originate in primary care settings, where patients are initially evaluated, referred to specialists, and subsequently entered into a centralized waiting list system that allocates cases to hospitals based on geographic proximity and service availability. Many other cases, however, enter the system through spontaneous demand, such as emergency room visits or inter-hospital transfers. Within hospitals, the assignment of surgical cases to surgeons reflects a combination of logistical, clinical, and administrative considerations. Patients in critical condition who require immediate intervention are typically treated by the surgical teams on duty, leaving little room for discretion. By contrast, elective cases, as well as certain urgent cases in which patients are stable enough to wait several hours or overnight, afford hospitals greater flexibility in scheduling. In these instances, assignment decisions may be influenced by a mix of formal scheduling systems, informal coordination among providers, and day-to-day operational dynamics, which can result in some surgeons systematically performing more procedures than others.

This allocation process severely limits patients' ability to choose their surgeon even for elective procedures, and there is thus little room for patient preferences to determine volume outcomes of physicians. It is typically the case that the identity of the surgeon is unknown to the patient until the day of the procedure, and patients do not have the option to reject specific surgeons. Refusing to accept the assigned surgeon may lead to significant delays in treatment, as patients would need to wait for another available slot, which may take weeks or months. If a new slot does not become available within a reasonable time, the patient might need to repeat preoperative exams, which themselves have often long and specific waiting lists depending on the case. For patients arriving at the hospital with emergency conditions, the scope for choosing physicians is naturally more limited. In these situations, time is critical and immediate medical attention is essential to preserve life, so waiting to select a specific physician could cause harmful delays.

[9] The SUS was established by the 1988 Federal Constitution, grounded in the principles of universality, integrality, and equity. It consolidated earlier public health initiatives by unifying fragmented programs into a single national system, marking an unprecedented shift in health policy. Before its creation, formal sector workers received care through the Ministry of Health, while the rest of the population relied largely on philanthropic institutions or out-of-pocket spending. Today, approximately 75 percent of Brazilians depend on the SUS for medical care, with the remainder covered by private insurance plans (Paim et al., 2011).

2.2 Data

We use comprehensive administrative records that cover the full universe of hospital admissions within the Brazil's Unified Health System from 2008 through the first quarter of 2024, with a brief overview presented here and a detailed description available in Appendix B. These data originate from multiple registries maintained by the Ministry of Health, Federal Council of Medicine, Brazilian Medical Association, and National Medical Residency Commission. Using unique identifiers, we link patient-level data to medical procedures and the physicians performing them, achieving an approximately 99 percent match rate between patients, doctors, and treatments. The dataset includes rich clinical and demographic information, such as patients' age, sex, race, admission and discharge dates, comorbidities, and diagnostic codes classified under the International Classification of Diseases (ICD-10). For each hospital stay, we observe all procedures performed, the admission type (elective or unplanned), patient mortality, and detailed payment records for each procedure, including reimbursements for professional services and other hospital-related costs under the national fee schedule.

We restrict the sample to surgical hospitalizations, linking each case to the surgeon who performed the operation. Low-complexity skin, subcutaneous tissue, and mucosal procedures (e.g., mole or cyst removal, biopsies) are excluded, as they are typically outpatient and do not require hospital admission, although they occasionally appear in our data as short hospital stays. After these restrictions, the data comprise approximately 66 million surgeries between 2008 and 2024. Alongside the operative records, the dataset contains detailed professional and background information for all physicians, with updates available on a monthly basis. This includes physicians' roles, working hours, facility affiliations, medical school attended, and years of experience. While the administrative registers offer extensive professional information, they do not explicitly record physicians' gender. To address this issue, we supplement our data with records from the *Instituto Brasileiro de Geografia e Estatística* (IBGE), which enable us to infer physicians' gender from their first names, using patterns documented in the 2010 demographic census. We classify a first name as male or female if at least 50 percent of individuals with that name identify with the corresponding gender. Given that first names in Brazil are almost exclusively linked to a single gender, this approach produces a gender assignment that is virtually error-free.[10]

We analyze both surgeons and interventional cardiologists, the latter performing less invasive procedures. Throughout the paper, we collectively refer to them as surgeons. In our baseline analysis, we construct a panel dataset at the surgeon-hospital-quarter level, capturing operative volume, total procedure-generated revenues (including both surgical and non-surgical), specialty, and total weekly hours worked, among other variables. Depending on the analytical objectives, we employ alternative specifications, such as aggregating at the surgeon-quarter level, using annual rather than quarterly frequency, or limiting the panel to surgeons' primary affiliations. To examine correlates of gender gaps, we merge this data with detailed information on the healthcare establishments where surgeons practice, incorporating characteristics such as geographic location, ownership type (for-profit, not-for-profit, and public), complexity levels, and infrastructure indicators including number of beds and availability of specialized equipment.

[10]The IBGE provides the dataset *Nomes do Brasil*, which contains a comprehensive list of first names recorded in the 2010 Brazilian census, along with gender-specific frequencies for each name. Given the strong gender-specific naming conventions in Brazil, the median gender association exceeds 100 percent. As a result, our findings remain virtually unchanged when using a more stringent 100 percent cutoff.

3 Basic Facts on Gender and Surgical Practice

3.1 Gender Gaps in Operative Volume

Panel A of Figure 1 presents the central motivating fact of this paper: a sizeable gender gap in operative volume. On average, female surgeons perform 4.4 fewer surgeries per quarter per hospital than their male counterparts, corresponding to a raw gap of approximately 27 percent. This gap is observed across most of the largest surgical specialties (Panel B) and is notably larger for elective procedures (Panel C), where hospitals have more discretion over case assignments. Moreover, the gap emerges early in surgical careers, widening rapidly within the first two years after specialization before stabilizing (Panel D). This suggests that most of the observed volume gap between female and male surgeons is explained by factors that affect them immediately upon starting their careers.

These differences in operative volume have consequences in terms of service revenue generated. Because the payment rates are nationally standardized per service and do not vary with physician characteristics, any observed differences in total payments reflect variation in the volume and type of services performed. Figure 3 shows that while female surgeons generated about R$ 7920 in professional services, male surgeons earned about R$ 10200, resulting in a gap of 28 percent. This figure is similar to the 35-percent gap observed among physicians in Brazilian labor survey data, which includes fixed basis salaries and earnings from all sources beyond the publicly funded healthcare sector.[11] Comparable patterns are also evident outside the Brazilian context. In the United States, for instance, multiple studies have estimated a gender pay gap among physicians ranging from 20 to 50 percent (Ly et al., 2016; Seabury et al., 2013; Weeks et al., 2009; Zeltzer, 2020).

To assess the role of the operative volume gap in explaining this payment disparity, we non-parametrically reweigh the volume distribution of women to match that of men, following the approach developed by DiNardo et al. (1996). We first sort surgeons into 50 equal-sized bins based on operative volume and compute average surgeon payments by gender within each bin. We then calculate the counterfactual payment levels for female surgeons by taking the average of their observed payments within each volume bin, weighted by the number of male surgeons in each bin. The third bar of Figure 3 displays the resulting counterfactual payment level. Controlling for differences in volume rises the average payment for female surgeons by approximately 25 percent, narrowing the raw gender gap by more than 86 percent. This indicates that differences in operative volume account for most of the gap in service revenue and are thus an important source of income gender inequality among physicians.

3.2 Surgeon Characteristics and the Gender Volume Gap

One potential explanation for the volume gap is that it stems from inherent differences in individual attributes between female and male surgeons that affect their volume outcomes, including fields of practice, labor supply, work schedules, and case complexity. In this section, we take a step toward evaluating this hypothesis.

[11] We use the *Pesquisa Nacional por Amostra de Domicílios Contínua*, a nationwide quarterly labor survey conducted by Brazil's national statistics bureau. The survey has been carried out continuously since the first quarter of 2012. We draw on data from the waves conducted between 2012 and the first quarter of 2024 to compute the gender earnings gap among physicians per quarter per establishment.

Controlling for Observables. Table 1 shows that female and male surgeons differ in several dimensions, including training, experience, and contractual agreements. In Table 2, we explore the role of these variables in driving the volume gap by regressing volume outcomes on a male surgeon indicator and then sequentially including different covariates. Our unit of analysis is a surgeon practicing in a given hospital during a specific quarter.[12] Column 1 replicates the raw gender gap of 4.4 surgeries. Column 2 shows that controlling for surgeon specialty fixed effects increases the estimated volume gap to approximately 5, suggesting that omitting specialty leads to an underestimate of the gap due to the disproportionate concentration of women in high-volume specialties such as gynecology and obstetrics. When we control for year-of-professional-registration×specialty fixed effects in the next column, the estimated coefficient drops to 4.1 (column 3). This suggests that part of the gender gap in operative volume is due to differences in entry cohorts within specialties.

Column 4 controls for a rich set of medical school fixed effects, but there is no material impact on the estimated gap. Conversely, controlling for the type of contractual agreement (e.g., independent contractor versus civil servant) reduces the estimated volume gap by about 10 percent (column 5). This is consistent with a greater representation of women in contractual arrangements that offer fewer opportunities or limited incentives for high-volume output. In column 6, we examine the role of labor supply by controlling for hours worked per week. To account for potential non-linearities in the relationship between hours worked and volume, we break down hours worked into 20 bins of equal width. Including a full set of dummies for these bins has no practical effect on our results. We reache the same conclusion when we look at the operative volume scaled by the number of hours worked. This limited role for labor supply is consistent with the time-intensive nature of the surgical career, which leave little scope for differences in time devoted to clinical practice. Two obvious hypotheses that may explain the gender gap conditional on total hours worked are differences in when surgeons practice and in types of cases undertaken, both of which we explore below.

Days of Practice. If female surgeons tend to operate more frequently on specific days of the week or during off-peak months when surgical volume is typically lower, their operative output would naturally be reduced. To examine this possibility, Appendix Table A.3 estimates the gender gap in operative volume at the surgeon-day level and control for a full set of date (day-of-week×calendar-month×year) fixed effects. With this rich set of controls, we are comparing female and male surgeons practicing on the exact same date. We infer a surgeon's days of practice based on whether they are observed performing at least one surgical or non-surgical procedure on a given hospital-day. As one can infer from the figure, the volume gap remains essentially unchanged when we include the full set of day fixed effects, suggesting that differences in work schedules are unlikely to explain the observed disparity in operative volume.

Sub-Areas of Practice. While the volume gap remains largely unchanged after controlling for specialty fixed effects, it is possible that female and male surgeons practice in different, though related, sub-areas within hospitals that entail varying levels of surgical activities. For example, some female surgeons may concentrate their work in obstetric surgeries, whereas some male surgeons with identical training backgrounds

[12] The regressions are estimated at the surgeon-hospital level for consistency with our main analysis on the role of hospitals in driving the volume gap presented below, but the conclusions are the same when we collapsed the data at the surgeon level. For the interested reader, these results are presented in Appendix Table A.1.

may focus more on breast-related surgeries. These differences may result from organizational mechanisms that potentially limit females' operative output, but they could also reflect personal preferences or other individual-level considerations. We examine this question by applying an Oaxaca-style decomposition of the volume gender gap into within- and between-area components of practice:

$$\bar{y}^M - \bar{y}^F = \underbrace{\sum_{l \in \ell} (\pi_l^M - \pi_l^F)\bar{x}_l^M}_{\text{Between Area}} + \underbrace{\sum_{l \in \ell} \pi_l^F (\bar{x}_l^M - \bar{x}_l^F)}_{\text{Within Area}}$$

In this expression, $\bar{x}_l^g$ is the mean operative volume of surgeons of gender g in area l, and π_l^g is the share of surgeons of gender g performing procedures in that area.[13]The between-area component captures the extent to which the volume gap arises from differences in the sub-areas where female and male surgeons perform surgeries. The within-area component, by contrast, reflects differences in operative volume between male and female surgeons working within the same sub-area. The definition of sub-areas is based on a standardized coding system used by the Brazilian Ministry of Health, which groups related surgical procedures into clinically meaningful categories, allowing for consistent comparisons across distinct yet related areas of surgical practice. Under this classification, obstetric and breast-related procedures, for instance, are categorized into distinct sub-areas.

Our calculations indicate that approximately 106 percent of the volume gender gap is attributable to differences in operative volume within the same sub-area of practice. This suggests that, if anything, segregation contributes to marginally mitigating the gender gap in surgical output.

Case Complexity. A related possibility is that female surgeons may handle more complex patients. Such cases often demand not only extended time in the operating room, but also more intensive pre- and post-operative management, limiting the overall number of surgeries a surgeon can perform.[14] However, the evidence points to the opposite: male surgeons are more likely to manage more complex cases. We reach this conclusion by predicting each patient's in-hospital mortality risk using a Random Forest algorithm (Appendix C) and estimating the effect of the male surgeon indicator on this risk. Appendix Table A.4 shows that the estimated coefficient of interest is positive and statistically significant. This result holds even when we control for surgery fixed effects and thus compare surgeons performing the exact same procedures. Overall, there is no evidence that the volume gap is driven by female surgeons managing more complex cases.

3.3 Where Does the Surgical Work of Female Surgeons Go?

If women surgeons spend as much time in the hospital as their male counterparts, yet perform fewer surgeries, where does their time go? Qualitative studies suggests that women may disproportionately shoulder a set of "invisible" or undervalued responsibilities that absorb time without formal recognition (Lim et al., 2021). Such tasks may include responsibilities tied to institutional compliance, routine documentation, and coordination

[13]Starting from the overall gender volume gap $\bar{y}^M - \bar{y}^F = \sum_{l \in \ell} \pi_l^M \bar{x}_l^M - \sum_{l \in \ell} \pi_l^F \bar{x}_l^F$, we add and subtract the term $\sum_{l \in \ell} \pi_l^F \bar{x}_l^M$ to obtain the decomposition above.

[14]While many aspects of patient care are handled by other members of the medical team, surgeons often participate directly in preoperative planning (e.g., reviewing exams, obtaining informed consent) and postoperative follow-up (e.g., managing complications or adjusting treatment plans). This is especially true in complex cases. These non-operative tasks can be time-consuming and reduce the number of surgeries a surgeon can perform.

of schedules and resources. Beyond administrative duties, women may be asked more often to collaborate in surgical cases led by others, both intraoperatively and perioperatively.[15]

Empirically, assessing the extent to which female surgeons disproportionately spend time on low-value, time-consuming tasks is challenging, precisely because most of these activities are invisible and not formally recorded. In our context, health authorities fortunately record the total hours physicians devote to nonclinical activities, encompassing administrative duties, management, and other functions not directly tied to patient care.[16] We can use this information, along with data on surgeons' participation in cases where they are not the primary operator, to gain insights into the time devoted to lower-value tasks. Consistent with qualitative evidence, Table 3 shows that female surgeons spend more time on nonclinical hours and participate more frequently as collaborators in cases led by others, both in intra-operative and peri-operative phases, compared with their male counterparts. This suggests that the lower surgical output of female surgeons may in part reflect the disproportionate time they spend on administrative and collaborative responsibilities.

3.4 Summary

Overall, the results in this section reveal an economically meaningful gender gap in operative volume, one that is widespread, present upon entry into surgical practice, notably larger in procedures with greater institutional discretion, and that cannot be explained by a basic set of observable surgeon characteristics. This gap in surgical output is accompanied by a disproportionate involvement of female surgeons in nonclinical and lower-value collaborative tasks. One hypothesis is that these differences in the surgical practice reflect differences in the work environment felt by female and male surgeons face. This could range from hospital-wide rules or guidelines to highly localized norms that create uneven opportunities even within the same institution. However, this interpretation should not be taken for granted. If female and male surgeons differ in unobserved factors that are important determinants of volume outcomes, such as preferences for practice styles or unmeasured surgical skill, one may find a gender gap in volume outcomes even in the absence of any environmental effect. In the next part of the paper, we turn to a more direct investigation of the role of environmental factors in shaping the gender gap in operative volume.

4 The Role of the Work Environment

In this part of the paper, we investigate how the practice environment to which surgeons are exposed affects the volume gender gap. We first examine how the patterns observed in a hospital are transmitted to newly minted specialists entering practice there. We then exploit variation generated by surgeons transitioning across different hospitals to infer the causal effect of each hospital and perform counterfactual exercises.

[15] Intraoperative collaborations typically involve assisting during the surgery itself, such as handling instruments, suturing, or supporting specific steps of the procedure under the direction of the lead surgeon. Perioperative collaborations refer to tasks performed before and after the operation, including preoperative planning, preparing the surgical field, and managing postoperative care. While essential to surgical practice, these roles generally position the surgeon as a supporting rather than leading contributor.

[16] In practice, this variable is something of a "black box," since it aggregates diverse responsibilities without specifying their content. It is also likely that some tasks fall outside its scope—for example, corresponding with hospital staff, completing paperwork for surgical scheduling, or organizing departmental meetings. These activities resemble "secretarial" duties that surgeons may be asked to perform but are unlikely to be fully captured in reported nonclinical hours.

4.1 Gender Gap Among Newcomers

Panel A of Figure 2 shows that gender gaps vary widely across hospitals, with the median at 3 surgeries and the interquartile range spanning nearly 7 to 0.05 surgeries.[17] This variation is correlated with general attributes of hospitals, including clinical complexity, location, and ownership (Panel B). One possible explanation for this wide variation is that hospitals differ in the work environment they offer and, as a result, in the task performance observed across surgeons. As a first empirical step, we examine whether newly affiliated surgeons inherit the gender gap historically observed at their hospital.

To fix ideas, let y_{ich} denote the operative volume of newly affiliated surgeon i of gender $g \in \{M, F\}$ entering in hospital h in year c:

$$y^g_{ich} = \gamma^g_h + \alpha^g_i + \tau^g_c + \eta^g_{ich}$$

where τ^g_c denotes a cohort-specific effect, γ^g_h captures gender-specific hospital effects, α^g_i reflects surgeon ability or practice style, and η^g_{ich} is an idiosyncratic error term. Taking the difference between males and females average volumes for each cohort and hospital yields:

$$\bar{y}^M_{ch} - \bar{y}^F_{ch} = \underbrace{(\gamma^M_h - \gamma^F_h)}_{\text{Work Environment}} + \underbrace{(\bar{\alpha}^M_h - \bar{\alpha}^F_h)}_{\text{Sorting}} + \tilde{\tau}_c + \tilde{\eta}_{ch}$$

Of particular interest are the first and second terms on the right-hand side of the equation. The former captures differences in the work environment between male and female surgeons, including institutional practices, norms, and constraints that may systematically advantage one group over the other. Because these environmental forces could in theory differ within hospitals, we view the hospital-level parameter as capturing the average effect of potentially distinct local environments operating within each institution. The third term reflects any gender differences in the average characteristics of the incoming surgeons at that hospital affecting their volume outcomes.

