Equal Entry, Unequal Careers: Women in Medicine in the Long Run

Fatima Aqeel*, Hadia Majid†

August 30, 2026

Abstract

Twenty years after women's representation in an occupation increases, what does gender integration in the occupation look like? We study a 1992 reform that eliminated gender quotas in Pakistani medical schools and substantially increased the share of female doctors. Combining original data on over 700 doctors with an exposure-based triple-differences design, we find that greater exposure to the reform widened several gender gaps among post-reform cohorts. Female doctors in more exposed units allocate relatively more time to teaching than to clinical practice, have more fragmented consultation schedules, and charge lower fees for outpatient visits than their male peers and comparison groups. While patient volume consulted with does not differ, consultation style differs, with female doctors reporting spending less time with each patient. We show that gaps are concentrated among mid-career

*Corresponding author. Department of Economics, College of the Holy Cross. Email: faqeel@holycross.edu. We are indebted to Gallup Pakistan for data collection, and are deeply grateful to study participants for their time and cooperation. We thank seminar discussants at Colgate University, Boston University, Lahore University of Management Sciences (LUMS) IDEAS conference and the Lahore School of Economics CREB conference for feedback. Special thanks to Sam Bazzi, Dilip Mookherjee, Daniele Paserman, Carolina Castilla, Nidhiya Menon for comments. Eraj Tufail Arbab, Leena Salman Butt, Samay Gupta, Asra Waseem Qureshi, Adeen Saeed, Hadia Shehzad, and Yuexi Zhang provided superb research assistance. The Colgate University Research Council generously provided data collection grant support. Figures may be printed in black and white. Declarations of interest: none. The data were collected by the authors and are available upon request.

†The Lahore University of Management Sciences.

physicians, when work and family demands may be particularly salient, and that hours spent practicing are positively linked with fee pricing. Thus, even among women who remain in medicine, increased access produced doctors who are unequal in their practice and contributions.

Keywords: Development; economics of gender.

JEL Classifications: O10; J16

1 Introduction

Women's access to education has drastically improved over the past few decades (Medley, 2016), with cases such as Title IX, the inclusion of women in US military academies (JUSTIA US Supreme Court, 1996), and changes in testing policies, such as in Japan and Singapore, as examples. While these landmark policies increased women's access to schooling, there is limited follow-up on their long-run effects on women's integration into occupations.

Moreover, existing research suggests that as women enter an occupation, average pay and prestige may decrease (Harris, 2022; Goldin, 2014; AAMC, 2018). Cases leading to greater female access therefore also provide an opportunity to study the consequences of occupational feminization, if any exist.

We study long-run gender integration and occupational changes following a large, plausibly exogenous increase in women's entry in medicine. In 1992, Pakistani courts struck a 20% cap on women's seats in medical schools, ruling it discriminatory and replacing it with merit-based admissions. The female share of medical students increased immediately by 10 percentage points (p.p), and overall from 28% to 55% (Aqeel, 2024). We leverage this setting to study two questions: if a policy shock allows more women into higher education, how integrated are women in the associated occupations in 20 years? And, second, does the occupation change overall as women's representation increases?

Our empirical strategy exploits uneven increases after the reform in the female share of doctors across specialties and provinces. We show that specialty-province units with a *lower* pre-reform female share experienced a 0.6 p.p. *greater* increase in the female share following the reform than units that started at a higher baseline, potentially due to catch-up effects. We, therefore, use the pre-reform female share of doctors by specialty and province ("exposure") in an exposure-based triple-differences design. Specifically, we compare post-reform male and female doctors in units with higher exposure to the reform, while differencing out gaps in units less exposed to the reform and pre-existing gender differences. Event studies provide support for the parallel trends assumption. Further, the pre-reform female share is predetermined, and we argue, plausibly exogenous conditional on specialty and province fixed effects.

Our main data consists of a survey of over 700 randomly sampled Pakistani doctors that we collected between 2022 and 2026. We focus on two sets of outcomes: first, outpatient consultation fees and salaries, which captures the price of doctors' services[1] and physician pay, and second, time use at work, such as hours spent practicing. Medical doctor registration data are used to calculate the female share of doctors in specialty-province units and to examine doctor retention, and finally, we benchmark selected findings using data from online healthcare portals.

We find, first, increased gender gaps in time use at work. Surviving pre-reform female doctors teach a greater number of hours than male doctors. Among post-reform cohorts in more exposed units, however, this teaching gap is about 40% higher. Post-reform female doctors also allocate more time to teaching relative to clinical practice relative to their male peers and comparison groups, with gaps being greatest among mid-career post-reform physicians. These findings are broadly consistent with labor theories of women and men taking on different roles within occupations (Blau and Winkler, 2017), and inequalities within rather than between occupations.

Second, from a healthcare provision stand-point, the supply of doctors' time is structured differently. Post-reform female doctors in more exposed units do not report consulting with fewer patients. Instead, they report spending less time with each patient, by about 5 fewer minutes, or 50%, in the most exposed units than their male counterparts and comparison groups. They are also more likely to have scheduled breaks between consultation hours which results in relatively fragmented consultation schedules.

Less time spent with patients may reflect efficiency, or the need to balance competing time demands, among other reasons. However, the change in the female-male fee gap suggests a decreasing value of post-reform female doctors' service, which weighs against a pure efficiency interpretation. Surviving pre-reform female doctors in more exposed units enjoy a *fee premium* over pre-reform men, with an 8.3 log point higher fee for every 1 p.p. of exposure. In contrast, among post-reform cohorts, a 1 p.p. increase in exposure narrows the female-to-male fee ratio by 5.8% relative to the comparison groups, so that the initial differential becomes smaller by 72%.[2] Overall, post-reform female doctors in more exposed units spent more time teaching, maintained

[1] For other studies that use fee as a metric of the pricing and gender gaps in physicians' services, see: Feldstein (1970); Merali et al. (2021); La Brooy et al. (2025); Richardson et al. (2006) as examples.