It is useful to rewrite this equation in terms of the predicted gender gap, calculated using information from all preceding cohorts. Letting $\bar{y}^g_{(c-1)h}$ be the average operative volume for gender g among all cohorts entering hospital h prior to cohort c, we arrive at the following estimating equation:

$$\bar{y}^M_{ch} - \bar{y}^F_{ch} = \tilde{\alpha}_c + \beta[\bar{y}^M_{(c-1)h} - \bar{y}^F_{(c-1)h}] + \theta_{ch}$$

where $\theta_{ch} = (\bar{\alpha}^M_{ch} - \bar{\alpha}^F_{ch}) + \tilde{\eta}_{ch}$ and $\beta = (\gamma^M_h - \gamma^F_h)/[\bar{y}^M_{(c-1)h} - \bar{y}^F_{(c-1)h}]$. Conceptually, β captures the proportion of the observed gap that can be attributed to differences in the practice environment. The higher β, the greater the role of environmental factors. We aggregate the data for each cohort at the hospital-by-year-of-entry level to avoid observations with too few female and male surgeons entering a given hospital. Additionally, we exclude cells below and above the 1 and 99 percentiles of the predicted gap distribution to reduce the influence of extreme values.

Panel A of Figure 4 presents a binned scatter plot of the observed gender gap among newly affiliated surgeons against the predicted gender gap. It shows clearly that hospitals with larger predicted gender gaps also tend to exhibit larger gaps among new surgeons. The slope of the fitted line is 0.175 (standard error =

[17] These differences across hospitals are not merely a consequence of variation in overall surgical volume. As illustrated in Appendix Figure A.2, the absolute and relative (log-transformed) measures of the volume gap are strongly positively correlated, with a near-linear relationship.

0.038), indicating that a one-unit increase in the historical gap is associated with a 0.175 unit increase in the gap among newcomers. The magnitude of this relationship is economically and statistically significant. It implies that moving from the 25th to the 75th percentile of the predicted gap distribution is associated with an increase of 2.9 surgeries in the observed gap, corresponding to roughly 40 percent of the mean gap.

A causal interpretation of this finding requires the assumption that any sorting of surgeons into hospitals is based on features that do not differentially affect men and women. For instance, if both male and female surgeons choose hospitals based on general attributes, such as geographic location, available resources, or market size, these factors would cancel out in the difference $\bar{y}^M_{(c-1)h} - \bar{y}^F_{(c-1)h}$ and would not generate bias. In practice, however, it is possible that male and female surgeons choose hospitals based also on gender-specific considerations. If female surgeons systematically avoid hospitals perceived as hostile to women, for example, our estimates may partially capture selection effects. This motivates our more granular analysis below.

Variation Across Surgical Categories. To disentangle selection from environmental effects, we take a more granular look at the data by exploring variation across narrowly defined surgical procedures within hospitals. We rely on the Ministry of Health's classification system, which organizes surgical procedures into roughly 1,000 distinct categories. As an example, consider the broad cardiovascular group. Within this group, coronary artery bypass surgery, transcatheter aortic valve replacement, pacemaker implantation, and valve repair are classified as separated types of interventions. For each surgeon affiliated with a given hospital, we count all surgeries performed across every surgical subcategory in which the hospital had activity, assigning a zero to subcategories where the surgeon performed none.

Using these data, we examine the extent to which gender gaps historically observed within a specific surgical subcategory at a hospital predict the gender gap among newly arriving surgeons in that same subcategory. The core idea behind our approach is that, while surgeons may systematically decide where to practice based on hospital-wide characteristics, it is less plausible that they make such decisions based on hyper-specific procedure-level dynamics. Sorting of this kind would require surgeons to encyclopedic knowledge of how correlates of gender biases vary between, say, laparoscopic versus open cholecystectomies within the same hospital, and to act upon it in a gender-specific manner. Absent such specific sorting, the patterns in the data would allow us to infer the causal effects of local practice environments. The reasons why the practice environment may vary between clinically similar surgical categories may include informal norms regarding which surgeons are trusted with which types of procedures, varying incentives to pursue high-return procedures, and differences in how surgical teams are staffed.

To define how clinically close two surgical categories are, we measure the proportion of surgeons performing procedures in one category who also perform surgical procedures in the other category. This approach is conceptually similar to the method used by Bell et al. (2019) to classify similarity between patents. As an example, this approach yields a ranking where inguinal hernia repair is the closest surgical category to cholecystectomy, followed by appendectomy, and laparotomy. We use this ranking to test the extent to which the predicted gender gap in a given surgical category explains not only the actual gap in that same category but

[17] The original classification system from the Ministry of Health includes roughly 1,600 surgical procedures commonly observed between 2008 and 2024, excluding minor and superficial interventions. We refine this classification by grouping procedures with only minor variations. For instance, we treat coronary artery bypass surgeries with and without cardiopulmonary bypass as a single category. This aggregation yields approximately 1,000 distinct surgical categories.

also the gaps observed in its immediate neighborhood of clinically related categories. Under the assumption that factors determining any gendered selection of surgeons into hospitals do not vary meaningfully across clinically similar procedures within a hospital, any differences in the observed transmission of gender gaps can be interpreted as causal effects of local practice environment.

Panel B of Figure 4 presents estimates of the transmission parameter by distance between surgical categories, where comparisons involving the same category correspond by construction to perfect similarity. We include a comprehensive set of fixed effects for surgical category, hospital, and cohort to control for systematic differences across procedures, hospitals, and time periods.[18] Our estimates show that a one-surgery increase in the predicted gap leads to a 0.8-surgery increase in the actual gap for procedures in exactly the same category. The point estimate falls abruptly to 0.12 for the nearest neighboring category and then declines steeply as we move further from the focal category, eventually reaching values below 0.02. Our estimates are precise enough to confirm that the discontinuous jump between the focal and adjacent category is statistically significant.

Appendix Table A.5 shows that these findings are not driven by hospitals highly specialized in a narrow range of procedures, which may increase the risk of differential sorting. Specifically, the table shows that the results remain almost unchanged when we exclude hospitals in the top quartile of the surgical concentration distribution, as measured by the Herfindahl-Hirschman index. The results are also little affected when we instead exclude the surgical categories with the highest volume either within each hospital or within each hospital and broad surgical group (e.g., cardiovascular surgeries).

If the aforementioned identification assumption holds, the difference in the estimated coefficients between the focal and adjacent surgical categories implies that the practice environment plays a significant role. Our estimates imply that moving from the 25th to 75th percentile of the predicted gap distribution would increase the actual gender gap by approximately 80 percent ($\approx 0.12 \times (0.80 - 0.12)/0.10$). The specificity of this pattern also provides clues about the mechanisms at play. Mechanisms such as hospital-wide policies would be unlikely to generate such localized effects. Rather, the results paint a nuanced picture involving dynamics that occur within hospitals, such as team-specific practices, and intra-specialty cultures.

4.2 Using Movers to Identify Causal Effects

While our within-hospital, across-procedure analysis provides compelling evidence that the practice environment has causal effects, translating these insights into policy may be difficult because targeting such granular dynamics is inherently complex. We therefore pursue a complementary identification strategy that leverages variation across hospitals and allows us to identify which hospitals, on average, systematically amplify or attenuate gender gaps in operative volume. We first exploit this variation to estimate hospitals' average causal effects, and then use these estimates to quantitatively evaluate the sources of the volume gap.

4.2.A Framework and Identification

Framework. Our empirical draws on the well-established mover design implemented in a variety of contexts, including the role of firms in wage setting (Card et al., 2013; Gerard et al., 2021), the estimation of place effects

[18] As in the hospital-level analysis, we exclude hospital-surgery-category-cohort cells falling below or above the 1 and 99 percentiles of the predicted gap distribution to reduce the influence of extreme values.

on health care utilization (Finkelstein et al., 2016; Molitor, 2018), and the contribution of neighborhoods to children's outcomes (Chetty and Hendren, 2018; Chyn and Katz, 2021). In the canonical labor economics setting used to estimate firm effects on earnings, the idea is to track how a worker's earnings evolve after transitioning from one firm to another. In our context, the idea is to track how the volume gap between female and male surgeons change as they simultaneously move between hospitals. Our approach therefore requires observing two types of movers simultaneously: female and male surgeons. Intuitively, if hospitals have causal effects, women moving to hospitals with larger gender gaps would experience a decline in operative volume, while men making the same move would see a relative increase.

Formally, let surgeon i's operative volume at time t be defined as:

$$y_{it}^{g} = \gamma_h^g + \alpha_i^g + \tau_t^g + \eta_{it}^g \tag{1}$$

This model mirrors the decomposition introduced in Section 4.1, where operative volume is expressed as the sum of surgeon-specific, hospital-specific, and time-varying components. When a set of female and male surgeons move from an origin hospital o to a destination hospital d between periods 0 and 1, the conditional change in the gender gap is given by:

$$(\mathbb{E}[y_{id1}^M|\alpha_i^M] - \mathbb{E}[y_{id1}^F|\alpha_i^F]) - (\mathbb{E}[y_{io0}^M|\alpha_i^M] - \mathbb{E}[y_{io0}^F|\alpha_i^F]) = (\gamma_d^M - \gamma_d^F) - (\gamma_o^M - \gamma_o^F)$$

This expression illustrates that the pre-post-move difference in gender gaps depends in part on the differences in the practice environments faced by surgeons when they move between origin and destination hospitals. As such, we can exploit these different sources of variation in a triple difference estimation strategy. The first difference is over time, comparing surgeons' operative volume before and after switching hospitals. The second difference is across hospitals, as the magnitude of the change in the practice environment varies depending on where surgeons are from and moving to. The third difference is across genders, because the practice environment may advantage one gender over the other.

Building on this intuition, we first algebraically rearrange equation (1) to express the model in terms of post-move and gender interactions:

$$y_{it} = Post_t \cdot (\gamma_d^F - \gamma_o^F) + \left[(\gamma_d^M - \gamma_d^F) - (\gamma_o^M - \gamma_o^F)\right] \cdot Post_t \cdot D_m + \tilde{\alpha}_i + \tilde{\tau}_{tg} + \eta_{it}$$

The term $Post_t$ is an indicator equal to one for the post-move period, while D_m is an indicator for male surgeons.[19] We can then conveniently multiply and divide the first two terms on the right-hand of this equation by the origin-destination differences in gender gaps, $\Delta_d^{\text{M-F}} - \Delta_o^{\text{M-F}}$, to obtain:

$$y_{it} = \underbrace{\left[\frac{\gamma_d^F - \gamma_o^F}{\Delta_d^{\text{M-F}} - \Delta_o^{\text{M-F}}}\right]}_{\theta} \cdot \left[\Delta_d^{\text{M-F}} - \Delta_o^{\text{M-F}}\right] \cdot Post_t + \underbrace{\left[\frac{(\gamma_d^M - \gamma_d^F) - (\gamma_o^M - \gamma_o^F)}{\Delta_d^{\text{M-F}} - \Delta_o^{\text{M-F}}}\right]}_{\beta} \cdot \left[\Delta_d^{\text{M-F}} - \Delta_o^{\text{M-F}}\right] \cdot Post_t \cdot D_m$$
$$+ \ \tilde{\alpha}_i + \ \tilde{\tau}_{tg} + \eta_{it}$$

This reformulation makes explicit two types of interaction terms that are central to our empirical strategy. The

[19]The two-way model equations by gender and period can be recovered by applying the identities:

$$\tilde{\alpha}_i = (\alpha_i^F + \gamma_o^F) \cdot (1 - D_m) + (\alpha_i^M + \gamma_o^M) \cdot D_m$$
$$\tilde{\tau}_{tg} = \tau_0^F \cdot (1 - D_m) + \tau_0^M \cdot D_m + (\tau_1^F - \tau_0^F) \cdot Post_t \cdot (1 - D_m) + (\tau_1^M - \tau_0^M) \cdot Post_t \cdot D_m$$

This preserves the gender-specific structure of the original specification.

first term is the interaction between the origin-destination difference in the gender gap and an indicator for the post-move period, $[\Delta_d^{\text{M-F}} - \Delta_o^{\text{M-F}}] \cdot Post_t$. It captures the component of institutional change that affects both genders equally and is proportional to the origin-destination difference in gender gaps. The second term is a triple interaction between the origin-destination difference in the gender gap, the post-move indicator, and a dummy for male surgeons, $[\Delta_d^{\text{M-F}} - \Delta_o^{\text{M-F}}] \cdot Post_t \cdot D_m$. This triple interaction captures relative importance of the practice environment in driving the differences in gender gaps across origin and destination hospitals. It is parsimoniously summarized in the share parameter β. If all of the difference in the observed gender gaps between origin and destination hospitals reflects sorting of different types of surgeons into certain hospitals, estimates of β would tend to 0. Alternatively, if all of the differences in the gender gaps reflects causal effects of the institutional environment, then we would expect β to converge toward 1. It answers the counterfactual question of how much of the volume gap between the two hospitals would be reduced if surgeons were randomly reallocated between them.

The triple-difference design can be usefully interpreted as comparing difference-in-differences estimates obtained separately for male and female surgeons. In this setup, the parameter of interest corresponds to the interaction between the post-move indicator and the difference in gender gaps between origin and destination hospitals. If the hospital environment were gender-neutral, the two estimates would be identical, and their difference would converge to zero. Alternatively, if the practice environment systematically favors men, male movers would experience positive effects from transitioning to more male-advantaging hospitals, whereas female movers would experience negative effects. By differencing these gender-specific estimates, the triple difference estimator nets out any confounding factor common to both genders and isolates the contribution of the practice environment to observed gender gaps.[20]

Identification. We can estimate the triple difference model above using data on all surgeons who switched hospitals across different periods and origin-destination pairs. A complication with this approach is that the timing of the move differs across surgeons. A recent literature has established that conventional estimators based on two-group/two-period before-after comparisons may be biased in the presence of heterogeneous treatment effects (De Chaisemartin and d'Haultfoeuille, 2020; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; Goodman-Bacon, 2021). Our approach to avoid biases from "forbidden" comparisons is a stacked difference-in-differences design (Wing et al., 2024; Butters et al., 2022; Cengiz et al., 2019). We first construct a separate panel dataset for each surgeon that switched hospital in quarter k and then append all these sub-samples into a single file to form the final stacked sample. We estimate the parameters of interest using this stacked sample

[20] Formally, suppose we estimate the following difference-in-differences model separately for each gender:

$$y_{it}^g = \theta^g \cdot \left[\Delta_d^{\text{M-F}} - \Delta_o^{\text{M-F}}\right] \cdot Post_t \; + \; \tilde{\alpha}_i + \; \tilde{\tau}_t + \eta_{it},$$

where θ^g captures the average effect of moving to a hospital with a different gender gap for surgeons of gender g. The triple-difference estimator can then be obtained as the difference between the two coefficients, $\beta \equiv \theta^{\text{M}} - \theta^{\text{F}}$. Under the structure implied by the two-way fixed effects model in equation (1), this parameter simplifies to

$$\beta = \theta^{\text{M}} - \theta^{\text{F}} = \frac{[(\gamma_d^M - \gamma_d^F) - (\gamma_o^M - \gamma_o^F)]}{\Delta_d^{\text{M-F}} - \Delta_o^{\text{M-F}}},$$

which directly maps the difference-in-differences estimates for each gender into the share of the origin-destination difference in gender gaps that can be attributed to the causal effect of the institutional environment.

and control for a full set of event fixed effects interacted with individual and time fixed effects.

Under this stacked approach, the estimating equation takes the form:

$$y_{itk} = \theta \cdot \left[\Delta^{\text{M-F}}_{d(ik)} - \Delta^{\text{M-F}}_{o(ik)}\right] \cdot \mathbb{1}\{t \geq k\} + \beta \cdot \left[\Delta^{\text{M-F}}_{d(ik)} - \Delta^{\text{M-F}}_{o(ik)}\right] \cdot \mathbb{1}\{t \geq k\} \cdot D_m \; + \; \tilde{\alpha}_{ik} + \; \tilde{\tau}_{tgk} + \eta_{itk} \quad (2)$$

In this specification, we replace $Post_t$ with $\mathbb{1}\{\cdot\}$, an indicator equal to one for quarters at or after the time of the switch. Note that the gender gap, $\Delta^{\text{M-F}}$, is indexed by i and k because each surgeon may move at a different time k and between a distinct pair of hospitals. Crucially, the inclusion of event fixed effects interacted with both surgeon and time-gender fixed effects implies that the parameter of interest is identified by the comparison of surgeons who switched hospital in the same quarter but from different origin-destination pairs.

Assumption 1 (Parallel Trends). *If surgeons had not switched hospitals, the difference in operative volume between male and female surgeons would have evolved similarly across origin-destination hospital pairs with different gender gap differentials*

Assumption 1 is essentially a exogenous mobility condition. Note that this assumption allows for selection into hospitals based on time-invariant unobservables, such as surgical ability, preferences for practice styles, and intrinsic motivation. Only time-varying factors would be a concern. Since all individuals in the sample are movers, selection into moving per se is not an issue. Instead, the identifying variation comes from differences in the $type$ of move (i.e., the magnitude of the change in gender gap), not the decision to move. One could imagine scenarios in which hospitals with smaller gender gaps also serve larger patient populations and thus have higher overall surgical volumes, making moves to such hospitals mechanically associated with increases in operative volume for all surgeons. However, such hospital-level factors that affect female and male surgeons equally are captured by the two-way interaction on the right-hand of (2) and thus do not threaten identification. More generally, what would pose a threat to identification is the presence of time-varying shocks that are (i) correlated with the magnitude of the origin-destination gender gap differences, (ii) aligned with the timing of the move, and (iii) affect operative volume in a gender-specific way. This is a relatively narrow condition.

Even under this specially demanding specification, we do not take the identification condition for granted. If women are especially likely to move to lower gender gap hospitals at moments when they are experiencing unobserved increases in operative volume, such as during phases of natural skill consolidation, our estimates could overestimate the importance of environmental factors in driving the gender gap. To transparently assess the plausibility of the common trends assumption, we estimate a flexible version of equation (2) that allows β to vary by quarter relative to the timing of the event. If movers with larger and smaller destination-origin differences in gender gaps exhibit similar trends in operative volume prior to the move, and only diverge afterward, this provides strong evidence that such changes were driven by the workplace rather than by unobservable factors.

Assumption 2 (Separability). *A surgeon's operative volume can be expressed as a function additively separable in surgeon and hospital effects:* $y^g_{it} = \alpha^g_i + \gamma^g_h + \tau^g_t + \eta^g_{it}$

This is an economically appealing assumption because it delivers a tractable framework in which individual and institutional contributions can be cleanly separated and estimated using standard linear methods. However, it is fairly restrictive and may not hold if there are surgeon-hospital match effects. Such match effects may arise if some surgeons thrive only under specific institutional cultures. For example, certain surgeons may perform better in hospitals with highly collaborative decision-making processes, while others excel in environments with strict hierarchies and clear delegation of tasks. Institutional norms around communication, tolerance for risk, or performance monitoring can differentially interact with surgeons' personalities, training backgrounds, or working styles. In terms of identification, what ultimately matters is whether surgeons select into hospitals in ways that are systematically related to this kind of interaction effects. Provided that such selection does not occur, the two-way fixed effect model (1) continues to provide a valid approximation.

As discussed in Card et al. (2013), a major and testable implication of absence of sorting on match effects is that moves between firms with different average outcomes should produce symmetric changes in individual outcomes. In our context, this means that if the two-way fixed effect model (1) provides a valid approximation and there is no systematic selection on interaction effects between surgeon characteristics and hospital effects, then the average change in operative volume for surgeons moving to lower-volume hospitals should closely mirror, in magnitude but opposite in direction, the change experienced by those moving to higher-volume hospitals. To test for symmetry, we classify hospitals into four quartiles according to their average operative volume and track surgeons' volumes before and after moves from the lowest-volume quartile to each quartile and vice versa. We residualize operative volume by removing surgeon-specific and time-specific trends to account for general changes in volume outcomes unrelated to hospital moves. As shown in Appendix Figure A.3, moves across hospital volume quartiles produce roughly equal and opposite changes in operative volumes, lending support to the additive separability assumption in our analysis.

4.2.B Sample and Results

Estimation Sample. The estimation sample consists of surgeons who changed their hospital affiliation between 2008 and 2024. Because some physicians practice in more than one hospital simultaneously, we focus on a surgeon's primary affiliation. We define the primary hospital as the facility where a surgeon allocates the largest share of their total working hours in a given quarter.[21] Our analysis include all movers, including those who moved more than once in the study period.[22]For surgeons with multiple moves, we include each move by stacking the data and treating every surgeon-event as a separate unit of analysis, as in previous studies involving multiple events per unit (e.g., Chetty et al., 2014; Lafortune et al., 2018; Atal et al., 2024).[23,24] We restrict the panel to a symmetric window of 3 quarters before and after the switch to capture the immediate jump resulting from changes in the workplace environment with sufficiently fine temporal precision, akin to a regression discontinuity design.

The final estimation sample includes about 146,000 observations, with 20,970 unique surgeon-event pairs. To construct the origin-destination difference in gender gaps, we use a leave-one-out strategy that omits surgeon i's own observations from both origin and destination observations in the estimation sample. This avoids any

[21]In our data, 61.9 percent of surgeons are observed working in a single hospital within a given quarter, 28.3 percent work in exactly two hospitals, and 9.8 percent are affiliated with more than two.

[22]Among movers, 40.6 percent experienced a single move, 33.4 percent moved twice, and 26 percent moved more than twice.

[23]For instance, a surgeon who moves twice contributes two time series, one for each move.

[24]An alternative approach would be to restrict the sample to surgeons who move only once during the study period, thus avoiding repeated observations of the same individual. While this strategy may simplify the analysis, we follow the more inclusive approach to maximize statistical precision and leverage all available variation in exposure to different hospital environments.

mechanical correlation between volume outcomes and the independent variable of interest. Appendix Figure A.4 plots the distribution of the origin-destination difference in gender gaps separately for female and male surgeons, revealing a great deal of variation. Both distributions are centered near zero and roughly symmetric, indicating that moves from hospitals with high gender gaps to those with low gaps are about as common as moves in the opposite direction. The standard deviation of the distribution is larger for male surgeons (20 versus 17 surgeries), suggesting that men tend to experience slightly greater changes in gender-related institutional environments when they switch hospitals.

Results. Before presenting the triple-difference results, it is instructive to examine the underlying difference-in-differences patterns separately for male and female surgeons. Specifically, we estimate the difference-in-differences version of equation (2), replacing the post-move indicator in the interaction with the origin-destination gender gap difference by a full set of event-time dummies. Figure 5 presents these results. For both genders, the pre-move coefficients hover near zero with no discernible trend. Upon transitioning to hospitals with larger gender gaps, average outcomes begin to diverge. Female surgeons experience a relative decline in operative volume after switching to hospitals with larger gender gaps, while male surgeons exhibit a mirror-image response but in the exact opposite direction. The discrete change in the coefficients occurs immediately at the time of the move and remains stable in the following quarters, a pattern strongly consistent with the common trends assumption. If surgeons were gradually building experience, acquiring new surgical skills, or changing their practice styles in ways that influence operative volume, we would expect a smoother or more progressive trajectory rather than a discrete jump aligned with the timing of the move. For selection to explain our findings, unobserved shocks to volume outcomes would need to occur exactly in the quarter of the move, and scale with the destination-origin difference in average outcome in a gender-specific way.

The same figure overlays results from the triple-difference event-study specification, which captures the evolving gap between male and female point estimates from the underlying difference-in-differences models at each event time. The figure shows that the gender gap between surgeons moving to higher gender gap hospitals widens immediately following the move, in line with the dynamic pattern observed separately for male and female surgeons. The average point estimate from the triple difference model is approximately 0.5, suggesting that half of the cross-hospital variation in gender gaps reflects causal effects of practice environments.

Threats to Validity. We now focus on the parametric, parsimonious model (2) to summarize magnitudes in table format and perform specification checks parsimoniously. The results are reported in Table 4. Column 1 presents the results from the baseline specification (2), which controls for surgeon-event fixed effects as well as time-by-gender-by-event fixed effects. The triple difference coefficient is estimated at 0.49 and is highly statistically significant at the 1 percent level. A possible objection to the interpretation of the results is that male and female surgeons may systematically sort into different types of hospitals, even when moving to institutions with similarly high gender gaps. For example, male surgeons might disproportionately move to large, well-resourced hospitals, such as regional referral centers with higher surgical throughput, while female surgeons may be more likely to move to smaller or more resource-constrained facilities. In this case, the estimated effects could reflect differences in institutional capacity or infrastructure rather than the causal influence of gendered practice environments per se.

Column 2 investigates the relevance of these potential confounding factors. Specifically, we reestimate the baseline specification but add interactions between origin-destination difference in hospital characteristics and a full set of time fixed effects. The hospital characteristics include overall patient volume, ownership (public versus for-profit), complexity level, urban location, and employment structure (average hours worked and the proportion of surgeons under different contract types). If our results were driven by systematic sorting into hospitals with different levels of capacity or institutional structure, the inclusion of these detailed, dynamically interacted controls should reduce the magnitude of the estimated coefficient. In practice, the triple difference coefficient remains remarkably stable at 0.49. Column 3 goes a step further by introducing triple interactions that allow the effects of the origin-destination difference in hospital characteristics to vary flexibly over time and by surgeon gender. Once again, the coefficient of interest remains almost intact.

Column 4 complements this evidence by adding individual-level controls for hours worked and contract type. If male surgeons tend to move to hospitals with more demanding work arrangements, this could explain the gender-specific pattern we find above. Admittedly, these controls may be endogenous, as the practice environment itself could lead male and female surgeons at the same hospital to end up under different work arrangements.[25] Yet, these controls have no material impact on the estimated coefficient and its standard error, reinforcing that the results are unlikely to be explained by gender-specific sorting into hospitals with varying work intensity demands or contractual arrangements. The robustness of our estimates to these controls also suggests that gendered labor supply responses are unlikely to be a relevant channel through which differences in the practice environment impact volume outcomes.

Another possible threat to identification is that either male or female surgeons may be differentially likely to move to between hospitals when entering specialties with different volume profiles. In this case, the observed changes in operative volume could reflect shifts in case mix or underlying demand associated with specialization choices, rather than the causal effect of the practice environment. However, this possibility is made somewhat less plausible given the dynamic pattern documented in Figure 5. To explain the observed effects, specialization-related shifts would have to coincide exactly with the timing of the move and generate an immediate, discontinuous and lasting shift in volume outcomes. This specific dynamic is difficult to reconcile with the typically gradual nature of specialization processes. Moreover, it is unclear why such specialization-related shifts would align with the magnitude of the origin-destination difference in gender gaps *and* do so in a gender-specific manner. In any case, column 5 adds a surgeon's specialty fixed effects. Consistent with the demanding nature of our design, these controls have essentially no impact on our results. The estimated coefficient of interest stands at 0.49 and remains significant at the 1 percent level.

Overall, the findings of this section suggest that the work environment plays an important role in explaining the variation in the gender gap across hospitals. The specific dynamic of these effects around the move provide some insights on possible mechanisms. First, the fact that the effects emerge immediately, rather than gradually over time, rules out slow-moving channels that depend on repeated interactions, such as gradual changes in

[25] For example, surgeons may be steered into different arrangements depending on how hospitals allocate operating-room blocks, schedule on-call duties, or assign administrative and teaching responsibilities. A female surgeon asked to assume extensive educational or committee work might end up on a regular employment contract with protected nonclinical time, whereas a male colleague offered prime operating blocks could hold an independent provider arrangement with higher clinical volume. Such contractual differences are themselves manifestations of a gendered work environment that channels male and female surgeons into distinct patterns of task performance.

peer support or the buildup of administrative barriers. Instead, the timing is consistent with mechanisms that operate immediately upon entry in the hospital, such as differences in how surgical opportunities are distributed or prioritized across surgeons. Second, the approximate symmetry in magnitudes between the gains for male surgeons and the losses for female surgeons implies that these effects are not primarily driven by changes in a surgeon's capacity or efficiency, but rather by a reallocation channel. That is, the male gains in surgical volume at more male-advantaging hospitals come at the expense of female peers, consistent with a zero-sum dynamic shaped by organizational factors.

4.2.C Decomposition

High-versus-Low Gender Gap Hospitals. Having established that hospitals influence the gender gaps in volume outcomes, we now proceed to perform decompositions of the gender gaps in their environmental and sorting components. To do so, we first estimate γ^g and α_i from the two-way fixed effects model (1) separately for female and male surgeons. With these estimates at hand, our first exercise decomposes the difference in the gender gap between two groups of hospitals, R and R', as follows:

$$(\bar{y}^M_{R'} - \bar{y}^F_{R'}) - (\bar{y}^M_R - \bar{y}^F_R) = \Big[\sum_{h\in R'}(\hat{\gamma}^M_h \pi^M_h - \hat{\gamma}^F_h \pi^F_h) - \sum_{h\in R}(\hat{\gamma}^M_h \pi^M_h - \hat{\gamma}^F_h \pi^F_h)\Big] + \big[(\bar{\hat{\alpha}}^M_{R'} - \bar{\hat{\alpha}}^F_{R'}) - (\bar{\hat{\alpha}}^M_R - \bar{\hat{\alpha}}^F_R)\big]$$

The first term on the right-hand is the hospital component, whereas the second term captures gender differences in fixed surgeons' characteristics across both groups of hospitals. The parameter π^g_h represents the share of surgeons of gender g practicing in hospital h. Following the logic of the classic Oaxaca decomposition (Oaxaca, 1973), the hospital component can be broken down into components reflecting differences in the work environment and sorting patterns across hospitals:[26]

$$\sum_{h\in R'}(\hat{\gamma}^M_h \pi^M_h - \hat{\gamma}^F_h \pi^F_h) - \sum_{h\in R}(\hat{\gamma}^M_h \pi^M_h - \hat{\gamma}^F_h \pi^F_h) = \overbrace{\sum_{h\in R}\hat{\gamma}^M_h(\pi^M_h - \pi^F_h) - \sum_{h\in R'}\hat{\gamma}^M_h(\pi^M_h - \pi^F_h)}^{\text{Institutional Sorting}}$$
$$+ \underbrace{\sum_{h\in R}\pi^F_h(\gamma^M_h - \gamma^F_h) - \sum_{h\in R'}\pi^F_h(\gamma^M_h - \gamma^F_h)}_{\text{Work Environment}}$$

In this equation, *institutional sorting* captures the extent to which differences in gender gaps are explained by differences in the distribution of female and male surgeons across lower- versus higher-volume hospitals. It answers the question of how the gender gap would change if male and female surgeons were assigned to hospitals in the same proportions. This channel contrasts with the *work environment* component, which captures variation in gender gaps arising from differences in the conditions or opportunities faced by male and female surgeons within the same hospital. It answers the question of what the gender gap would be if male and female surgeons faced the same hospital environment, conditional on their observed allocation across hospitals.

In estimating the objects of interest from the two-way model (1), we use the full sample of surgeons, including both movers and nonmovers, to increase precision. In the two-way fixed effects framework, hospital

[26] To obtain this expression, we add and subtract $\hat{\gamma}^M_h \pi^F_h$ for both groups of hospitals R and R'.

effects can only be identified up to a normalizing constant within each connected set of hospitals that have movers in common (Abowd et al., 1999).[27] Following standard practice in the literature (e.g., Gerard et al., 2021), we restrict our analysis to the largest connected set within each gender group, which includes 97 percent of surgeon-quarter observations for females and 98 percent for males. We restrict our analysis to hospitals that employ both male and female surgeons, which is necessary to estimate gender-specific volume effects within each hospital. The additive decomposition will yield consistent estimates provided that the exogenous mobility and separability assumptions are satisfied, which is supported by the evidence in the previous subsections.

Table 5 presents the results of the decomposition. Column 1 compares hospitals above and below the median in terms of gender gaps, while columns 2–4 focus on increasingly extreme comparisons, contrasting hospitals in the top quartile, top decile, and top 5 percent of the gender gap distribution to those in the bottom quartile, decile, and 5 percent, respectively. Across all specifications, differences in hospital effects account for the majority of the variation in gender gaps between hospital groups, ranging from 62 to 68 percent of the total gap. Within the hospital component, the practice environment explains nearly all of the variation. In contrast, the institutional sorting component is consistently close to zero, and negative in all cases. These results suggest that gender disparities in surgical volume across hospitals are primarily driven by differences in the environment that male and female surgeons face within the same hospital, rather than by differences in the types of hospitals where they practice.

The prominent role of the practice environment in the decomposition is consistent with the triple-difference estimates presented in the previous subsection. The decomposition implies a somewhat larger magnitude for this channel, likely because the triple-difference approach averages across all movers and thus reflects a broader set of comparisons. This pattern suggests that gender disparities in practice conditions are not confined to the subset of movers, but are also evident in the broader cross-section of hospitals.

Impacts on Mean Gap. The decomposition above is useful because it highlights the mechanisms driving variation in gender gaps across hospitals. However, this analysis is inherently local, as it sheds light on differences between subsets of hospitals, but does not directly speak to the magnitude or composition of the gender gap in the overall sample. To understand the aggregate drivers of the gender gap in surgical volume across all hospitals, we now turn to a decomposition of the $mean$ gap. To explore this question, it is useful to note that the average difference in operative volume between male and female surgeons can be decomposed as follows:

$$\bar{y}^M - \bar{y}^F = \sum_h \hat{\gamma}_h^M(\pi_h^M - \pi_h^F) + \sum_h \pi_h^F(\hat{\gamma}_h^M - \hat{\gamma}_h^F) + \hat{\alpha}^M - \hat{\alpha}^F$$

Analogous to the local decomposition above, the first two terms isolate the roles of institutional sorting and the work environment, now evaluated at the full-sample level. The final term captures residual differences attributable to surgeon-specific characteristics, such as preferences for practice styles, preferences, and surgical productivity.

Unlike the local decomposition, interpreting the mean decomposition poses the additional challenge that

[27] Note that our decomposition is invariant to the normalization. The decomposition involves differences in gender gaps between subsets of hospitals. Because it compares relative differences, any normalization constant added to all hospital effects within each gender group cancels out when taking these differences. This renders the decomposition invariant to the choice of reference hospital.

the hospital effect for any given hospital is only identified relative to a reference hospital or set of hospitals. While this does not affect the sorting component, the size of the practice environment term can be sensitive to how the model is normalized.[28]For example, if the reference set includes hospitals where male surgeons tend to have better volume outcomes, the normalized average hospital effect for men will be lower, making the hospital component appear smaller. Identification therefore requires a set of hospitals where female and male surgeons face similar environmental conditions or opportunities (i.e., $\gamma_h^M \approx \gamma_h^F$).

In the literature decomposing earnings across different groups, researchers typically use the restaurant or service sectors as the reference group, motivated by the economic intuition that their low-surplus nature limits wage-setting power and firm-specific rent extraction (Card et al., 2013; Gerard et al., 2021). In our setting, identifying an economically meaningful reference group is less straightforward because differences in performance or task allocation outcomes are more difficult to benchmark against a clearly neutral baseline. We therefore adopt an approach that combines institutional considerations with empirical patterns observed in the data. In particular, we define the reference group as the set of public hospitals exhibiting the small and statistically insignificant gender gaps in operative volume. The rationale is that, as is often the case in the public sector in general, public organizations tend to operate under more rigid administrative structures and standardized procedures, limiting the scope for bias in the allocation of tasks. The evidence in Figure 2 documenting that the volume gender gap is the lowest in public hospitals is consistent with this reasoning. By further restricting the reference group to public hospitals with negligible observed gender gaps in operative volume, we increase the chances of isolating settings where workplace dynamics are effectively gender-neutral.

Of course, this data-driven approach is not a panacea and the normalization assumption may not hold in practice. Even in settings with standardized procedures and negligible gaps, subtle gender-biased practices may still be at play. However, given the descriptive evidence that the observed volume gap favors male surgeons in most hospitals and the substantial work environment effect suggested by the local decompositions, it is plausible that any violation of this assumption arises because male surgeons face a more favorable environment in the reference group. This will tend to cause the estimated effect of the practice environment to appear smaller than it truly is.

With this caveat in mind, column 5 of Table 5 presents the results. We define public hospitals with negligible gender gaps as those where the estimated gap is less than 0.5 surgeries in absolute value and statistically indistinguishable from zero.[29] As one can infer from the table, over 80 percent of the overall gender gap in operative volumes is attributed to hospital effects, with most of this effect driven by the work

[28]The sorting component remains unaffected by the normalization choice because the distribution shares π's are probability distributions that sum to one across all hospitals. This implies that the difference in group-specific distributions across hospitals always sums to zero: $\sum_h(\pi_h^M - \pi_h^F) = \sum_h \pi_h^M - \sum_h \pi_h^F = 1 - 1 = 0$. Formally, suppose the estimated hospital effects $\tilde{\gamma}_h^M$ are normalized by subtracting a constant μ, so that $\gamma_h^M = \tilde{\gamma}_h^M - \mu$. Then the sorting term is given by $\sum_h \gamma_h^M(\pi_h^M - \pi_h^F)$, which can be rewritten as:

$$\sum_h(\tilde{\gamma}_h^M - \mu)(\pi_h^M - \pi_h^F) = \sum_h \tilde{\gamma}_h^M(\pi_h^M - \pi_h^F) - \mu\sum_h(\pi_h^M - \pi_h^F)$$

Since the second term is zero, the sorting component is invariant to the normalization constant μ.

[29]Appendix Figure A.5 assesses the sensitivity of our estimates of the work environment effect to alternative definitions of the "small" gap threshold. For thresholds below 0.5, the contribution of the work environment fluctuates around our baseline estimate of 58 percent, with no clear upward or downward trend. However, as the threshold increases beyond 0.5, this share declines monotonically and nearly linearly. This pattern is consistent with the idea that more lenient thresholds are more likely to include hospitals with a practice environment biased toward male surgeons, introducing a downward bias in the magnitude of the parameter of interest.

environment (58.1 percent) rather than sorting across hospitals (23.0 percent). While the sign and magnitude of the sorting effect differ from those estimated in the local decompositions, these two sets of estimates are not directly comparable. The consistently small sorting effects in the local comparisons indicate that gender-based selection into hospitals does not vary substantially between low- and high-gender gap hospitals. By contrast, the mean decomposition reflects the aggregate level of sorting in the data, which may still be sizeable even if it does not explain cross-hospital differences in gender gaps.