[2] Salary gaps change insignificantly.

patient volume but less time facing each patients, and experienced a decrease in pay relative to men, the pre-reform gap and gaps in less exposed units. Male and female doctors receive similarly costly medical training in medical college and beyond. Our results indicate that while the reform increased the female share of doctors, the doctors produced continue to differ in practice and the areas of medicine to which they contribute.

In contrast, we find limited changes in specialty-province units overall as the female share of doctors rises, which is consistent with female doctors adapting to the field rather than the field adapting to them. Doctors in more exposed units are not significantly different in average fee or output than those in less exposed units in the long run. This is unlike recent findings which show female entry reduces average pay in a field (Harris, 2022; Oldenziel, 2001; Li, 2024), and may be because the female share of doctors remained a minority, at one-third of doctors in the registration data.

We examine two main mechanisms for our findings. The first is of time-intensive "greedy" jobs (Goldin, 2014) creating gender gaps at certain career stages, even as more women enter these fields. Consistent with this explanation, fee gaps are largest among mid-career post-reform women, graduating between 2007-2014, who also occupy lower academic positions relative to their male peers and the comparison groups. Meanwhile, gaps are minimal among the older post-reform cohorts. Additionally, these fee gaps are greatest in units characterized by greater median physician hours (Center for Studying Health System Change, 2024), and fee is correspondingly positively related to practice hours. Together, these findings are consistent with more exposed units being especially punitive for mid-career women, who may face dual time burdens.[3]

The second mechanism is of composition changes. Doctor registration data suggests that a weak retention gap between male and female doctors emerges around 5-10 years after obtaining a license in more exposed units relative to the comparison groups. But there is no significant gap among those who remain fifteen years after license issue. Next, using Pakistan's Labor Force Surveys (1990-2012) we show that gender gaps in hours worked and the probability of working among pre-reform cohorts, measured contemporaneously when they were young, are similar to

[3]While we do not have panel data on individuals and all their children, we collected data on the individual's age when their first child is born. The average age at which a first-child was born is 30 years in our sample, while the average age of mid-level post-reform women is 38 years.

the gaps we observe among survivors of these cohorts in our survey. Second, there could also be selection changes at entry into specialty-province units. Survey data suggest that women's selection worsened slightly but insignificantly relative to men in more exposed units and relative to the comparison groups. When controlling for observable changes in quality, fee gaps worsen by 0.05 log points and teaching gaps remain similar. Nevertheless, we cannot fully rule out the effect of selection changes, and interpret our findings as the long-term gender gap among cohorts differently exposed to the reform, while examining evidence for each of the main mechanisms.

We contribute to the literature in the following ways. Our first contribution is to the question of whether equalizing access to education is sufficient in eliminating labor-market gaps for historically excluded groups, including women.[4] The persistence of gender differences we find despite increased entry suggests a significant role of occupational organization (Bolotnyy and Emanuel, 2022; Adams et al., 2025; Kumar et al., 2026; Le Barbanchon et al., 2021; Collis and Van Effenterre, 2025; Folke and Rickne, 2022), which can create segmentation and inefficiency (Cain, 1976; Deakin et al., 2013; Hsieh et al., 2019). One specific feature of occupational organizational is of time-greedy jobs (Goldin, 2014; Cortés and Pan, 2023; Wasserman, 2017; Kim and Moser, 2025; Bertrand et al., 2010; Cortés and Pan, 2019; Harrington and Kahn, 2026; Gicheva, 2013; Deschacht et al., 2025), where the inability to work certain hours leads to a penalty. This paper extends existing literature by showing that even when a policy shock increases women's representation, in the long-term, organizational features may contribute to outcome gaps. We also provide evidence from high-skilled workers and the price of service provision in the medical profession.

Second, we contribute to research on how changes in gender composition affect individuals' outcomes. Similar to Calkins et al. (2023), we find that women's outcome gaps are worse in mixed settings than in more female-dominated ones, though we study work rather than college outcomes. Existing work finds that greater exposure to women need not worsen male outcomes (Greenberg et al., 2026) and can improve women's access to senior positions and influence, but

[4]There is vast research on affirmative action policies, which favor an underprivileged group. See for example, Carlana et al. (2024), Zhao et al. (2025) as some recent examples. Though our paper shares some similarities with this literature, in our context, a previous policy that favored men was changed to equal access for men and women.

that these gains are mediated by female-friendly firms and female leaders (Hampole et al., 2026; Karpowitz et al., 2024). We show that absent any major mediating factors, a country-wide, policy-driven increase in female representation preserves or increases gender gaps.

Third, we contribute to research on how an occupation itself may change due to "feminizing", as its share of women grows. Existing research suggests that occupations can experience pay declines or a reduction in male employment as they become more female-dominated (Harris, 2022; Oldenziel, 2001; Li, 2024). This may reflect falling prestige (Goldin, 2014) or a shift toward non-pecuniary benefits (Lee and Thompson, 2019; Wiswall and Zafar, 2018; Mas and Pallais, 2017). In contrast, we do not find changes in units overall as female representation rises, which may be because female representation remained below a tipping point (Pan, 2015).