4.2.D The Work Environment and the Payment Gender Gap

What would gender earnings inequality look like if all hospitals had a gender-neutral practice environment? To approximate this counterfactual scenario, we combine estimates of the relationship between operative volume and overall revenue with the mean decomposition results from the previous subsection. Specifically, we compute the counterfactual gender gap in payments using the following expression:

$$\bar{s}^{Mc} - \bar{s}^{Fc} = \underbrace{(\bar{s}^{M} - \bar{s}^{F})}_{\text{Actual Payment Gap}} - \underbrace{\hat{\beta}_{sy} \sum_{h} \pi^{F} (\hat{\gamma}_h^{M} - \hat{\gamma}_h^{F})}_{\text{Work Environment}}$$

Here, $\bar{s}^{gc}$ and $\bar{s}^{g}$ denote the counterfactual and observed mean revenue for surgeons of gender g, respectively. The summation term captures how much of the observed earnings gap is attributable to differential practice environments, as measured by hospital fixed effects from the previous decomposition. The coefficient $\hat{\beta}_{sy}$ is obtained from a regression of individual surgeon revenue on surgical volume, and reflects how differences in volume translate into payment differences. It implicitly assumes that this relationship does not differ by gender, which we find institutionally reasonable given that payment rates are nationally standardized and do not vary with a surgeon's gender. Empirically, we are unable to reject the null hypothesis of no gender difference in the volume–revenue gradient.

Our calculations indicate that eliminating the practice environment channel would reduce the gender payment gap by approximately 60 percent. These estimates are based on averages from physician-hospital-quarter-level data. Assuming this share remains proportionally similar when using data aggregated at the physician–quarter level, the gender gap in service payments would decline by roughly R$ 2000. For perspective, this decline corresponds to about 15 percent of the overall gender gap in physicians' earnings from all sources, as measured in labor survey data. This estimate should be viewed as conservative, as our data do not cover any payments made outside the publicly funded healthcare system, and there is no basis to expect the influence of the work environment to be smaller in those settings. Even so, a contribution of 15 percent to the overall gap is economically large.

4.2.E The Characteristics of Male-Biased Workplaces

We next turn to a characterization of hospitals with more and less male-biased work environments. Using the estimated gender-specific hospital effects, we compute the work environment effect for each hospital as $\hat{\gamma}_h^{M} - \hat{\gamma}_h^{F}$ and relate it to observable hospital characteristics. We focus primarily on observables that capture differences in resources, workforce composition, and local market conditions. While the results do not necessarily imply causal relationships, they represent a first step toward identifying and understanding the features of hospitals with more male-centric workplaces.

The results are reported in Figure 6. Each row represents a different covariate. In Panel A, we estimate a separate OLS regression for each hospital characteristics. In Panel B, the estimated coefficients are from a multivariate OLS regression, where the covariates were selected using LASSO applied to the full set of covariates (Belloni and Chernozhukov, 2013).[30] To facilitate interpretation, both the outcome and the regressors are standardized to have a mean of zero and a standard deviation of one. We weight the observations by the number of surgeons in each hospital to account for differences in the precision with which the hospital effects are estimated.

Most patterns are intuitive. While the majority of coefficients are statistically significant in the bivariate regressions, only five retain significance in the post-lasso estimates. Among these covariates, the strongest predictors are surgeon workforce size and overall surgical volume. Hospitals that tend to exhibit more male-centric practice environments are typically those with fewer surgeons overall, and higher surgical volumes. This suggests that gender disparities are more pronounced in settings where organizational slack is limited and performance pressures are high. In hospitals with fewer surgeons, the work environment may be more informal, with decisions concentrated in the hands of a few senior individuals, facilitating the persistence of gender-biased allocation practices. Hospitals operating under pressure, in turn, may allocate surgical cases based on perceived efficiency, inadvertently reinforcing stereotypes that portray male surgeons as more capable or productive.

We also find that more male-centric hospitals tend to have a lower share of female surgeons and receive lower scores on a reveled preference measure of firm value for female surgeons. We compute the revealed preference index by combining three measures that capture how surgeons surgeons sort across hospitals: poaching rate, PageRank index, and retention rate. The poaching rate measures how frequently a hospital attracts female surgeons from other hospitals, serving as a proxy for its relative desirability (Bagger and Lentz, 2019). The PageRank index, adapted from network analysis, captures how central a hospital is in the observed mobility network of female surgeons, weighting moves from more desirable hospitals more heavily (Sorkin, 2018).[31] Because these measures are highly correlated, we generate an index by averaging their standardized values.

One of the most salient findings from this analysis is that hospitals perceived as less attractive from female surgeons tend to be the same settings where women face a more unfavorable practice environment in terms of access to surgical opportunities. Of course, this relationship may reflect reverse causality. Those workplaces characterized by exclusionary practices, lack of support, or unfriendly to women in general may be less interested in hiring women. To shed further light on these patterns, Appendix Table A.6 reports results separately for each component of the firm value index. We find that the gender gap in hospital effects is significantly correlated with all components of the index. When we repeat the analysis using the corresponding

[30]The Least Absolute Shrinkage and Selection Operator (LASSO) is a regression-based method that simultaneously estimates coefficients and selects variables. It does so by applying a penalty that shrinks less relevant coefficients toward zero, effectively excluding them from the model. This allows one to handle settings with many potential controls, improving prediction accuracy and mitigating overfitting while retaining a parsimonious set of covariates (Belloni and Chernozhukov, 2013).

[31]These measures have gained considerable popularity in labor economics in recent years. Recent applications analyze the macroeconomic consequences of gender inequality (Morchio and Moser, 2024), study the reallocation effects of minimum wage policies (Dustmann et al., 2022), examine how collective bargaining can create more female-friendly jobs (Corradini et al., 2025), and investigate the role of worker voice in the workplace (Harju et al., 2025).

measures for male surgeons, the estimates are of opposite sign, though only the coefficient on retention is statistically distinguishable from zero.

The lower gender gap in hospital effects observed in hospitals with a higher share of female surgeons may reflect not only their greater attractiveness to women but also the possibility that a stronger female presence enhances women's bargaining power within the organization. When female surgeons represent a larger share of the workforce, they may be better positioned to advocate for more equitable task allocation, challenge biased practices, and influence institutional norms. This potential mechanism is consistent with broader evidence suggesting that underrepresented groups can reduce bias and improve workplace fairness when they reach a critical mass. For example, Corradini et al. (2025) show that union efforts focused on improving conditions for women had a larger impact in workplaces where women were initially a lower share of workers, suggesting that low representation may constrain women's ability to advocate for better conditions absent institutional support.

4.3 Female Managers as Agents of Change?

Given the sizable role of workplace environment in driving gender gaps in surgical tasks, it is natural to ask whether these environments are immutable or can be actively reshaped from within by actors at the top of the organizational hierarchy. A growing literature has documented that female representation in leadership produces positive spillovers on other women's career trajectories in lower ranks (Matsa and Miller, 2011; Kunze and Miller, 2017). An underexplored hypothesis is that female managers may expand women's access to higher-value tasks. This could occur through both direct intervention in task allocation decisions and broader cultural change that challenge entrenched stereotypes responsible for disproportionately restricting women's roles. The descriptive patterns in Section 4.2.E provides suggestive support for this idea by showing that hospitals with greater female representation exhibit less male-biased workplace environments. However, this correlation may reflect reverse causality: hospitals with a practice environment that is already gender-neutral may be more likely to promote women into leadership positions over the long run. In what follows, we take a step further toward assessing the plausibility of a leadership effect by tracking gender-specific changes around the timing of female managerial appointments.

To explore whether hospitals become more gender-neutral following female managerial appointments, we employ an event-study design that compares the gender-specific evolution of surgical activity before and after a woman assumes a managerial role. Since such effects may take time to materialize, we conduct the analysis at an annual frequency. We match each treated hospital to a not-yet treated hospital based on pre-event characteristics. Appendix Table A.7 demonstrates that treated and matched control hospitals are similar across a broad range of observed characteristics.[32] The sample composed of 5,990 hospital-event pairs and respective matched control units.

To avoid biases due to variation in treatment timing, we adopt a stacked event-study approach analogous

[32] Appendix Table A.7 shows that treated hospitals and their matched controls are virtually indistinguishable across observed characteristics. By contrast, both groups differ substantially from hospitals that never experience a female appointment: treated and matched hospitals tend to be larger, more complex institutions in terms of procedures performed, more likely to operate as teaching or not-for-profit hospitals, and embedded in more urban and better-resourced areas.

to that used in Section 4.2. The estimating equation takes the form:

$$y_{itk} = \alpha_{ik} + \lambda_{tk} + \sum \beta_\tau \cdot \mathbb{1}\{\tau = t - k\} \cdot \text{Treated}_{ik} + \eta_{itk} \quad (3)$$

The variable $Treated$ is a binary indicator that equals one for observations hospitals experiencing the event at period k, and zero for the matched not-yet treated hospitals. The coefficient β measures the average effect of having a female manager on the outcome variable, y_{itk}. The key control variables of interests are the hospital-by-event α_{ik} and year-by-event τ_{tk} fixed effects. By including this rich set of fixed effects, estimates of the parameter of interests are not confounded by avoiding biases from problematic comparisons that would emerge in a standard two-way fixed effects models when treatment timing varies and treatment effects are heterogeneous (Wing et al., 2024).

The event-study coefficients and respective 95 percent confidence intervals are presented in Figure 7. Panel A displays the average surgical volume separately by gender, while Panel B focuses on the changes in the gender gap over time. The observations are weighted by the number of surgeons in each hospital. The analysis reveals no evidence of divergent trends between treated and control hospitals before the event, consistent with the underlying identification assumption that both groups would have continued to evolve similarly in the post-event period absent the treatment. The estimated event-time coefficients remain close to zero and statistically insignificant throughout the post-treatment period.

These average null effects may mask heterogeneous responses across hospitals with different starting points in terms of gender gaps. In hospitals with large gender gaps, gender biases are arguably more visible, harder to justify, and potentially more uncomfortable for staff and leadership alike. This salience may trigger stronger motivation or external pressure to address the disparities following the appointment of a female manager. Moreover, hospitals with wider disparities could rely on more discretionary or informal practices, creating greater room for the new manager to promote equity. To explore this possibility, we extend equation (3) by interacting the treatment indicator with hospitals' pre-treatment gender gap.

The results of this exercise are presented in Panels C and D of Figure 7. The figure plots estimates of θ_τ with respective 95 percent confidence intervals. Panel C shows that following the appointment of a female manager, the surgical volume of female surgeons gradually increases, while that of male surgeons declines in parallel. These effects are economically and statistically significant for both genders. Panel D displays the evolution in the gender and shows that the differential decline is statistically distinguinshable from zero. Our estimates indicate that hospitals in the top quartile of the baseline gap distribution saw the volume gap shrink by about 13 surgeries, a 27 percent reduction relative to their baseline average. Importantly, as one can also infer from the figure, these effects are not driven by pre-existent differential trends.

Overall, the findings provide suggestive evidence that while female managers on average do not uniformly reduce gender disparities in surgical activity, they can act as effective agents of change in hospitals where gender gaps are especially pronounced.

5 Tradeoffs between Gender Bias and Economic Efficiency?

Could gender gaps in surgical output be justified by principles of economic efficiency? The success of a surgery rests crucially on thorough preoperative planning, the surgeon's technical skill, and their ability to manage complications timely (Hamdorf and Hall, 2000). The observed volume gaps may be efficient if they arise from a process in which patients are allocated to the most competent available surgeons. Establishing the extent to which the current outcomes reflect such efficient sorting or organizational biases is important, as they have distinct implications. If female surgeons achieve better patient outcomes than their male peers using the same inputs, then correcting gender biases becomes compelling not only on grounds of equity but also of economic efficiency. But if instead male surgeons achieve better outcomes, then there could be a trade-off between equity and efficiency. We close the paper with an empirical investigation of this issue.

Some scholars argue that female surgeons follow clinical guidelines more closely, offer more patient-centered care, and communicate more effectively than male surgeons, with implications for patient outcomes (Silliman et al., 1999; Hershman et al., 2008). Several studies provide evidence consistent with this view by showing that female surgeons achieve better clinical outcomes (Wallis et al., 2017, 2023; Saka et al., 2024), but others find no meaningful differences associated with a surgeon's gender (Tsugawa et al., 2018). These mixed findings likely reflect the challenges that non-random patient-surgeon sorting poses for causal identification. As documented in Section 3, male surgeons tend to take on more complex cases compared to their female counterparts. Hence, simple comparisons of male and female surgeons are likely biased toward finding a female advantage in clinical performance even in the absence of a true causal relationship.

To overcome this identification challenge, we pursue a strategy that compares female and male surgeons in settings with limited discretion in case assignment. Specifically, we focus on patients admitted to the hospital through the emergency department and who are on the margin of survival. In high-risk situations, patients require immediate surgical attention, and even marginal differences in the quality of care delivered could translate into large differences in patient outcomes. Case allocation in such situations typically follows institutional protocols, such as surgeon availability or rotation schedules. Because minimizing treatment delays is critical in emergency care, surgeons cannot afford to wait for preferred patient profiles and have no option but to treat patients as they come. Therefore, focusing on "at-risk" patients substantially limits the scope for omitted variable bias, an approach that is similar in spirit to recent work examining the effects of financial incentives and peer groups (Batty and Ippolito, 2017; Silver, 2021). Following Silver (2021), we define at-risk patients as those in the top decile of the predicted mortality distribution, which is obtained using the random forest model described in Appendix C.[33] The estimation sample includes approximately 4.1 million at-risk patients, who have a mortality rate of 16 percent and account for more than two-thirds of all deaths.

Using this subsample, we regress multiple patient outcomes on an indicator for whether the surgeon is male. We control for hospital-by-day-of-week and hospital-by-month-by-year fixed effects to account for hospital-specific seasonal patterns in the types of patients treated. We also include fixed effects for surgical procedures interacted with surgeon specialty, which allows us to compare male and female surgeons with similar training backgrounds executing the same operation. Even after controlling for this comprehensive set of fixed effects, notable variation in surgeon gender remains across cases. A regression of surgeon gender on these covariates yields a R^2 of only 0.21, leaving almost 80 percent of the overall variation available for identification. Consistent with the limited scope for discretion in case assignment among at-risk cases,

[33] The random forest model predicts each patient's in-hospital mortality risk based on demographics, diagnosis, and admitting hospital. Random forests are particularly suited to this task because they flexibly capture complex interactions and nonlinearities that standard linear models may miss. The model was trained on the full universe of hospital admissions, generating individualized risk scores that allow us to identify the top decile of patients most at risk of dying during their stay.

Appendix Figure A.6 presents a balance checks showing that male and female surgeons treat clinically similar patients based on predetermined characteristics, including age, sex, race, comorbidities, and recent readmission history. The estimate coefficients are not only statistically indistinguishable from zero, but small in magnitude. The largest standardized coefficient is as small as 0.0025 (in absolute value).

The main results of this section are presented in Table 6. We examine both the quality and quantity dimensions of healthcare production. On the quality side, we consider mortality within a patient's hospital stay, the incidence of complications, and 30-day readmission. On the quantity side, we consider length of stay and overall inpatient spending. We further disaggregate spending into distinct categories, including diagnostic procedures, therapeutic procedures (inside and outside the operating room), medical devices and supplies, and room and board. These outcomes were selected to capture not only the clinical effectiveness of care but also the intensity of resource use across different stages of the hospitalization process.

Across all quality and quantity metrics, we do not observe statistically significant differences between female and male surgeons. The point estimates are consistently close to zero, with no discernible pattern favoring either group. For example, the estimated effect on mortality is just 0.1 percentage points, compared to a mean mortality rate of 16.3 percent among patients treated by female surgeons. Moreover, the coefficients are precisely estimated and can be bounded to a tight interval around zero. Indeed, we are able to rule out differences larger than 2 percent in the vast majority of outcomes. Including pre-determined patient characteristics does not materially alter the estimates, consistent with the absence of baseline imbalances. Moreover, Appendix Table A.8 shows that our results are robust to alternative definitions of at-risk patients. The estimated effects remain near zero and statistically insignificant when restricting the sample to patients above the 80th, 85th, 95th, or even 98th percentiles of baseline risk, as opposed to only those above the 90th percentile.

In sum, we find no evidence that male surgeons outperform their female counterparts in high-stakes surgical situations where discretion in case selection is limited. These results suggest that policies intended to address gender biases in the surgical workplace are unlikely to compromise patient outcomes or impose additional costs on the healthcare system.

6 Concluding Remarks

Our evidence shows that firms can generate inequality not only through wage-setting, but also by shaping the environments in which revenue-generating tasks are allocated and performed. We explore this question in the context of surgical specialists, where core tasks can be accurately measured, clearly affects career outcomes, and gender inequalities are widespread. Using rich administrative data from Brazil, we show that female surgeons perform 27 percent fewer surgical procedures than male counterparts. This gap is observed across specialties, emerges immediately upon entry in the surgical career, and account for most of the gap in payments directly tied to medical procedures. Rigorous empirical designs, which exploit surgeon mobility across hospitals and procedures, allow us to rule out individual factors and sorting as primary explanations. Instead, we provide causal evidence that local practice environments within hospitals are responsible for more than half of the gap. This effect operates through a reallocation of opportunities: male surgeons gain volume at the direct expense of their female colleagues within the same workplace. Women are not indifferent to gender biases in the workplace. Hospitals with more gender biases are systematically associated with lower scores in a revealed-preference measure of firm value based on the flows of female surgeons across hospitals, suggesting a link between organizational culture and market value.

These findings carry significant implications for both theory and policy. They reveal that gender gaps can arise and persist through organizational dynamics even in the absence of overt discrimination in wage setting. While wages may look neutral "on paper," inequality can still emerge through differences in access to tasks that are directly or indirectly rewarded. Crucially, these disparities occur without any corresponding difference in patient outcomes or resource efficiency between male and female surgeons. This suggests that redistributing surgical volume to reduce gender gaps would likely constitute a Pareto improvement, enhancing equity without sacrificing quality or increasing costs. More broadly, our results highlight the need to look beyond traditional explanations for wage inequality and focus on the often-invisible institutional factors that govern the allocation of work and opportunity.

The implications reach beyond healthcare. In many high-skill settings where pay and promotion track realized output—sales, professional services, R&D, academia—task allocation is discretionary, creating scope for environments to amplify disparities. Our results suggest that policy and managerial levers that improve the fairness and transparency of task assignment can raise equity. The moderating role of female managers in the most male-biased environments further underscores the importance of leadership and governance for task access. We view measuring and correcting environment-driven frictions in task realization as a core complement to the literature on firm pay premia, and a priority for future work that maps the micro-mechanisms and tests scalable interventions across sectors.

Bibliography

Abowd, John M., Francis Kramarz, and David N. Margolis, "High wage workers and high wage firms," *Econometrica*, 1999, *67* (2), 251–333.

Antecol, Heather, Kelly Bedard, and Jenna Stearns, "Equal but inequitable: Who benefits from gender-neutral tenure clock stopping policies?," *American Economic Review*, 2018, *108* (9), 2420–2441.

Atal, Juan Pablo, José Ignacio Cuesta, Felipe González, and Cristóbal Otero, "The economics of the public option: Evidence from local pharmaceutical markets," *American Economic Review*, 2024, *114* (3), 615–644.

Azmat, Ghazala and Rosa Ferrer, "Gender gaps in performance: Evidence from young lawyers," *Journal of Political Economy*, 2017, *125* (5), 1306–1355.

Babcock, Linda, Maria P Recalde, Lise Vesterlund, and Laurie Weingart, "Gender differences in accepting and receiving requests for tasks with low promotability," *American Economic Review*, 2017, *107* (3), 714–747.

Bagger, Jesper and Rasmus Lentz, "An empirical model of wage dispersion with sorting," *The Review of Economic Studies*, 2019, *86* (1), 153–190.