The remaining paper is structured as follows. In section 2 we describe the policy change in medical schools and the structure of healthcare in Pakistan. Section 3 describes our data and identification strategy. Section 4 discusses our main findings, followed by robustness checks in section 5. Finally, we present our conclusions in section 6.

2 Background

Health care and medical school in Pakistan: As of 2022, healthcare in Pakistan accounted for approximately 1.2 percent of GDP, reflecting longstanding constraints in health system capacity. The number of registered doctors has increased from 48 doctors at Pakistan's independence in 1947 to over 274,000 doctors in 2022 according to data from the Pakistan Medical Dental Council (PMDC). Admission to medical school, previously based solely on secondary academic performance, has since 2000 been determined by a weighted combination of entrance exam scores, secondary school grades, and middle school grades. Following five years of medical training, graduates may register with an M.B.B.S. degree[5] or pursue postgraduate training through a residency ("house jobs") and fellowships. While medical colleges are required to provide paid house jobs to their own graduates, these positions are not guaranteed for external

[5]M.B.B.S stands for Bachelor of Medicine, Bachelor of Surgery. It is equivalent to being a Doctor of Medicine.

applicants. Placement into medical residency is primarily based on exit exams (College of Physicians and Surgeons Pakistan, or FCPS, exams) (Aga Khan University Hospital (AKUH), 2026). The PMDC issues a temporary 1-year residency license to medical graduates after medical school, and a permanent license after successful completion of a residency program.

The Policy Change: The quota for women in medical school was started in 1948 to reserve 20 percent of medical college seats for women, and 80 percent for men. Over time, however, as more women applied to medical college, the number of female applicants outgrew female reserved seats (Masood, 2019).

The quota was removed after a 1989 court case when a group of female applicants to medical school, led by Shrin Munir, alleged that they were denied admission despite scoring higher than several men. A court ruling decided that the old admissions policy was unconstitutional, stating:

"The selection procedures violated [sections of] the Pakistani Constitution prohibiting discrimination [...] on the basis of sex" - Shrin Munir and others v. Government of Punjab through Secretary Health, Lahore and another (PLD 1990 SC 295).[6]

A few months later, a similar case was won in the province of Sindh.[7] Subsequently, the admissions policy was changed in 1991 to being based purely on merit. Women and men had an equal opportunity of being admitted to medical school for all cohorts starting in 1992. The first cohort affected by the reform is expected to have graduated from medical school and registered for their initial license in about 1996-1997.

The reform remained controversial for many years. In 2014, there were efforts to limit female students' admission to medical schools once again because female doctors are more likely to quit the labor force than male doctors (The Express Tribune, 2021), but this failed in courts for constitutional reasons.

Health Care Delivery: Healthcare is primarily delivered through public and private hospitals. Although the private sector comprises a smaller share of total facilities (35 percent of hospitals in

[6]Available through: https://sahsol.lums.edu.pk/node/17087
[7]https://pakistanconstitutionlaw.com/p-l-d-1990-karachi-342/

2022), it serves a disproportionately larger share of the urban population (Aqeel, 2024). Public hospitals (also known as government hospitals) are government funded and provide low-cost or free care (Khalid et al., 2021). They typically experience high patient-loads (Hussain et al., 2019). In some cases, a "public-private partnership" consists of the private sector operating a public hospital to improve hospital revenues (Pasha and Nasar, 2003).

One other distinction in this context is of teaching hospitals. Both public and private hospitals can be teaching hospitals, which are hospitals with their own affiliated medical schools and residency programs. Doctors hold academic titles, and also produce academic research (Painter, 2017).

Fee and Salary Setting: In public hospitals, fees are fixed by government regulations and typically vary by hospital type (e.g., tertiary care, district-level) and specialization. Outpatient services are often free or require a small fee, such as to consult with specialists. However, in practice, fees can vary due to difficulty in fee regulation (Dawn Newspaper, 2011). Further, certain arrangements allow doctors at government hospitals to charge a fee, such as by allowing doctors to hold private clinics at their government hospital in the evenings ("institutionalized private practice") (The News, 2024).[8] Private hospitals, on the other hand, offer doctors more autonomy. In large private hospitals, senior specialists may retain flexibility to set their own rates. In small private clinics, doctors typically have full discretion in setting their fees. Similarly, fees at a secondary practice (if a doctor has one) are often set autonomously by doctors, and are market-driven based on factors such as seniority and specialty. In many cases, part of the fee earned is taken by the hospital while the remaining is allotted to the doctor.

Salaries at public hospitals are based on a basic pay scale that sets a salary range for public officers of different salary grades. For example, medical doctors starting in the public health system could start in Grade 17, and in 2017 their salaries would range from Rs. 25,000 to Rs. 65,000 (90 USD to 234 USD in 2026).[9] Salaries set by private hospitals also depend on factors such as doctors' specialty, experience and location, but, as with fee, are less regulated.[10]

[8] A second is the public-private partnership model described earlier.

[9] This information was collected from the Cabinet Division, Government of Pakistan. See for example: https://cabinet.gov.pk/SiteImage/Downloads/basic-pay-scales-civil-servant.pdf

[10] This information was obtained from Gallup Pakistan, which has previously worked extensively with

3 Data and Empirical Framework

3.1 Data

We collected three rounds of survey data on doctors across Pakistan for a total of 769 doctors. The surveys were carried out by Gallup Pakistan from May-September, 2022, May-September, 2023 and March-August, 2026. In the first round, we collected data on 420 doctors, in the second round on 268 additional doctors, and 81 additional doctors in the third round. We also included questions about position, secondary school exit test scores and background traits such as birth place and parental education in the second and third rounds, in addition to questions from round 1.

For each survey round, we started with a list of doctor names from doctor registration data available on the website of the PMDC. We randomly sampled doctors from this list and provided Gallup with ten times as many names as our target survey sample size. The Gallup team obtained contact information on these randomly sampled doctors and approached them via email and social media. If there was no response, the doctors were contacted by phone or in person. Eighty percent of the surveys were completed in person, 13 percent on the phone, and the remaining through other means such as email or social media. The average response rate averaged 85% across rounds.

In table A1, we present summary statistics on our full sample. The average doctor in our sample is 45 years old. About 79% of doctors were born in an urban area and 93% practice medicine in an urban area. This partly reflects the fact that specialists practice mainly in urban areas (80% of specialists and 87% of female specialists in PMDC registration data are registered in urban areas), but may also reflect selection into our survey. Fifty-one percent of the sample consists of female doctors, about a third is married, and 45% practice in the public sector. About 28% of doctors also have a secondary private practice.[11] Ninety-four percent of doctors in our sample were practicing at the time of our survey, and among pre-reform cohorts, 86% of female doctors and 93% of male doctors were practicing at the time of the survey. We discuss selective

doctors, and informal discussions with local doctors. It was also verified where possible with newspaper coverage and/or government documents.

[11] This is 23% for female doctors.

attrition below when describing our identification strategy and again in subsection 4.1.3.

The average doctor in our sample spends more hours a week in practice than in research or teaching. Clinical practice includes activities such as direct patient care, charting and documentation, conducting procedures, and reviewing tests. Time use is reported in categorical intervals. Nearly half of doctors report spending 21-40 hours per week in clinical practice, while 53 percent report no teaching and another 30 percent report teaching 1–10 hours per week.

3.2 Identification

Our main interest is in estimating medical practice changes and gender gaps due to an increase in the female share of doctors. An ideal experiment would randomize the change in female share across specialty-province units, while holding doctor selection constant. We could then compare cohorts of female and male doctors across units of varying treatment intensity. In most cases, however, an increase in the female share of doctors occurs because an occupation is becoming welcoming to women. Changes in medical practice itself, therefore, could lead to the expanded entry of women, or a third factor, such as demand for certain types of workers, may affect both (Black and Juhn, 2000).

We address these endogeneity concerns by leveraging an admissions school reform and variation in the extent of exposure across units to the reform. As mentioned earlier, a Pakistani medical admissions school policy in 1992 replaced a quota for female students with merit-based admissions. Subsequently, the proportion of female students who graduated from medical schools in about 1997 rose by 10 percentage points within a year (Aqeel, 2024). We use 1997 as the year in which medical doctors affected by the reform are expected to have graduated and entered the workforce, after 5 years of medical school. In section 5, we test the robustness of our results to varying the timing of when cohorts affected by the reform are expected to enter the workforce.

Figure 1 plots the change in the share of female doctors by the pre-reform share using PMDC data. The pre-reform female share is defined as the share of female doctors among cohorts who registered to practice between 1952 and 1997. The post-reform female share is the female share among cohorts who registered between 1998-2018, and the change in the female share is the difference between the two. The left graph presents shares calculated by specialty (including

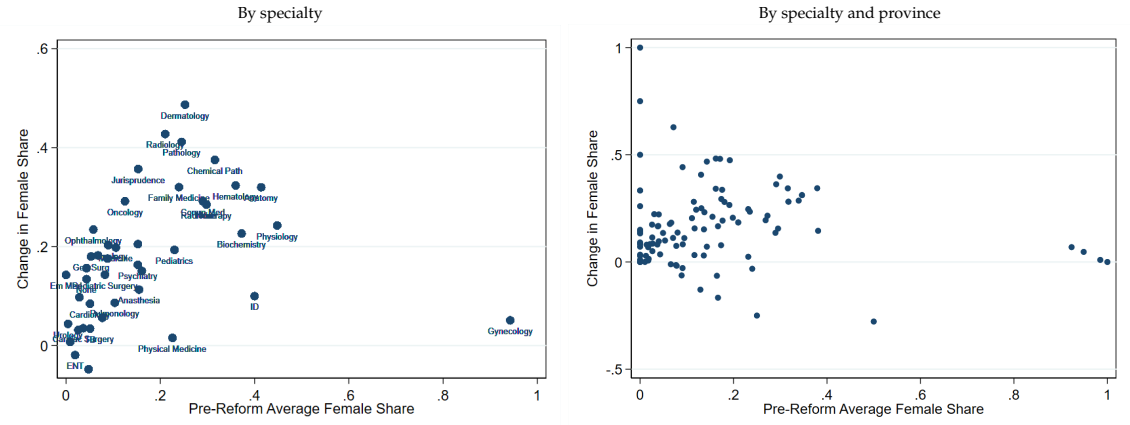
non-specialists)[12] The right graph presents the relationship between change in female share by pre-reform shares at the specialty and province level, and figure 2 further presents a binscatter of this relationship conditional on specialty and province fixed effects. In both specialty-province graphs, specialty and province pairs which started at a lower baseline of female share of doctors experienced a *greater* change in the female share of doctors in the unit.

This relationship is potentially driven by catch-up effects across specialty-province pairs. Before the reform, there was variation in the female share of doctors in the same specialty across provinces, and within the same province, across specialties. In specialty-province units with a low pre-reform female share, there was a greater difference in the female share among outgoing and incoming cohorts of doctors. We therefore refer to units with a lower female share of doctors before the reform "high exposure" units because these are units where the female share of doctors changed to a greater extent.