Barth, Erling, Sari Pekkala Kerr, and Claudia Olivetti, "The dynamics of gender earnings differentials: Evidence from establishment data," *European Economic Review*, 2021, *134*, 103713.

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

Bayard, Kimberly, Judith Hellerstein, David Neumark, and Kenneth Troske, "New evidence on sex segregation and sex differences in wages from matched employee-employer data," *Journal of labor Economics*, 2003, *21* (4), 887–922.

Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen, "Who becomes an inventor in America? The importance of exposure to innovation," *The Quarterly Journal of Economics*, 2019, *134* (2), 647–713.

Belloni, Alexandre and Victor Chernozhukov, "Least squares after model selection in high-dimensional sparse models," *Bernoulli*, 2013, *19* (2), 521–547.

Bertrand, Marianne and Kevin F Hallock, "The gender gap in top corporate jobs," *Ilr Review*, 2001, *55* (1), 3–21.

_ , Claudia Goldin, and Lawrence F Katz, "Dynamics of the gender gap for young professionals in the financial and corporate sectors," *American economic journal: applied economics*, 2010, *2* (3), 228–255.

_ , Sandra E Black, Sissel Jensen, and Adriana Lleras-Muney, "Breaking the glass ceiling? The effect of board quotas on female labour market outcomes in Norway," *The Review of Economic Studies*, 2019, *86* (1), 191–239.

Biernat, Monica and Kathleen Fuegen, "Shifting standards and the evaluation of competence: Complexity in gender-based judgment and decision making," *Journal of Social Issues*, 2001, *57* (4), 707–724.

Breiman, Leo, "Random forests," *Machine learning*, 2001, *45* (1), 5–32.

Bruns, Benjamin, "Changes in workplace heterogeneity and how they widen the gender wage gap," *American Economic Journal: Applied Economics*, 2019, *11* (2), 74–113.

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

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

Card, David, Ana Rute Cardoso, and Patrick Kline, "Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women," *The Quarterly journal of economics*, 2016, *131* (2), 633–686.

_ , Francesco Devicienti, Maria Christina Rossi, and Andrea Weber, "The Gender Gap in Career Trajec-

tories: Do Firms Matter?," *NBER Working Paper*, 2025, (w33730).
_ , Jörg Heining, and Patrick Kline, "Workplace heterogeneity and the rise of West German wage inequality," *The Quarterly journal of economics*, 2013, *128* (3), 967–1015.
Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer, "The effect of minimum wages on low-wage jobs," *The Quarterly Journal of Economics*, 2019, *134* (3), 1405–1454.
Chaisemartin, Clément De and Xavier d'Haultfoeuille, "Two-way fixed effects estimators with heterogeneous treatment effects," *American economic review*, 2020, *110* (9), 2964–2996.
Chetty, Raj and Nathaniel Hendren, "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects," *The quarterly journal of economics*, 2018, *133* (3), 1107–1162.
_ , John N Friedman, and Jonah E Rockoff, "Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates," *American economic review*, 2014, *104* (9), 2593–2632.
Chyn, Eric and Lawrence F Katz, "Neighborhoods matter: Assessing the evidence for place effects," *Journal of Economic Perspectives*, 2021, *35* (4), 197–222.
Cook, Cody, Rebecca Diamond, Jonathan V Hall, John A List, and Paul Oyer, "The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers," *The Review of Economic Studies*, 2021, *88* (5), 2210–2238.
Corradini, Viola, Lorenzo Lagos, and Garima Sharma, "Collective bargaining for women: How unions can create female-friendly jobs," *The Quarterly Journal of Economics*, 2025, p. qjaf024.
Coudin, Elise, Sophie Maillard, and Maxime Tô, "Family, firms and the gender wage gap in France," Technical Report, IFS Working Papers 2018.
Couronné, Raphael, Philipp Probst, and Anne-Laure Boulesteix, "Random forest versus logistic regression: a large-scale benchmark experiment," *BMC bioinformatics*, 2018, *19* (1), 270.
Cruz, Gabriel and Tomás Rau, "The effects of equal pay laws on firm pay premiums: Evidence from Chile," *Labour economics*, 2022, *75*, 102135.
Cullen, Zoë and Ricardo Perez-Truglia, "The old boys' club: Schmoozing and the gender gap," *American Economic Review*, 2023, *113* (7), 1703–1740.
DiNardo, John, Nicole M Fortin, and Thomas Lemieux, "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 1996, *64* (5), 1001–1044.
Dustmann, Christian, Attila Lindner, Uta Schönberg, Matthias Umkehrer, and Philipp Vom Berge, "Reallocation effects of the minimum wage," *The Quarterly Journal of Economics*, 2022, *137* (1), 267–328.
Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams, "Sources of geographic variation in health care: Evidence from patient migration," *The Quarterly Journal of Economics*, 2016, *131* (4), 1681–1726.
Flabbi, Luca, Mario Macis, Andrea Moro, and Fabiano Schivardi, "Do female executives make a difference? The impact of female leadership on gender gaps and firm performance," *The Economic Journal*, 2019, *129* (622), 2390–2423.
Flory, Jeffrey A, Andreas Leibbrandt, and John A List, "Do competitive workplaces deter female workers? A large-scale natural field experiment on job entry decisions," *The Review of Economic Studies*, 2015, *82* (1), 122–155.
Flynn, Patricia M, John D Leeth, and Elliott S Levy, "The evolving gender mix in accounting: implications for the future of the profession," *Selections-Los Angeles-*, 1996, *12*, 28–39.
Gerard, François, Lorenzo Lagos, Edson Severnini, and David Card, "Assortative matching or exclusionary hiring? The impact of employment and pay policies on racial wage differences in Brazil," *American Economic Review*, 2021, *111* (10), 3418–3457.
Goldin, Claudia, "A grand gender convergence: Its last chapter," *American economic review*, 2014, *104* (4), 1091–1119.
_ and Lawrence F Katz, "A most egalitarian profession: pharmacy and the evolution of a family-friendly occupation," *Journal of Labor Economics*, 2016, *34* (3), 705–746.
_ , Sari Pekkala Kerr, Claudia Olivetti, and Erling Barth, "The expanding gender earnings gap: Evidence

from the LEHD-2000 Census," *American Economic Review*, 2017, *107* (5), 110–114.

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

Hamdorf, JM and JC Hall, "Acquiring surgical skills," *British Journal of Surgery*, 2000, *87* (1), 28–37.

Harju, Jarkko, Simon Jäger, and Benjamin Schoefer, "Voice at work," *American Economic Journal: Applied Economics*, 2025, *17* (3), 271–309.

Hershman, Dawn L, Donna Buono, Russell B McBride, Wei Yann Tsai, Kathy Ann Joseph, Victor R Grann, and Judith S Jacobson, "Surgeon characteristics and receipt of adjuvant radiotherapy in women with breast cancer," *JNCI: Journal of the National Cancer Institute*, 2008, *100* (3), 199–206.

Jagsi, Reshma, Elizabeth A Guancial, Cynthia Cooper Worobey, Lori E Henault, Yuchiao Chang, Rebecca Starr, Nancy J Tarbell, and Elaine M Hylek, "The "gender gap" in authorship of academic medical literature—a 35-year perspective," *New England Journal of Medicine*, 2006, *355* (3), 281–287.

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard, "Children and gender inequality: Evidence from Denmark," *American Economic Journal: Applied Economics*, 2019, *11* (4), 181–209.

Kunze, Astrid and Amalia R. Miller, "Women helping women? Evidence from private sector data on workplace hierarchies," *Review of Economics and Statistics*, 2017, *99* (5), 769–775.

Lafortune, Julien, Jesse Rothstein, and Diane Whitmore Schanzenbach, "School finance reform and the distribution of student achievement," *American Economic Journal: Applied Economics*, 2018, *10* (2), 1–26.

Lentz, Rasmus, Suphanit Piyapromdee, and Jean-Marc Robin, "The anatomy of sorting—evidence from danish data," *Econometrica*, 2023, *91* (6), 2409–2455.

Lim, Wen Hui, Chloe Wong, Sneha Rajiv Jain, Cheng Han Ng, Chia Hui Tai, M Kamala Devi, Dujeepa D Samarasekera, Shridhar Ganpathi Iyer, and Choon Seng Chong, "The unspoken reality of gender bias in surgery: a qualitative systematic review," *PLoS One*, 2021, *16* (2), e0246420.

Ly, Dan P, Seth A Seabury, and Anupam B Jena, "Differences in incomes of physicians in the United States by race and sex: observational study," *Bmj*, 2016, *353.*

Matsa, David A. and Amalia R. Miller, "Chipping away at the glass ceiling: Gender spillovers in corporate leadership," *American Economic Review*, 2011, *101* (3), 635–639.

Molitor, David, "The evolution of physician practice styles: evidence from cardiologist migration," *American Economic Journal: Economic Policy*, 2018, *10* (1), 326–356.

Morchio, Iacopo and Christian Moser, "The gender pay gap: Micro sources and macro consequences," Technical Report, National Bureau of Economic Research 2024.

Niederle, Muriel and Lise Vesterlund, "Do women shy away from competition? Do men compete too much?," *The quarterly journal of economics*, 2007, *122* (3), 1067–1101.

Oaxaca, Ronald, "Male-female wage differentials in urban labor markets," *International economic review*, 1973, pp. 693–709.

OECD, "Reporting Gender Pay Gaps in OECD Countries: Guidance for Pay Transparency Implementation, Monitoring and Reform," 2023. Accessed: 2025-08-25.

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

Palladino, Marco Guido, Alexandra Roulet, and Mark Stabile, "Understanding the role of firms in the gender wage gap over time, over the life cycle, and across worker types," 2021.

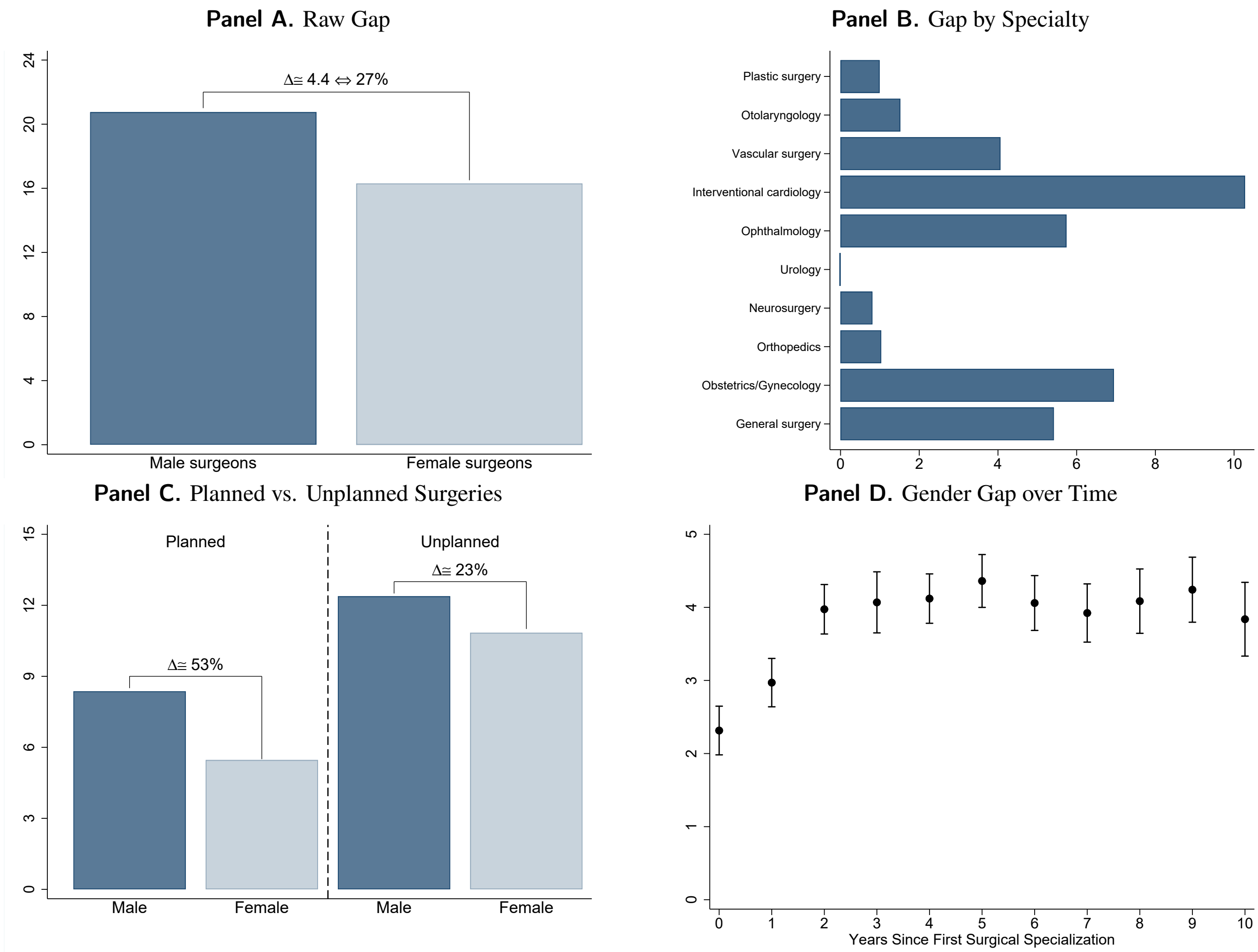
Reuben, Ernesto, Paola Sapienza, and Luigi Zingales, "How stereotypes impair women's careers in science," *Proceedings of the National Academy of Sciences*, 2014, *111* (12), 4403–4408.

Saka, Natsumi, Norio Yamamoto, Jun Watanabe, Christopher Wallis, Angela Jerath, Hidehiro Someko, Minoru Hayashi, Kyosuke Kamijo, Takashi Ariie, Toshiki Kuno et al., "Comparison of postoperative outcomes among patients treated by male versus female surgeons: a systematic review and meta-analysis," *Annals of Surgery*, 2024, *280* (6), 945–953.

Sarsons, Heather, "Interpreting signals in the labor market: evidence from medical referrals," *Job market*

paper, 2017, pp. 141–145.
Seabury, Seth A, Amitabh Chandra, and Anupam B Jena, "Trends in the earnings of male and female health care professionals in the United States, 1987 to 2010," *JAMA internal medicine*, 2013, *173* (18), 1748–1750.
Silliman, Rebecca A, Serkalem Demissie, and Susan L Troyan, "The care of older women with early-stage breast cancer: what is the role of surgeon gender?," *Medical care*, 1999, *37* (10), 1057–1067.
Silver, David, "Haste or waste? Peer pressure and productivity in the emergency department," *The Review of Economic Studies*, 2021, *88* (3), 1385–1417.
Sorkin, Isaac, "The role of firms in gender earnings inequality: Evidence from the United States," *American Economic Review*, 2017, *107* (5), 384–387.
_ , "Ranking firms using revealed preference," *The quarterly journal of economics*, 2018, *133* (3), 1331–1393.
Sun, Liyang and Sarah Abraham, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of econometrics*, 2021, *225* (2), 175–199.
Tsugawa, Yusuke, Anupam B Jena, E John Orav, Daniel M Blumenthal, Thomas C Tsai, Winta T Mehtsun, and Ashish K Jha, "Age and sex of surgeons and mortality of older surgical patients: observational study," *bmj*, 2018, *361.*
Vattuone, Giulia, "Worker Sorting and the Gender Wage Gap," Technical Report, Working paper 2023.
Wallis, Christopher JD, Angela Jerath, Khatereh Aminoltejari, Kirusanthy Kaneshwaran, Arghavan Salles, Natalie Coburn, Frances C Wright, Lesley Gotlib Conn, Zachary Klaassen, Amy N Luckenbaugh et al., "Surgeon sex and long-term postoperative outcomes among patients undergoing common surgeries," *JAMA surgery*, 2023, *158* (11), 1185–1194.
_ , Bheeshma Ravi, Natalie Coburn, Robert K Nam, Allan S Detsky, and Raj Satkunasivam, "Comparison of postoperative outcomes among patients treated by male and female surgeons: a population based matched cohort study," *Bmj*, 2017, *359.*
Weeks, William B, Tanner A Wallace, and Amy E Wallace, "How do race and sex affect the earnings of primary care physicians?," *Health Affairs*, 2009, *28* (2), 557–566.
Wing, Coady, Seth M Freedman, and Alex Hollingsworth, "Stacked difference-in-differences," Technical Report, National Bureau of Economic Research 2024.
Wynn, Alison T and Shelley J Correll, "Combating gender bias in modern workplaces," in "Handbook of the Sociology of Gender," Springer, 2018, pp. 509–521.
Zeltzer, Dan, "Gender homophily in referral networks: Consequences for the medicare physician earnings gap," *American Economic Journal: Applied Economics*, 2020, *12* (2), 169–197.
Zogg, Cheryl K, Lyndsay A Kandi, Hannah S Thomas, Mary A Siki, Ashley Y Choi, Camila R Guetter, Charlotte B Smith, Erica Maduakolam, Shreya Kondle, Sharon L Stein et al., "Comparison of male and female surgeons' experiences with gender across 5 qualitative/quantitative domains," *JAMA surgery*, 2023, *158* (2), e226431–e226431.

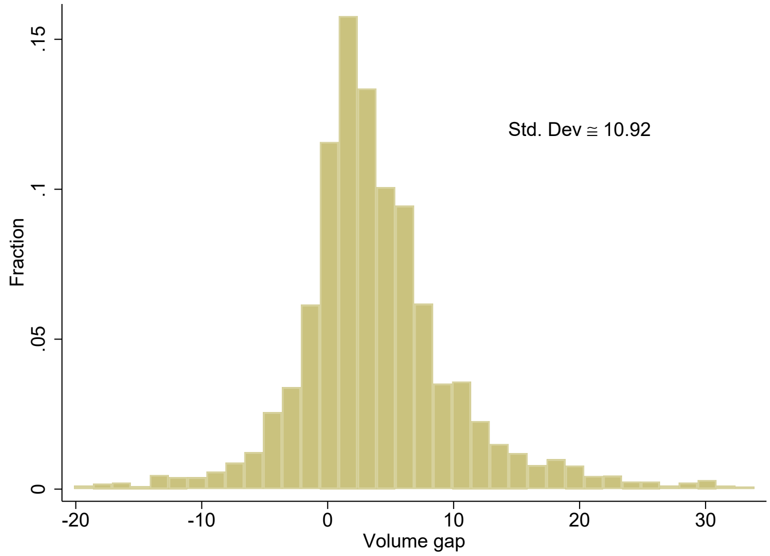
Figure 1: Raw Gap in Operative Volume



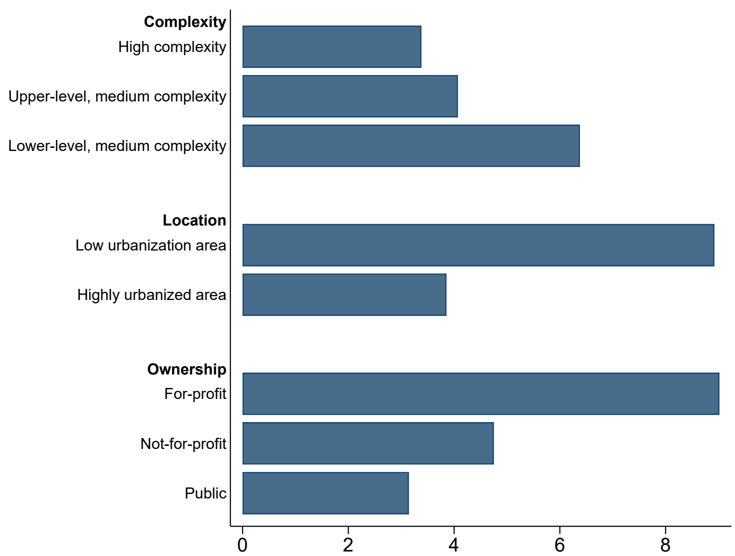
Notes. This figure compares the operative volume of female and male surgeons per hospital per quarter. Panel A shows the average volume gap. Panel B examines the volume gap across the largest specialties in terms of overall volume. Panel C compares the volume gap between planned and unplanned surgeries. Panel D explores the evolution of the volume gap since the first surgical specialty. To account for differences in cohorts, we control for the year of specialization fixed effects.