A first-stage estimate from regressing the change in female share on the pre-existing female share and conditional on specialty, province and graduation year fixed effects shows that a 1 percent increase in the pre-existing female share by specialty and province is linked with a 0.6 percentage point lower change in female share after the reform (table 1, column 1). The coefficient after removing the top 2% of outliers in the pre-reform female share is relatively stable (table 1, column 2).

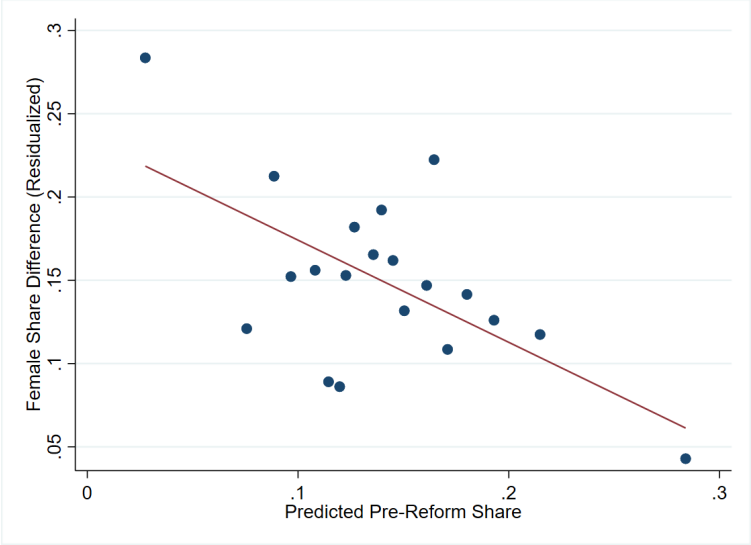
[12]The outlier specialty is Gynecology, where the female share before the reform was 95 percent and increased by only about 4 percentage points after.

Figure 1: Change in the Female Share of Specialties



Note: The graphs plots the relationship between the pre-reform average share of female specialists and the change in female share of specialists by specialty in the left graph and by specialty and province on the right graph. Data are from the PMC Medical Doctor Database between 1952 and 2018.

Figure 2: Change in Female Share against Pre-Reform Share, Partialling out Controls



Note: The graph plots a binscatter between the pre-reform female share and change in the female share of medical doctors, controlling for specialty and province fixed effects. Data are from the PMC Medical Doctor Database between 1952 and 2018.

Our main empirical strategy uses variation in the pre-reform share of female doctors by specialty and province in an intent-to-treat (ITT), or reduced form, triple-differences framework.[13] To examine gender gaps we estimate:

[13]For recent papers with pre-existing conditions proxying for the exposure to a reform see Topalova and Khandelwal (2011) and Harju et al. (2021).

	Δ Female Share	Δ Female Share
Pre-Reform Female Share	-0.61**	-0.60**
	(0.25)	(0.26)
Observations	132	129
Mean DV	0.15	0.15
F-statistic	16.9	16.3
R^2	0.53	0.53

Note: This table presents coefficients from a regression of the change in female share on the pre-reform share, or, "exposure" (calculated by province and specialty), controlling for specialty and province fixed effects. Outliers above the 98th percentile of the pre-reform share are excluded in the second column. Standard errors are clustered by province and specialty. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 1: First-Stage

$$\begin{aligned} Y_{ispt} = \alpha_0 + \alpha_1(\text{Exposure}_{sp} \times \text{Post}_t \times \text{Female}_i) + \alpha_2(\text{Post}_t \times \text{Female}_i) \\ +\alpha_3(\text{Exposure}_{sp} \times \text{Female}_i) + \alpha_4(\text{Exposure}_{sp} \times \text{Post}_t) + \\ \alpha_5 X_{ispt} + \alpha_s + \alpha_t + \alpha_p + v_{ispt} \end{aligned} \tag{1}$$

In equation 1, Y_{ispt} represents the outcome for medical doctor i in medical specialty s, province p and graduation cohort t.[14] Since Gynecology is an outlier in many ways, we exclude Gynecologists from our main analysis.[15] X_{ispt} includes controls for individual age and age-squared. α_t are graduation year fixed effects, α_s are specialty fixed effects, α_p are province fixed effects for Pakistan's four main provinces,[16] and v_{ispt} is an individual-level disturbance term. Standard errors are clustered by medical specialty and province. The coefficient, α_1 compares the gap in outcomes between post-reform female doctors and male doctors in higher capacity units

[14]To reduce the number of specialty fixed effects estimated given our sample size, we combine specialties into similar groups of specialties for a total of 20 specialty groups (from hereon, "specialty").

[15]This decreases our regression sample size to just over 500 observations, and of which 36% are female doctors.

[16]These are Sindh, Punjab, Khyber-Pakhtunkhua (NWFP) and Balochistan.

relative to the gap before the reform and the gap in low capacity units.[17]

Next we estimate event studies to investigate the parallel trends assumption:

$$Y_{ispt} = \beta_0 + \sum_{t=-\underline{T}}^{\bar{T}} (\gamma_1^t D_t^k \times Exposure_{sd} \times Female_i) + \gamma X_{ispt} + \alpha_s + \alpha_t + \alpha_p + u_{ispt} \quad (2)$$

Outcomes and fixed effects are defined similarly to in equation 1. The terms in X_{ispt} include age, age-squared and two-way interaction terms.[18] Graduation cohorts in the event study include those who graduated between 1980 and 2024,[19] and are grouped at the ends where the sample is thinner, to an initial interval between 1980-1987 and the final one between 2017-2022, and two year intervals for all graduation years in between. We interpret event study coefficients as reflecting gender gaps across cohorts in highly exposed units relative to less exposed units.

3.3 Identification Assumptions and Threats

Our first identifying assumption is of parallel trends in gender gaps in the absence of the reform. While this is unobserved, we provide suggestive evidence of parallel trends with event studies, described above. We also conduct sensitivity analyses with Honest DiD analysis Rambachan and Roth (2019), which depict how large a violation of parallel trends would have to be to negate our main results.

Second, exposure to the entry of female doctors is determined before the reform, and since specialty is a primary indicator of female entry, we assume that the pre-reform female share is plausibly exogenous conditional on specialty and province fixed effects.[20] One concern that our event studies do not fully rule out is of an omitted factor that is correlated with exposure. For example, high-exposure units may already have been becoming more time-intensive, and since time-intensity affects both female retention and gender gaps in outcomes, these units could be becoming different from low-exposure units over time. We control for trends in key specialty

[17] Note that the pre-reform female share is constant for specialty-province unit pairs, so we do not include specialty-province fixed effects.

[18] Given our sample size in each time interval, we reduce the number of interactions estimated by interacting exposure and female with a single post-reform indicator.

[19] There are very few doctors in pre-1980 cohorts.

[20] To the best of our knowledge, the reform was not anticipated by individuals, and was not widely reported on in newspapers prior to 1992.

traits (median income and median hours worked) to test whether this drives our results. Note also that, since units with a high pre-reform share received a *lower* change in the female share of doctors, any persistent cultural trends are likely to bias our estimates toward zero.

Third, since the reform did not have a staggered implementation, our identification is not based on a staggered differences-in-differences design. However, the reform induced variation in treatment intensity (the change in the female share of doctors), which we proxy for using the pre-reform female share. Based on recent work on difference-in-differences with continuous treatment, our analysis can recover the average treatment effect on the treated (ATT), but not the average treatment effect (ATE) which involves a strict parallel trends assumption (Callaway et al., 2021).[21]

Fourth, while our survey data have several self-reported responses, in section 5 we compare key variables, such as out-patient fees, with those available for comparable doctors on online platforms Oladoc.pk (similar to Zocdoc.com) and Marham.pk. These websites are used by doctors to book actual patient consultations for practice at hospitals, private practice, and for telehealth appointments. The information on these websites is therefore likely to be verified by sources such as hospitals themselves. While doctors available on online platforms are also self-selected, using multiple data sources can add to the strength of the results.

Fifth, while the doctor names we provided to the survey team were randomly selected, the doctors who responded to our survey were not random. We report differences between survey respondents and the full set of registration doctors in table A2. While gender and graduation year (which we are most concerned about, since experience can affect fee) are balanced, there are differences in the coverage of provinces and specialties between our sample and registration data. As noted above, we include province and specialty fixed effects in all specifications. We also address concerns of selection into our survey using inverse probability weights (IPW) (Wooldridge, 2007), described in section 5.

Finally, selective attrition will change the composition of our sample, for example through the selection of doctors that remain in the labor force, and the selection of doctors who enter a

[21] The assumption is that units that receive a different dose would have the same outcome if they received *any* other dose (and not just no dose as in the control group).

specialty-province unit. For the former, we examine selective attrition using registration data, which provide the full-count of doctor registrations and the years that their license is valid for. For the latter, we directly examine selection for pre-determined characteristics that we collected data on, such as quality of medical school attended, parental education, and secondary school exit exam scores, and show these did not change significantly. Separately, we show that our results are robust to controlling for background traits we collected data on. Nevertheless, we cannot rule out the role of selection changes entirely, and we interpret our findings as the current gender gap among doctors, comparing cohorts differentially exposed to the reform.

4 Results

Hours Worked and Time Use: Time use indicates how individuals may allocate time between work activities. The clearest difference in time use among post-reform female doctors in more exposed fields is not a reduction in labor supply, but a change in how working time is allocated. Among pre-reform cohorts, female doctors spend more hours teaching than their male counterparts (table 2, column 5). Among post-reform cohorts, however, the teaching gap is almost 40% higher in more exposed relative to less exposed units, as shown by the triple differences coefficient. The event study in figure 3 shows an increase in teaching hour gaps among post-reform cohorts while pre-reform coefficients are statistically insignificant. This result is also robust to the Honest DiD sensitivity analysis of Rambachan and Roth (2019) (figure A3). The estimated effect remains statistically different from zero when post-treatment violations of parallel trends are allowed to be as much as 1.4 times the magnitude of the pre-treatment violations.

The higher allocation of time to teaching may partially crowd out of hours spent practicing, and partially reflect an increase in total hours worked. Practice hours are lower (column 4), though the estimate is small and statistically insignificant (table 2 column 4), while the event study shows mostly negative but imprecise post-reform effects (figure A1, panel a). Nevertheless, the ratio of teaching to practice hours rises after the reform among post-reform cohorts and relative to the comparison groups (figure 3, panel b).[22] On the other hand, while the coefficient on the maximum

[22]Because weekly teaching and clinical-practice hours are reported in intervals, we construct an

number of hours worked[23] is positive, it is statistically insignificant (table 2 column 3 and figure A2). [24]

Importantly, the reallocation of time does not appear to reduce reported patient volume. Post-reform women in more exposed units report consulting with similar out-patients per day as their counterparts and comparison groups (table 2, column 1). Instead, the adjustment appears to occur partly within patient-facing time: women report spending less time with each out-patient relative to men and comparison groups (table 2, column 2). The triple differences coefficient indicates a decrease in the interval of time spent which corresponds to about 5 minutes fewer with each patient.[25]

A second adjustment occurs in the organization of the working day. The consultation timings for post-reform women in more exposed units are more likely to include scheduled breaks, and therefore be fragmented (figure 4, panel c). Combined with results on relatively unchanged maximum hours worked and out-patient volume, this pattern does not suggest withdrawal from work. Rather, it is consistent with a different organization of working time.