Figure 2: Volume Gender Gaps across Hospitals

Panel A. Distribution of Volume Gaps

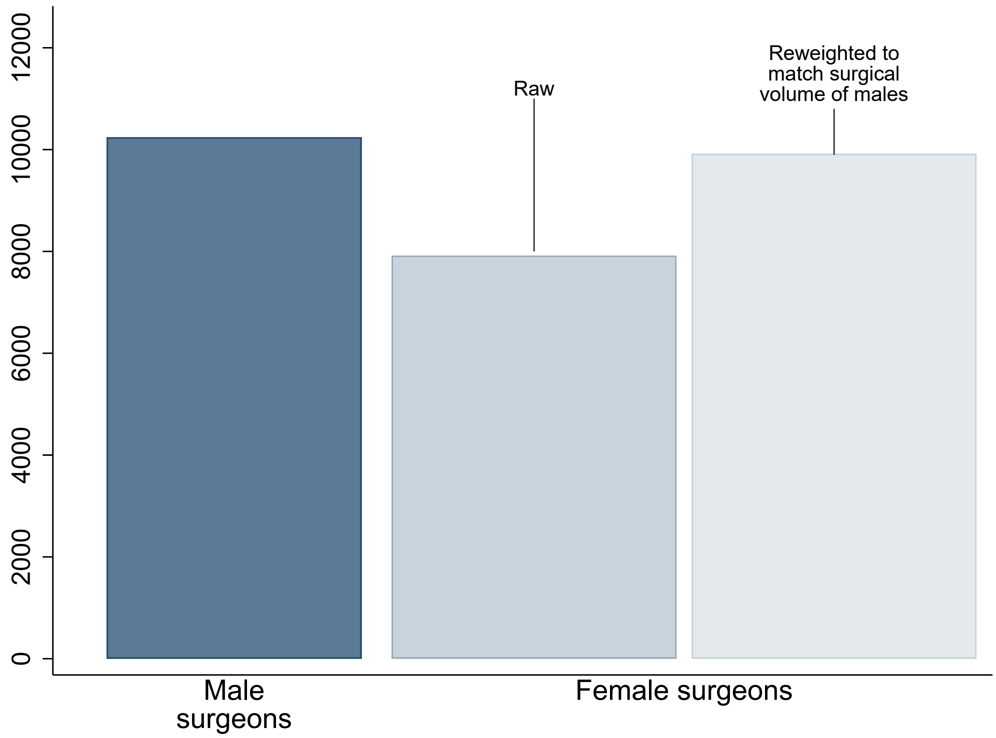


Panel B. Gaps by Hospital Characteristics



Notes. This figure illustrates the distribution of the volume gender gap across hospitals. Panel A shows a histogram of the gap, weighted by the number of surgeons in each hospital, excluding hospitals below the 1st and above the 99th percentiles of the distribution. Panel B reports the gap separately by complexity level, hospital location, and ownership status.

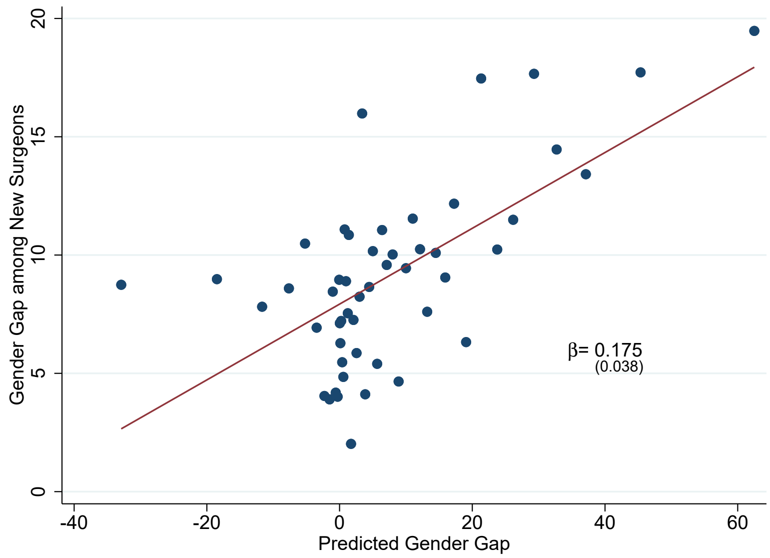
Figure 3: Gender Gap in Payments for Inpatient Procedures



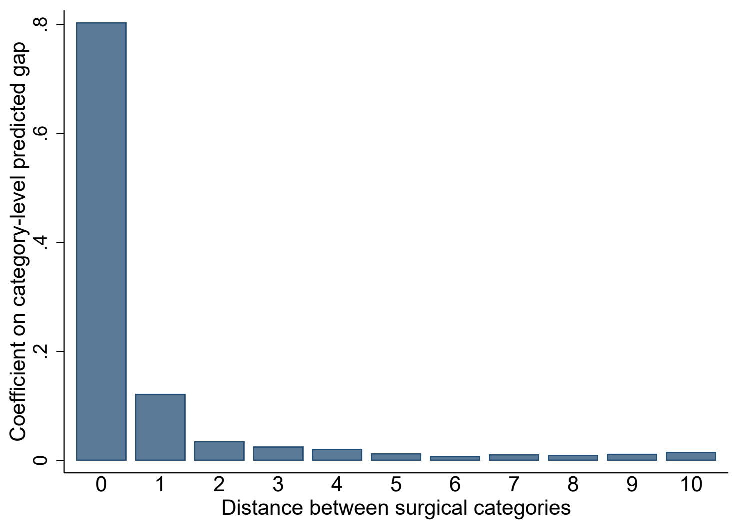
Notes: This figures compares payments for medical procedures performed by female and male surgeons. To assess the role of the operative volume gap in explaining this payment disparity, we non-parametrically reweigh the volume distribution of women to match that of men, following the approach developed by DiNardo et al. (1996). We first sort surgeons into 50 equal-sized bins based on operative volume and compute average surgeon payments by gender within each bin. We then calculate the counterfactual payment levels for female surgeons by taking the average of their observed payments within each volume bin, weighted by the number of male surgeons in each bin. All values are expressed in constant August 2024 Brazilian Reais.

Figure 4: Gender Gap among New Surgeons vs Predicted Gender Gap

Panel A. Hospital-by-cohort level

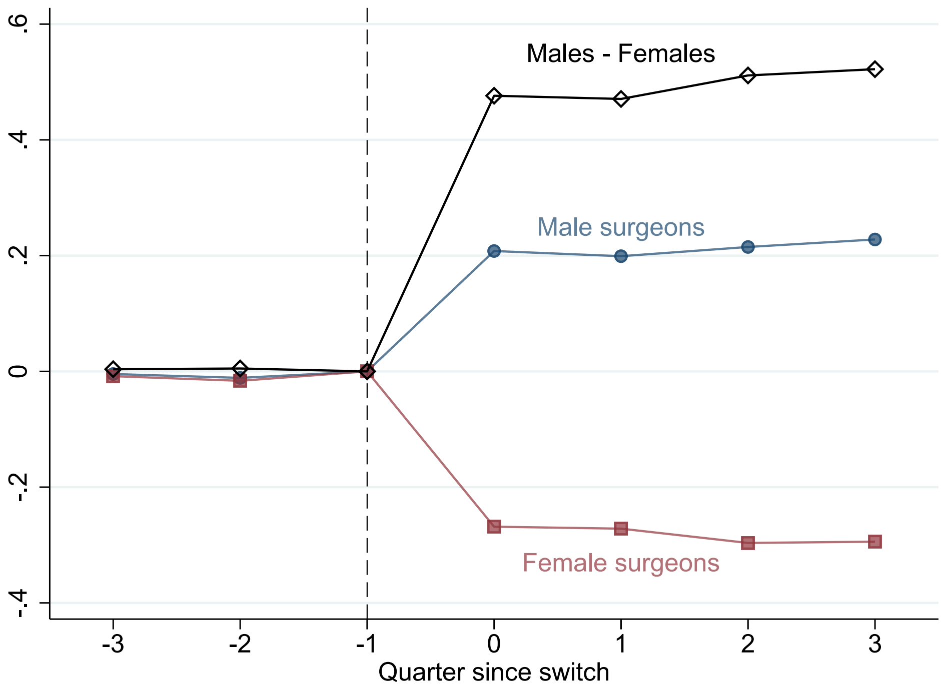


Panel B. Hospital-by-category-by-cohort level



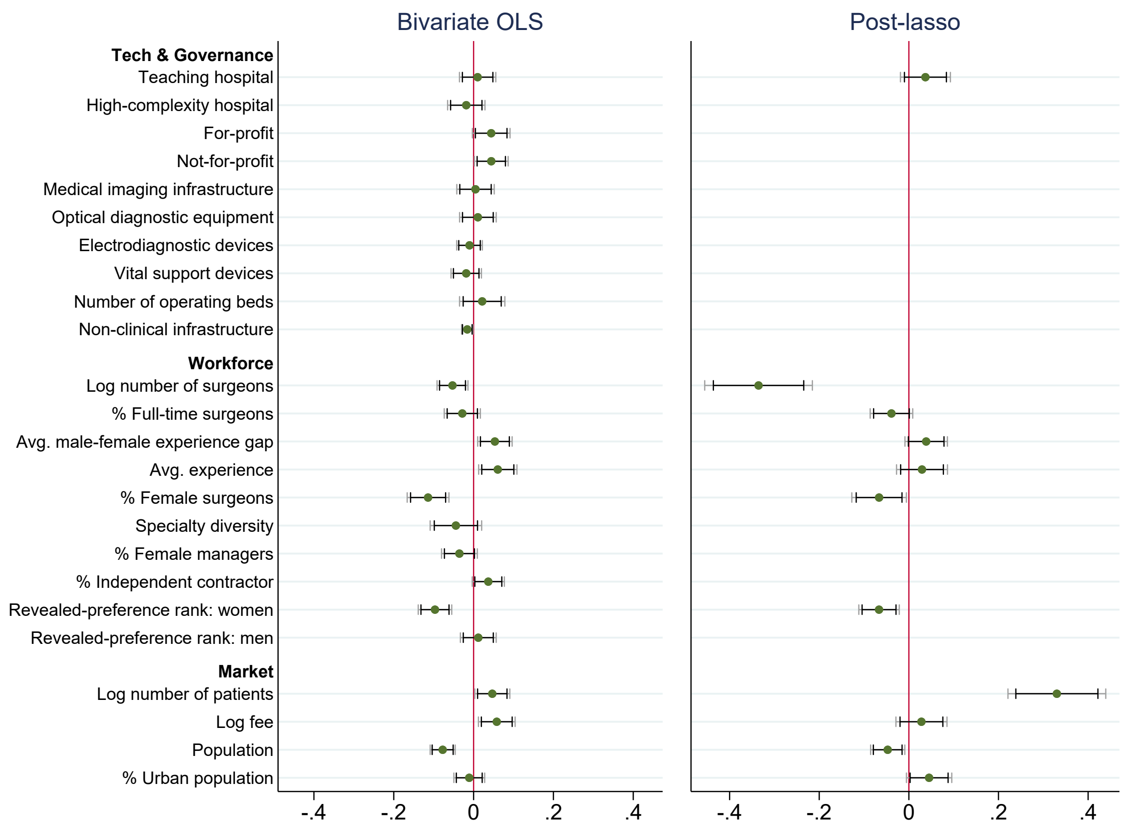
Notes. Panel A shows the relationship between the gender volume gap among newly affiliated surgeons and the predicted gap based on all previous cohorts. We aggregate the data for each cohort at the hospital-by-year-of-entry level to avoid observations with too few female and male surgeons entering a given hospital. Additionally, we exclude cells below and above the 1 and 99 percentiles of the predicted gap distribution to reduce the influence of extreme values. Panel B presents results from estimating this relationship at the hospital-by-procedure-by-cohort level. This regression controls for hospital, procedure and cohort fixed effects. To define how clinically close two surgical categories are, we measure the proportion of surgeons performing procedures in one category who also perform surgical procedures in the other category.

Figure 5: Triple Difference Estimates



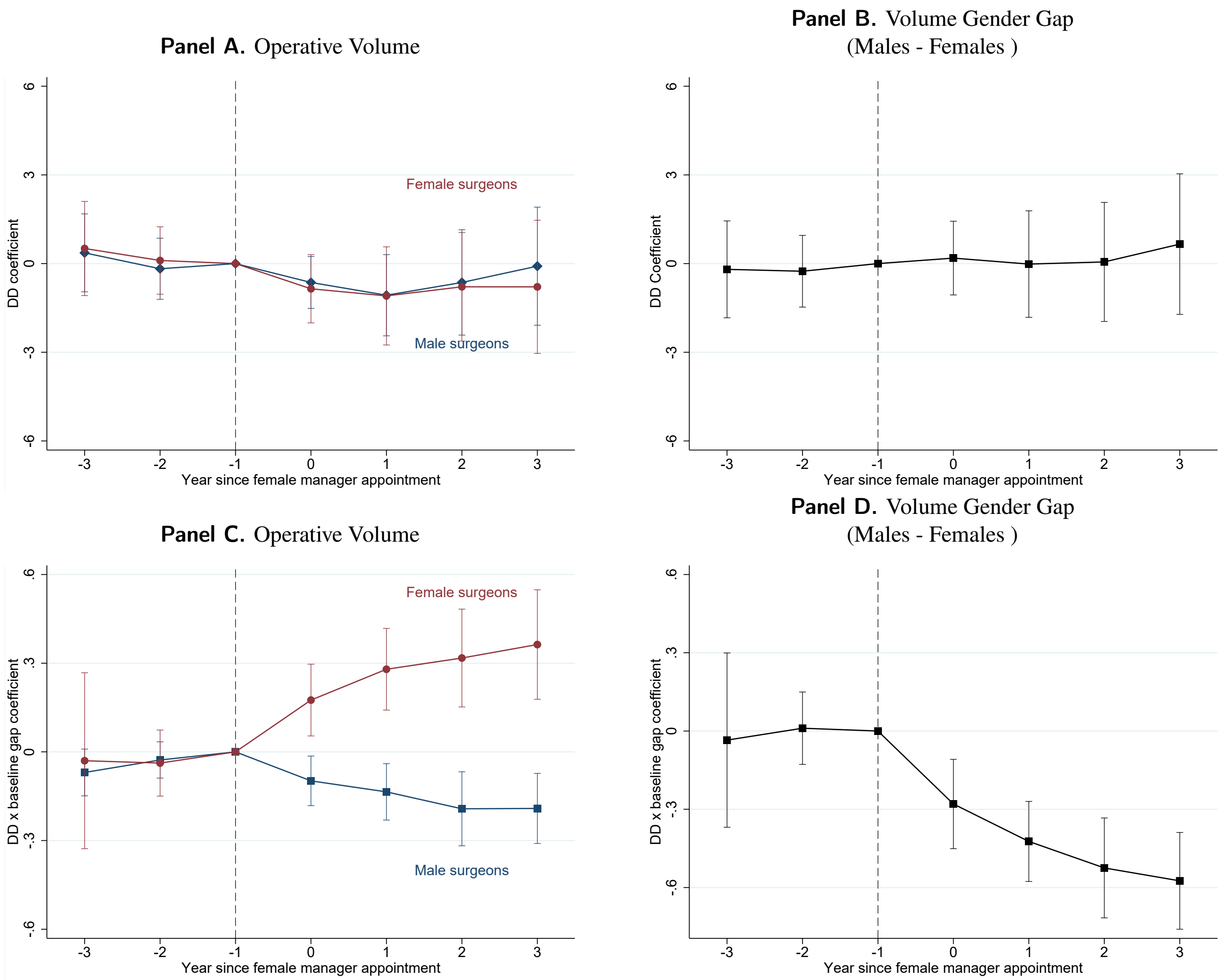
Notes: This figure presents results from estimating different versions of model (2). The black line corresponds to estimates from the triple difference specification. The navy and maroon lines plots the underlying difference- in-differences estimates separately for male and female surgeons. Specifically, we estimate the difference-in-differences version of equation (2), replacing the post-move indicator in the interaction with the origin- destination gender gap difference by a full set of event-time dummies. he estimation sample consists of surgeons who changed their hospital affiliation between 2008 and 2024. Our analysis include all movers, including those who moved more than once in the study period. For surgeons with multiple moves, we include each move by stacking the data and treating every surgeon-event as a separate unit of analysis.

Figure 6: Correlates of Work Environment Effects



Notes. This figure explores the determinants of hospital-level work environment effects. We construct these effects using the estimated hospital fixed effects by gender, defined as $\hat{\gamma}^M - \hat{\gamma}^F$. Panel A (left) displays point estimates together with heteroskedasticity-robust confidence intervals at the 90% and 95% levels from simple bivariate OLS regressions. Panel B (right) presents results from a post-Lasso multivariate specification. Both the dependent variable and all covariates are standardized to have mean zero and a standard deviation of one. To implement the post-Lasso approach, we first estimate a Lasso model including the full set of covariates, selecting the penalty parameter through 10-fold cross-validation that minimizes mean squared error. We then re-estimate a multivariate OLS model using only the covariates retained by the Lasso. The analysis covers approximately 3,200 hospitals.

Figure 7: Effects of Female Managers on Volume Outcomes



Notes. This figure reports the results of the stacked event-study model (3) along with 95 percent confidence intervals. In panels A and C, the dependent variable is the average operative volume for female and male surgeons. In panels B and D, the dependent variable is the gender volume gap, defined as the difference between the average operative volume of male and female surgeons. Observations are weighted by the number of surgeons in each hospital-year-event cell. For panels B and D, we re-estimate model (3) including an interaction between the event-time dummies and the baseline gender gap in operative volume. Standard errors are clustered at the hospital level to account for serial correlation in the residuals.

Table 1: Summary Statistics

	Men (1)	Women (2)
Specialty:		
General surgery	0.286	0.185
Obstetrics/Gynecology	0.163	0.466
Orthopedics	0.187	0.048
Neurosurgery	0.046	0.018
Urology	0.052	0.009
Ophthalmology	0.029	0.047
Interventional cardiology	0.043	0.030
Vascular surgery	0.034	0.028
Otolaryngology	0.027	0.034
Plastic surgery	0.027	0.023
Other	0.106	0.111
Professional experience	20.941	16.390
Work agreement:		
Independent contractor	0.598	0.519
Regular employment contract	0.177	0.199
Locum tenens	0.133	0.127
Civil servant	0.334	0.355
Weekly hours worked	28.333	27.275
Number of physicians	80986	39538

Notes. Summary statistics are for the overall sample. The data are at the surgeon-by-hospital-by-quarter level.