Gender Gaps in Fee and Salary Fee charged per out-patient is a key indicator of the price of a doctor's out-patient services, while salary differences represent the gender pay gap. In table 3, we examine fee charged per out-patient in Pakistani rupees (Rs.), the log of consultation fee at a primary practice (column 2) and secondary practice (column 5), the likelihood of no fee being charged (as is common in large public or charitable hospitals), and the salary interval of a doctors' salary in rupees.[26]

approximate teaching-to-practice intensity measure by assigning each interval its upper endpoint (maximum hours practicing and maximum hours teaching each week) and dividing reported teaching hours by reported clinical-practice hours. The measure should therefore be interpreted as an index of the relative allocation of time toward teaching rather than as an exact ratio of hours.

[23]The sum of the upper bounds of teaching, research and practice hours per week intervals.

[24]Research hours increase in figure A1, panel b, but the outcome is omitted from table 2 for brevity. Similarly, the number of consulting hours (measured in hours) do not change significantly, and results are available upon request.

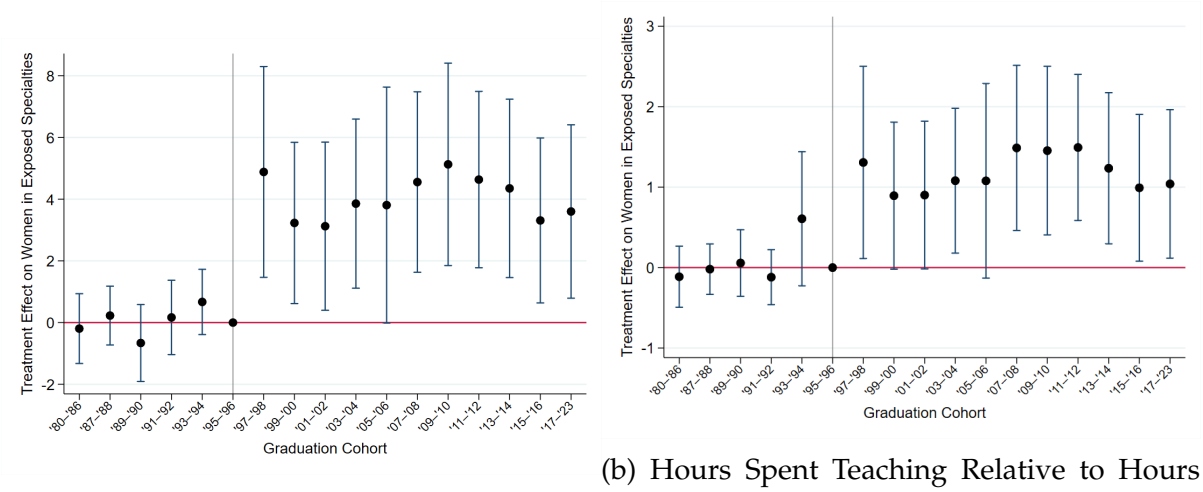
[25]The outcomes are the number of out-patients and time per out-patient at a primary practice only. Our results remain similar when including the number and time per out-patient at a secondary practice, but secondary practices typically do not include teaching responsibilities.

[26]Log(fee) is calculated as log(fee+1). Since the number of doctors that charge no fee is large, we did not trim the ends of the fee distribution. The average consultation fee in our data is Rs. 735 in our overall sample; Rs. 1019 in private practice and Rs. 383 in public practice, which is 4 percent and 1.8 percent of average monthly wages in Pakistan in 2021, respectively (CIEC, 2025).

	Number of OP	Time per OP	Maximum Hours	Practice Hours	Teaching Hours	Teaching /Practice
	(1)	(2)	(3)	(4)	(5)	(6)
Exposure x Female x Post	48.25	-2.24**	56.68	-0.73	4.51***	1.43**
	(35.88)	(1.01)	(37.65)	(1.99)	(1.50)	(0.64)
Female x Post	-39.59	1.29	-49.29	0.68	-4.17***	-1.33**
	(29.18)	(0.89)	(30.61)	(1.56)	(1.20)	(0.50)
Female	-23.55	0.58	39.10	-1.45	2.51*	0.92*
	(27.50)	(1.10)	(31.81)	(1.41)	(1.28)	(0.47)
Exposure x Post	-9.54	-0.17	2.45	1.45	-0.87	-0.27
	(19.36)	(0.87)	(18.12)	(1.08)	(0.77)	(0.29)
Exposure x Female	22.58	0.18	-49.07	1.03	-2.35	-0.83
	(33.87)	(1.23)	(39.53)	(1.81)	(1.52)	(0.59)
Exposure	-18.30	1.94	22.72	-0.74	0.35	0.09
	(31.42)	(1.17)	(26.37)	(1.10)	(1.22)	(0.38)
Observations	511	511	511	511	511	511
Mean DV	30.13	2.47	61.32	2.60	0.66	0.18
R^2	0.2	0.3	0.3	0.3	0.3	0.3