Table 2: Surgeon Gender and Operative Volume

	Dependent variable is number of surgeries:					
	(1)	(2)	(3)	(4)	(5)	(6)
Surgeon is male	4.446	5.037	4.109	4.069	3.690	3.686
	[0.221]***	[0.266]***	[0.267]***	[0.259]***	[0.262]***	[0.261]***
Female mean	16.3	16.3	16.3	16.3	16.3	16.3
Especialty FE		✓	✓	✓	✓	✓
Specialty-by-cohort FE			✓	✓	✓	✓
Medical school FE				✓	✓	✓
Contractual agreements					✓	✓
Weekly hours worked						✓
R^2	0.003	0.013	0.022	0.031	0.037	0.043
Observations	3,445,664	3,445,664	3,445,664	3,445,664	3,445,664	3,445,664

Notes. This table reports estimates of the gender gap in operative volume after progressively adding controls for observable characteristics. Column 1 shows the unconditional gap. Column 2 adds specialty fixed effects, while Column 3 further includes specialty-by-year-of-professional-registration fixed effects. Column 4 introduces a full set of medical school fixed effects. Column 5 controls for the type of contractual arrangement (independent contractor, regular employment contract, locum tenens, and civil servant). Column 6 accounts for working hours by dividing them into 20 equally sized bins and including a full set of bin dummies. The data are organized at the surgeon-by-hospital-by-quarter level, and standard errors are clustered at the hospital level. * $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

Table 3: Surgeon Gender and Lower-Value Tasks

		Cases as co-managing specialist			
	Nonclinical hours	Share of cases	Total cases	Cases involving intra-operative care	Cases involving peri-operative care
	(1)	(2)	(3)	(4)	(5)
Surgeon is male	-0.107	-0.051	-0.414	-0.289	-0.243
	[0.038]***	[0.004]***	[0.181]**	[0.147]**	[0.100]**
Female mean	3.32	0.327	8.15	5.55	2.86
Basic controls	✓	✓	✓	✓	✓
R^2	0.201	0.092	0.031	0.030	0.018
Observations	3,410,618	3,445,359	3,445,664	3,445,664	3,445,664

Notes. This table examines the gender gap in lower-value tasks. In Column 1, the dependent variable is total hours worked on nonclinical activities, including administrative duties, management, and other functions not directly related to patient care. Column 2 considers the share of surgical cases in which the surgeon appears as a co-managing specialist rather than the primary operator. Column 3 uses instead the count of such cases. Column 4 restricts this measure to co-management involving intra-operative care, while Column 5 focuses on peri-operative care. All specifications include the full set of covariates from Column 6 of Table 2. The data are organized at the surgeon-by-hospital-by-quarter level, with standard errors clustered at the hospital level.
* $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

Table 4: Triple Difference Estimates

	Dependent variable is number of surgeries					
	(1)	(2)	(3)	(4)	(5)	(6)
Triple difference	0.492	0.495	0.493	0.500	0.492	0.498
	[0.124]***	[0.121]***	[0.121]***	[0.123]***	[0.124]***	[0.121]***
Female mean	17.934	17.934	17.934	17.934	17.934	17.934
Δ Hosp. Char. $\times$ time FE		✓	✓			✓
Δ Hosp. Char. $\times$ time FE $\times$ gender			✓			✓
Employment conditions				✓		✓
Specialty FE					✓	✓
Basic controls	✓	✓	✓	✓	✓	✓
R^2	0.528	0.539	0.542	0.533	0.528	0.544
Observations	146,769	146,769	146,769	146,769	146,769	146,769

Notes. This table reports estimates from alternative specifications of the triple-difference model described in the text. The baseline controls include surgeon-by-event fixed effects, quarter-by-gender-by-event fixed effects, and an interaction between the post-move indicator and the origin–destination difference in gender gaps. Employment conditions add individual-level controls for hours worked and type of contractual agreement. The vector Δ Hosp. Char. captures origin–destination differences in hospital characteristics, including average hours worked, the distribution of contractual agreements among surgeons, ownership status, location, overall surgical volume, and hospital complexity category. Standard errors are clustered at the hospital level.

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

Table 5: Decomposition of Gender Gaps

	High-versus-Low Gender Gap Hospitals				
	Above/ below median	Top & bottom 25%	Top & bottom 10%	Top & bottom 5%	Mean Gap
	(1)	(2)	(3)	(4)	(5)
Share of difference due to:					
Hospital component	0.684	0.621	0.682	0.648	0.811
Practice environment	0.722	0.708	0.763	0.684	0.581
Institutional sorting	-0.038	-0.086	-0.081	-0.036	0.230
Endowment effect	0.316	0.379	0.318	0.352	0.189

Notes.This table reports the decomposition results described in Section 4.2.C. Columns 1–4 compare gender gaps across hospitals with different positions in the gap distribution. Column 1 contrasts hospitals above and below the median, while Columns 2–4 focus on increasingly extreme splits, comparing hospitals in the top quartile, decile, and top 5 percent with those in the corresponding bottom groups. Column 5 estimates the contribution of each component to the average gender gap in operative volume.

Table 6: Surgeon Gender and Patient Outcomes

	Coefficient on male surgeon					
	no controls		with controls		Female mean	Observations
Dependent variable:	(1)	(2)	(3)	(4)	(5)	(6)
Care Quality						
Major complication	-0.001	[0.002]	-0.001	[0.002]	0.160	4,544,813
Mortality	0.001	[0.001]	0.001	[0.001]	0.163	4,544,813
30-day readmission	-0.001	[0.001]	-0.001	[0.001]	0.157	4,544,813
Care Quantity						
Length of stay	0.005	[0.056]	0.007	[0.056]	9.31	4,544,813
Total spending	-18.033	[33.480]	-16.299	[33.586]	5891.3	4,544,813
Diagnostic spending	-3.611	[2.601]	-3.574	[2.579]	241.7	4,544,813
Therapeutic spending (outside OR)	0.385	[3.048]	0.690	[3.040]	283.5	4,544,813
Therapeutic spending (in OR)	-8.579	[12.603]	-8.301	[12.610]	2297.4	4,544,813
Spending on devices/supplies	-11.510	[10.047]	-11.533	[10.000]	879.9	4,544,813
Room and board spending	9.083	[25.251]	10.196	[25.263]	2168.0	4,544,813

Notes. This table investigates the relationship between surgeon gender and patient outcomes. Each row reports results from a separate regression. The sample is restricted to patients admitted through the emergency department who fall in the top decile of the predicted mortality distribution, where mortality risk is estimated using the random forest model described in Appendix C. All specifications include hospital-by-day-of-week, hospital-by-month-by-year, and surgery-by-specialty fixed effects. Columns 3 and 4 additionally control for patient characteristics, including age, gender, comorbidities, race, and hospital admission within the previous 30 days. Spending outcomes are expressed in constant August 2024 Brazilian Reais. Standard errors are clustered at the hospital level.
* $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

Online Appendix to "Organizations and the Production Side of Gender Inequality: Evidence from Surgeons"

Danyelle Branco, Bladimir Carrillo, Gustavo Cordeiro, Laisa Rachter

October 24, 2025

List of Figures 1

List of Tables 2

A Additional Figures and Tables 3

B Data Sources and Construction 15

B.1 Provider Data 15

B.2 Inpatient Procedure Registries 16

B.3 30-Day Readmissions 16

C Mortality Risk Prediction Using Random Forests 17

References 18

List of Figures

A.1 Gender Gaps in Operative Volume Across Hospitals in Logs 3

A.2 Gender Gaps in Operative Volume Across Hospitals: Level versus Logs 3

A.3 Change in Surgeon Volume across Quartiles of Mean Operative Volume 4

A.4 Distribution of Destination-Origin Difference in Volume Gender Gaps 4

A.5 Mean Decomposition: Robustness to Alternative Definitions of Reference Hospitals 5

A.6 Balance Tests: Surgeon Gender versus Patient Characteristics 6

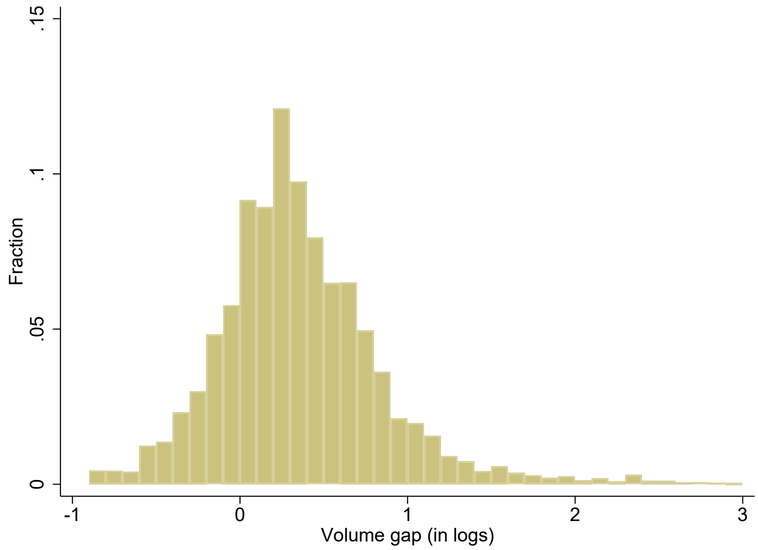
Contact information:

List of Tables

A.1 Surgeon Gender and Operative Volume: Surgeon Level Analysis 7
A.2 Surgeon Gender and Operative Volume per Hour Worked 8
A.3 Surgeon Gender and Operative Volume: Surgeon-by-Day Level Analysis 9
A.5 Predicted and Actual Volume Gender Gap Among Newly Affiliated Surgeons 11
A.6 Correlates of Practice Environment Effects: Components of the Revealed-Preference Measure of Firm Value 12
A.7 Treated and Untreated Units: Manager Analysis 13
A.8 Surgeon Gender and Patient Outcomes: Alternative "At-risk" Subsamples 14

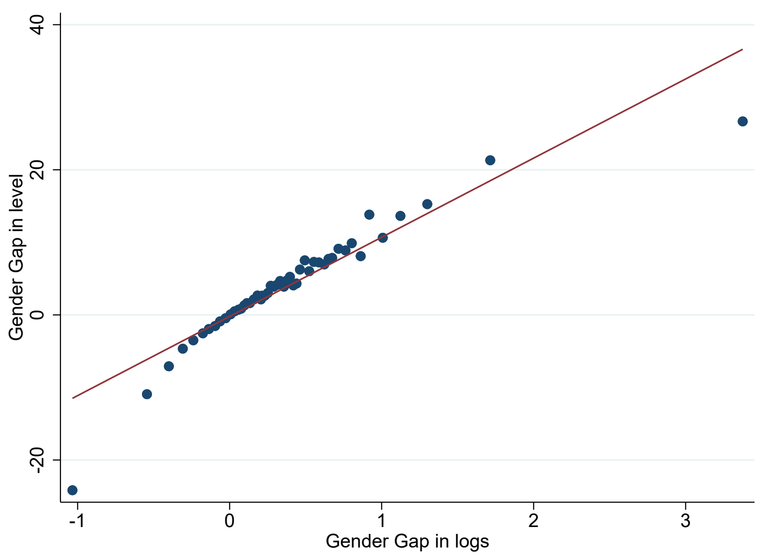
A Additional Figures and Tables

Figure A.1: Gender Gaps in Operative Volume Across Hospitals in Logs



Notes. This figure presents a histogram of the gender gap in log operative volume across hospitals.

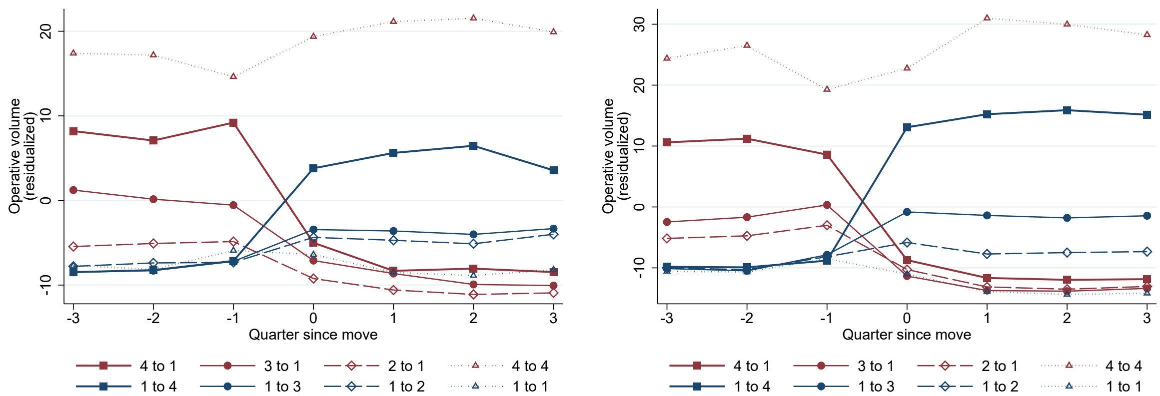
Figure A.2: Gender Gaps in Operative Volume Across Hospitals: Level versus Logs



Notes. This figure presents a binned scatter plot of the relationship between the within-hospital gender gap in operative volume measured in levels and in logs.

Figure A.3: Change in Surgeon Volume across Quartiles of Mean Operative Volume

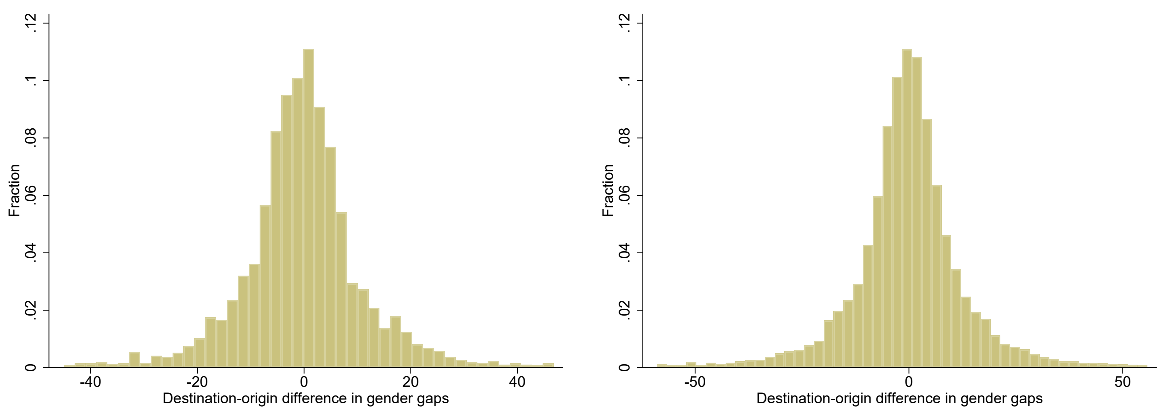
Panel A. Females **Panel B.** Males



Notes. To test for symmetry, this figure classifies hospitals into four quartiles according to their average operative volume and tracks surgeons' volumes before and after moves from the lowest-volume quartile to each quartile and vice versa. We residualize operative volume by removing surgeon-specific and time-specific trends to account for general changes in volume outcomes unrelated to hospital moves.

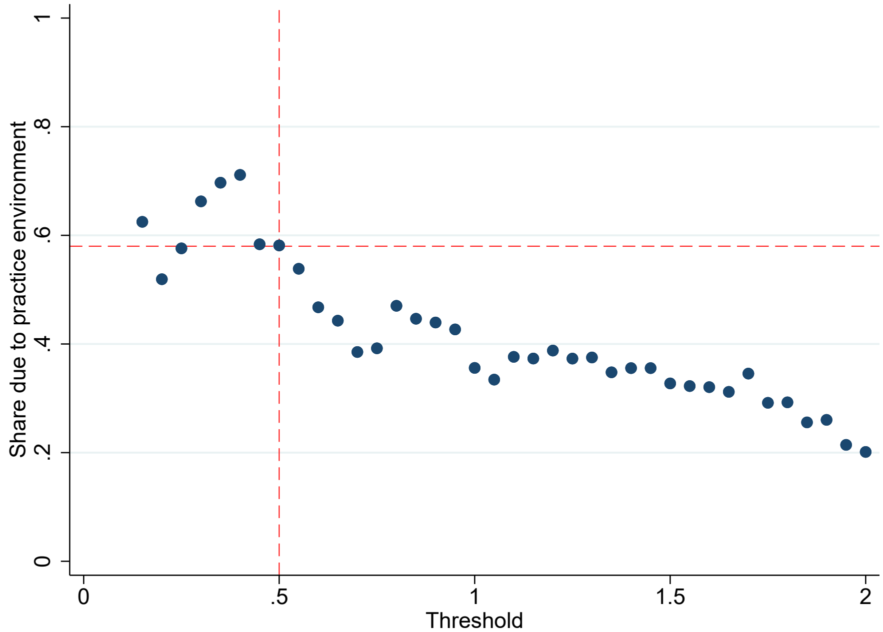
Figure A.4: Distribution of Destination-Origin Difference in Volume Gender Gaps

Panel A. Females **Panel B.** Males



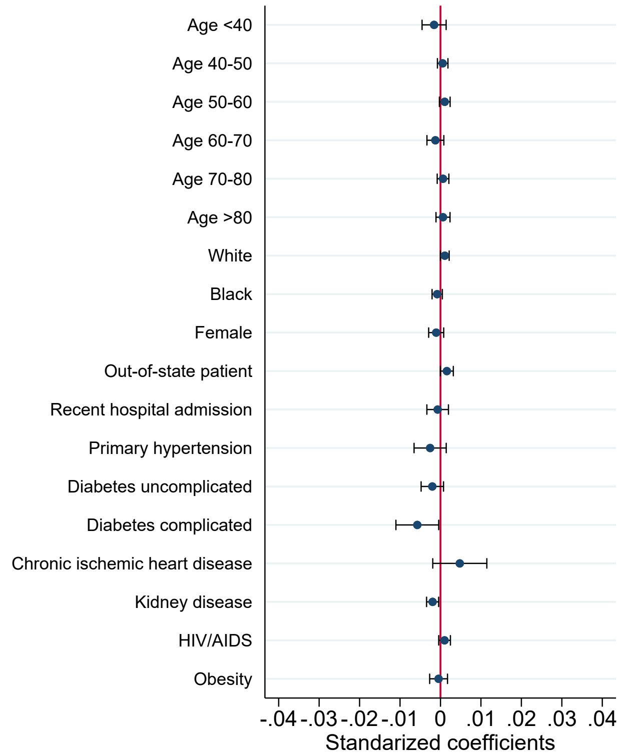
Notes. This figure plots the distribution of the origin-destination difference in gender gaps, separately for female and male movers.

Figure A.5: Mean Decomposition:
Robustness to Alternative Definitions of Reference Hospitals



Notes. This figure reports estimates of the contribution of the work environment effect to the mean gender gap under alternative normalization assumptions. In the baseline specification, the reference group is defined as public hospitals with small gender gaps in operative volume, where a gap is considered small if it is less than 0.5 surgeries in absolute value and statistically indistinguishable from zero. The figure explores alternative thresholds for defining a small gap.

Figure A.6: Balance Tests:
Surgeon Gender versus Patient Characteristics



Notes. Each coefficient comes from a separate regression in which a predetermined characteristic is the dependent variable. The key independent variable is an indicator for whether the surgeon is female. The sample is restricted to patients admitted through the emergency department who fall in the top decile of the predicted mortality distribution, with mortality risk estimated using the random forest model described in Appendix C. All specifications include hospital-by-day-of-week, hospital-by-month-by-year, and surgery-by-specialty fixed effects. Standard errors are clustered at the hospital level.

Table A.1: Surgeon Gender and Operative Volume:
Surgeon Level Analysis

	Dependent variable is number of surgeries					
	(1)	(2)	(3)	(4)	(5)	(6)
Surgeon is male	7.007	7.540	6.751	6.753	5.611	5.069
	[0.281]***	[0.325]***	[0.340]***	[0.336]***	[0.332]***	[0.327]***
Female mean	19.6	19.6	19.6	19.6	19.6	19.6
Especialty FE		✓	✓	✓	✓	✓
Specialty-by-cohort FE			✓	✓	✓	✓
Medical school FE			✓	✓	✓	✓
Contractual agreements					✓	✓
Weekly hours worked						✓
R^2	0.005	0.019	0.028	0.042	0.063	0.084
Observations	2,726,975	2,726,975	2,726,975	2,726,975	2,726,975	2,726,975

Notes. This table reports estimates of the gender gap in operative volume after progressively adding controls for observable characteristics. The data are organized at the surgeon-by-quarter level. Column 1 shows the unconditional gap. Column 2 adds specialty fixed effects, while Column 3 further includes specialty-by-year-of-professional-registration fixed effects. Column 4 introduces a full set of medical school fixed effects. Column 5 controls for the type of contractual arrangement (independent contractor, regular employment contract, locum tenens, and civil servant). Column 6 accounts for working hours by dividing them into 20 equally sized bins and including a full set of bin dummies. Standard errors are clustered at the surgeon level.
* $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

Table A.2: Surgeon Gender and Operative Volume per Hour Worked

	Dependent variable is number of surgeries per hour	
	(1)	(2)
Surgeon is male	0.338	0.250
	[0.025]***	[0.024]***
Female mean	16.4	16.4
Basic controls		✓
R^2	0.002	0.117
Observations	3,410,513	3,410,513

Notes. This table reports the relationship between surgeon gender and operative volume per hour worked. Column 2 includes the controls reported in Column 6 of Table 2. The data are structured at the surgeon-by-hospital-by-quarter level, with standard errors clustered at the hospital level.

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

Table A.3: Surgeon Gender and Operative Volume:
Surgeon-by-Day Level Analysis

	Dependent variable is number of surgeries		
	(1)	(2)	(3)
Surgeon is male	0.152	0.115	0.113
	[0.010]***	[0.011]***	[0.011]***
Female mean	0.862	0.862	0.862
Basic controls		✓	✓
Day-of-practice FE			✓
R^2	0.002	0.040	0.051
Observations	69,197,491	69,197,491	69,197,491

Notes. This table reports the relationship between surgeon gender and operative volume using data at the surgeon-by-day-by-hospital level. Columns 2 and 3 include the controls reported in Column 6 of Table 2. Standard errors clustered at the hospital level.
* $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

Table A.4: Surgeon Gender and Mortality Risk

	Dependent variable is predicted mortality	
	(1)	(2)
Surgeon is male	0.0006	0.0001
	[0.00014]***	[0.00005]*
Female mean	0.011	0.011
Procedure FE		✓
Basic controls	✓	✓
R^2	0.2734	0.6940
Observations	67,826,568	67,826,568

Notes. This table examines the relationship between surgeon gender and the predicted mortality risk of patients. The data are structured at the patient-by-surgeon level, with mortality predicted using the random forest model described in Appendix C. All specifications include the controls reported in Column 6 of Table 2, while Column 3 additionally includes surgical procedure fixed effects. Standard errors are clustered at the hospital level. * p < 0.1;** p < 0.05;*** p < 0.01.

Table A.5: Predicted and Actual Volume Gender Gap Among Newly Affiliated Surgeons

					Exclude largest category within			
	Baseline		No highly concentrated hospitals		Hospital		Hospital-surgical group	
	Same category	Top 10 closest categoeries	Same category	Top 10 closest categoeries	Same category	Top 10 closest categoeries	Same category	Top 10 closest categoeries
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Predicted gap	0.804	0.017	0.802	0.029	0.792	0.038	0.765	0.055
	[0.005]***	[0.002]***	[0.005]***	[0.004]***	[0.004]***	[0.005]***	[0.004]***	[0.009]***
Hospital FE	✓	✓	✓	✓	✓	✓	✓	✓
Surgical category FE	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.210	0.166	0.269	0.178	0.314	0.168	0.331	0.171
Observations	2,081,494	2,192,058	1,844,264	1,922,611	2,075,103	2,163,252	1,932,026	1,987,123

Notes. This table examines the relationship between predicted and actual gender gaps in operative volume among affiliated surgeons, measured at the hospital-by-year-of-entry-by-procedure level. The predicted gap is computed using all preceding cohorts. The "Same category" specification assesses the extent to which the gender gap in procedure p predicts the actual gap for that same procedure, while "Top 10 closest categories" examines how well the gap in p predicts the actual gap in the 10 most clinically similar procedures. Clinical similarity between procedures is measured by the proportion of surgeons performing one procedure who also perform the other. Columns 3 and 4 test robustness by excluding hospitals in the top quartile of surgical concentration (measured via the Herfindahl-Hirschman index). Columns 5 and 6 exclude procedures with the highest case volumes within the hospital, and Columns 7 and 8 further exclude procedures with the highest volumes within both the hospital and broad surgical group. Standard errors are clustered at the hospital level.

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

Table A.6: Correlates of Practice Environment Effects: Components of the Revealed-Preference Measure of Firm Value

	Dependent variable is $\hat{\gamma}_h^M - \hat{\gamma}_h^F$		
	(1)	(2)	(3)
PageRank index:			
Female surgeons	-0.057		
	[0.022]**		
Male surgeons	0.019		
	[0.026]		
Poaching rate:			
Female surgeons		-0.066	
		[0.023]***	
Male surgeons		0.032	
		[0.024]	
Retention rate:			
Female surgeons			-0.110
			[0.018]***
Male surgeons			0.097
			[0.020]***
R^2	0.003	0.004	0.014
Observations	2,429	2,771	3,183

Notes. This table reports the correlation between surgeon-specific components of the revealed-preference measure of hospital value and measures of the practice environment. The dependent variable is the estimated hospital-level gender gap in operative volume, $\hat{\gamma}_h^M - \hat{\gamma}_h^F$. Columns (1)–(3) correspond to regressions on different measures of the hospital environment: PageRank index (Column 1), poaching rate (Column 2), and retention rate (Column 3). Standard errors, reported in brackets, are robust to heteroskedasticity. The sample consists of all hospitals with at least one male and one female surgeon contributing to the measure.
* p < 0.1;** p < 0.05;*** p < 0.01.

Table A.7: Treated and Untreated Units: Manager Analysis

	Treated hospitals (1)	Matched hospitals (2)	All untreated (3)
Teaching hospital	0.250	0.251	0.0496
	[0.433]	[0.433]	[0.217]
High-complexity hospital	0.421	0.419	0.158
	[0.494]	[0.493]	[0.364]
For-profit	0.0489	0.0489	0.239
	[0.216]	[0.216]	[0.427]
Not-for-profit	0.417	0.409	0.349
	[0.493]	[0.492]	[0.477]
Medical imaging infrastructure	1.713	1.574	0.548
	[2.415]	[2.030]	[0.934]
Optical diagnostic equipment	1.005	0.870	0.282
	[2.747]	[2.271]	[1.365]
Electrodiagnostic devices	0.891	0.792	0.296
	[2.512]	[2.187]	[1.466]
Vital support devices	3.298	2.970	0.754
	[4.860]	[4.067]	[1.734]
Number of operating beds	45.81	41.91	15.89
	[60.70]	[50.57]	[25.82]
Non-clinical infrastructure	0.431	0.389	0.186
	[0.596]	[0.490]	[0.322]
Number of surgeons	44.79	43.78	9.145
	[59.26]	[48.18]	[18.38]
% Full-time surgeons	0.0642	0.0528	0.110
	[0.125]	[0.106]	[0.249]
Avg. male-female experience gap	4.246	4.349	4.886
	[9.547]	[7.583]	[11.62]
Avg. experience	19.69	19.59	24.25
	[6.850]	[4.991]	[9.953]
% Female surgeons	0.265	0.256	0.130
	[0.164]	[0.152]	[0.225]
Specialty diversity	0.395	0.378	0.674
	[0.261]	[0.251]	[0.303]
% Independent contractor	0.516	0.513	0.510
	[0.459]	[0.465]	[0.477]
Volume	2456.1	2366.0	673.5
	[2665.2]	[2436.4]	[1342.9]
Avg. fee	358.6	344.6	334.2
	[306.7]	[264.7]	[356.9]
Population	1269447.9	1110202.7	531449.4
	[2617106.4]	[2452600.1]	[1567055.9]
% Urban population	0.883	0.887	0.753
	[0.139]	[0.131]	[0.195]

Notes. This table reports summary statistics for treated hospitals, matched control hospitals, and all untreated hospitals in the manager analysis. Treated hospitals are those where a female manager assumes a managerial role, while matched hospitals are selected based on pre-event characteristics to resemble the treated hospitals. All variables are measured prior to the managerial appointment. Standard deviations are reported in brackets. The table demonstrates that treated and matched hospitals are broadly similar across infrastructure, staffing, surgical volume, and patient population characteristics.

Table A.8: Surgeon Gender and Patient Outcomes:
Alternative "At-risk" Subsamples

	Mortality risk above percentile:			
	80th	85th	95th	98th
Dependent variable:	(1)	(2)	(3)	(4)
Care Quality				
Major complication	-0.001	-0.001	-0.001	0.001
	[0.002]	[0.002]	[0.002]	[0.002]
Mortality	0.000	0.000	0.001	0.000
	[0.001]	[0.001]	[0.002]	[0.002]
30-day readmission	-0.000	-0.001	-0.000	-0.001
	[0.001]	[0.001]	[0.001]	[0.002]
Care Quantity				
Length of stay	0.024	0.020	0.009	0.010
	[0.045]	[0.048]	[0.065]	[0.076]
Total spending	-21.933	-18.279	4.782	41.230
	[27.469]	[30.242]	[38.944]	[49.764]
Diagnostic spending	-1.839	-2.675	-2.709	-1.492
	[2.809]	[2.647]	[2.407]	[2.462]
Therapeutic spending (outside OR)	0.256	0.113	5.037	7.494
	[2.724]	[2.815]	[3.481]	[4.849]
Therapeutic spending (in OR)	-17.245	-14.071	6.281	11.764
	[10.733]	[11.893]	[9.634]	[8.866]
Spending on devices/supplies	-10.238	-8.262	-12.961	1.527
	[11.274]	[11.139]	[7.047]*	[5.390]
Room and board spending	9.054	10.264	14.570	26.531
	[19.298]	[21.812]	[31.624]	[43.190]

Notes. This table replicates the analysis in Table 6 using alternative definitions of the "at-risk" sample. Instead of restricting to patients in the top decile of the predicted mortality distribution, we consider patients above the 80th, 85th, 95th, and 98th percentiles of predicted risk, as indicated in columns 1–4. Mortality risk is estimated using the random forest model described in Appendix C. Each row reports results from a separate regression. All specifications include hospital-by-day-of-week, hospital-by-month-by-year, and surgery-by-specialty fixed effects, and spending outcomes are expressed in constant August 2024 Brazilian Reais. Standard errors are clustered at the hospital level.
* $p < 0.1$;** $p < 0.05$;*** $p < 0.01$.

B Data Sources and Construction

Our analysis uses comprehensive administrative records that cover the full universe of hospital admissions within the Brazil's Unified Health System from 2008 through the first quarter of 2024. These data originate from multiple registries maintained by the Ministry of Health, the Regional Medical Council, and the Brazilian Medical Association. The final dataset contains detailed information on physicians' educational background, professional roles, and patient interactions. We next describe these data in detail.

B.1 Provider Data

We have access to background information on providers from the *Cadastro Nacional de Estabelecimentos de Saúde* (CNES). It is a rich source of data collected monthly since 2005 that catalogs every health facility in the country, whether public, private, or philanthropic, integrated with or independent of the Unified Health System. The registry serves as the foundational cadastral base for critical health information systems on outpatient and inpatient care, facilitating comprehensive health planning and effective management at all administrative levels. A major strength of these data is their universal nature and high-frequency observations, with minimal underreporting. It includes detailed attributes of each facility, such as geographic location, administrative classification (public, private for-profit, private non-profit), size indicators (number of beds, number of operating rooms), availability of specialized equipment, and the range of medical specialties offered. Additionally, CNES records information on the workforce at each establishment, including counts of medical professionals and other healthcare workers, their qualifications, and work intensity. We track doctors over time using unique individual identifiers, which allow us to merge these data with other administrative sources.

Other background information on physicians comes from the Federal Council of Medicine (CFM), Brazilian Medical Association (AMB), and National Medical Residency Commission (CNRM). The CFM is responsible for issuing and updating physicians' professional licenses, the AMB maintains records of specialty certifications granted by its affiliated medical societies, and the CNRM oversees the accreditation of residency programs and the registration of residents. The primary purpose of these registers is to ensure compliance with regulatory standards for medical practice. These datasets collectively provide detailed information on physicians' demographic characteristics, educational trajectories, residency training, and professional qualifications. They include records on medical school attended, year of graduation, residency programs completed, and board certifications, as well as other credentials relevant to clinical practice. Because they are maintained by official regulatory and professional bodies, these registers offer nationwide coverage and high reliability, enabling consistent tracking of physicians' careers over time and across different health care settings.

While the administrative registers offer extensive professional information, they do not explicitly record physicians' gender. To address this issue, we supplement our data with records from the *Instituto Brasileiro de Geografia e Estatística* (IBGE), which allows us to assign gender based on physicians' first names. The IBGE provides the dataset *Nomes do Brasil*, which contains a comprehensive list of first names recorded in the 2010 Brazilian census, along with gender-specific frequencies for each name. We classify a first name as male or female if at least 50 percent of individuals with that name identify with the corresponding gender. Given that first names in Brazil are almost exclusively linked to a single gender, this approach produces a gender assignment that is virtually error-free.

B.2 Inpatient Procedure Registries

For each patient admitted to a SUS-affiliated hospital, we observe the procedures completed by health care providers the from the *Serviços Profissionais files of the Sistema de Informações Hospitalares* (SIH-SP). The SIH-SP covers the universe of hospital procedures provided through the Unified Health System. All procedures have an identifier code created by the Ministry of Health to standardize service reporting, with multiple "layers" describing the procedures at different levels of detail. There exists one main procedure and there could be several secondary procedures per patient. The main procedure corresponds to the basic reason for the treatment, whereas the secondary procedures complement the main one. For example, in the case of a patient admitted for a hip replacement, the hip replacement surgery would be recorded as the main procedure, while secondary procedures might include anesthesia administration, blood transfusion, or postoperative physiotherapy sessions. The data are organized at the patient–procedure level, and for each procedure we observe the unique identifier of the health care professional who performed it.

Additional data on patients are drawn from the *Reduzida* files of the *Sistema de Informações Hospitalares* (SIH-RD). These files include information on date of birth, date of admission, date of discharge, diagnosis codes, race, sex, a hospital identifier, and an admission code that allows us to match these data to the inpatient procedure files described above. The diagnosis codes are divided into a primary and up to nine secondary diagnoses. The former is the health condition responsible for the admission and the latter are diseases that coexist at the time of the admission or develop during the hospital stay. We use the secondary diagnoses to create controls for comorbidities that commonly affect patients, such as hypertension, diabetes, chronic ischemic heart disease, kidney disease, among others. Key to our research design in Section 5 is that these data provide information on whether the patient was admitted to the hospital because of emergency health conditions. We link these patient-level data to the physician-procedure dataset using a unique patient case identifier available in both files. We successfully matched 100 percent of all cases.

B.3 30-Day Readmissions

In Section 5, we examine 30-day readmissions as an additional outcome of care quality. The inpatient datasets do not directly provide patient identifiers that allow tracking of individuals across multiple hospitalizations. Consequently, the occurrence of readmissions within a defined time window must be inferred. To construct a 30-day readmission measure, we implement a probabilistic matching procedure using the *Reduzida* files of the *Sistema de Informações Hospitalares* (SIH-RD). We first restrict the analysis to admissions in which the patient was discharged alive, since in-hospital deaths mechanically preclude readmission. For each index hospitalization, we search for any subsequent admission within 30 calendar days after the discharge date.

Because the SIH data do not contain a unique patient identifier, we match hospitalizations across time using an exact combination of (i) five-digit postal code of residence, (ii) race, (iii) exact date of birth, and (iv) gender. These variables are consistently reported and provide sufficient discriminatory power to minimize false positives in the linkage process.

A new admission is classified as a readmission if the matched record occurs in a hospital anywhere in the country within the 30-day window, regardless of whether the readmitting hospital is the same as the index hospital. In the rare case of multiple eligible admissions within the 30-day window, we retain only the first as the readmission event. This approach yields a harmonized, patient-level measure of short-term readmissions that can be consistently applied throughout the study period, while ensuring comparability across facilities and geographic regions.

C Mortality Risk Prediction Using Random Forests

This appendix outlines how we generated the mortality predictions used in our analysis. We predict mortality risk for each patient using a random forest model trained on a set of patient characteristics, including age, race, gender, three-digit level primary diagnosis, and admitting hospital. The outcome variable is an indicator for patients who died during their hospital stay.

Random forests are well-suited for this type of prediction task, especially in settings with complex interactions and nonlinear relationships among predictors. A random forest builds an ensemble of decision trees, each trained on a different bootstrap sample of the data. At each node of a tree, the algorithm considers a random subset of covariates and selects the one that best splits the data based on a chosen criterion. This randomness, both in the sampling of observations and in the selection of covariates, helps reduce overfitting and ensures that individual trees are de-correlated, making the aggregate prediction more robust. For a given input, each tree produces a prediction, and the forest's final output is obtained by averaging these predictions. By relying on data-driven recursive partitioning rather than pre-specified functional forms, the model can naturally capture higher-order interactions and nonlinear effects without the need for manual specification. Empirical comparisons show that random forests often outperform logistic regression in predictive accuracy (Couronné et al., 2018).

We constructed two versions of the model, one for the full sample of approximately 68 million patients and another restricted to the subsample of about 41 million emergency admissions. The former is used in Section 3 to evaluate whether female and male surgeons are in general assigned different patient profiles. The latter is used in Section 5 to identify patients on the margin of survival in a setting with limited scope for discretion in the assignment of surgeons to cases. Both models use 500 trees and a 50 percent subsampling rate in each bootstrap iteration, which offers a conservative and computationally efficient compromise. As discussed in Breiman (2001), prediction error tends to stabilize with a sufficiently large number of trees, and adding more trees typically improves stability without increasing overfitting. Mortality probabilities are computed using the out-of-bag prediction method, which yields unbiased estimates without requiring a separate validation set.

References

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