Note: This table presents reports on the number of out-patients (OP) consulted with by doctors per day (column 1), the time spent with each OP (column 2), maximum hours worked (column 3), hours worked practicing each week (column 4), and hours teaching (column 5) and hours teaching divided by hours practicing in column 6. All regressions control for age, age-squared and 2 way interaction terms. Additional fixed effects included are for province, graduation year and medical specialty group. Standard errors are clustered at the province and medical specialty group level. Stars represent statistical significance. Data are from a survey of Pakistani doctors conducted by the authors. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 2: Output



(a) Hours Spent Teaching

(b) Hours Spent Teaching Relative to Hours Spent Practicing

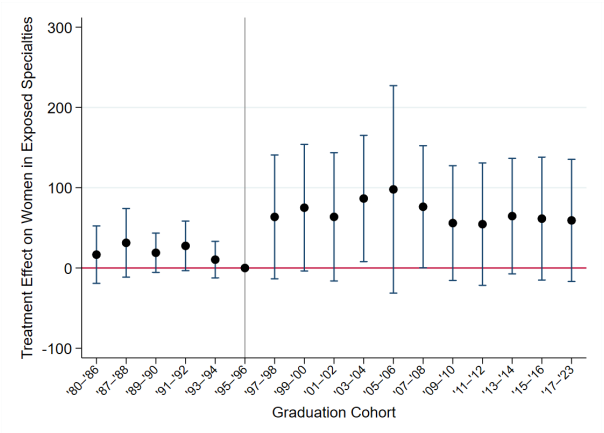
Figure 3: Event Study Estimates for Teaching Hours

Note: The graph plots an event study of hours spent teaching per week (a) and hours spent teaching relative to hours spent practicing (b) in intervals. Regressions include specialty, province and year fixed effects, and standard errors are clustered by specialty and province. The omitted period is 1995-1996. Data are from a survey of doctors by the authors.

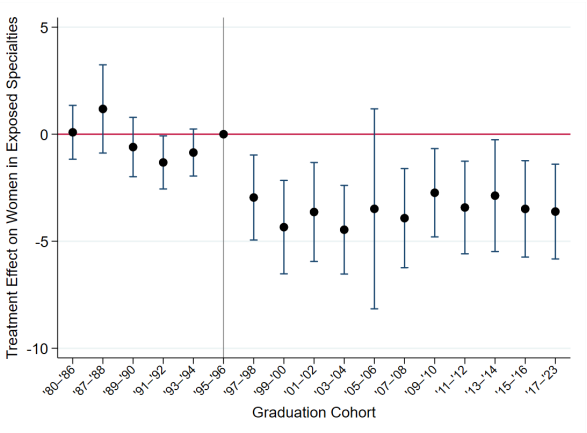
Surviving female pre-reform doctors in more exposed units have a *fee advantage* over their male counterparts (column 2) of the same graduation cohort, specialty, province and age. The "female premium" for these pre-reform cohorts is 8.2 log points relative to male doctors for a 1 p.p increase in exposure.

This relationship is significantly attenuated among post-reform cohorts. A 1 percentage-point increase in exposure reduces female fees relative to male fees by approximately 5.9 log points, relative to the pre-reform gender difference and the gap in less-exposed units, so that the overall gap between female and male fees shrinks by 72%. The point estimate for fees measured in levels is also negative (column 1), although less precise, consistent with the greater skewness in the raw fee distribution.

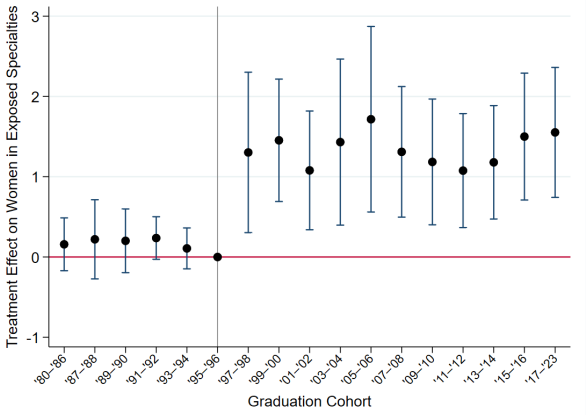
We find less evidence of corresponding changes in other forms of compensation. Salary and fees charged at a secondary practice (columns 4-5 of table 3) do not change significantly for post-reform female doctors in more exposed units relative to the comparison groups. There is, however, suggestive evidence on the extensive margin of fees (table 3, column 3): female doctors in more exposed units are marginally more likely to charge no outpatient fee after the reform.



(a) Number of Out-Patients per Day



(b) Time with each Out-Patient



(c) Scheduled Break Between Consulting Hours

Figure 4: Event Study Estimates for Patient Volume, Consultation Time, and Breaks Between Consulting Hours

Note: The figure plots event study estimates for three outcomes: the number of out-patients consulted with per day at a primary practice, time spent with each out-patient at a primary practice, and the probability of a scheduled break between hours of consulting time. Regressions include specialty, province, and year fixed effects, and standard errors are clustered by specialty and province. The omitted period is 1995-1996. Data are from a survey of doctors by the authors.

One possibility is that these fee differences reflect sorting across practice settings, since fees vary substantially across public and private, large and small, and teaching and non-teaching hospitals. We find little evidence for this explanation. Post-reform female doctors in more exposed units are no more or less likely to work in public, urban, or large practices relative to men, pre-reform gender differences, and less exposed units in our sample (table A3).

Figure 5 presents an event study of the female-male gap in log fee by cohort and relative to less exposed units. Notably, a pre-trend emerges among cohorts who graduated in 1991-1992. We address this in two ways. First, we show using Honest DID plots (Rambachan and Roth, 2019) (figure A3) that the pre-trend would have to be 1.13 times the size of the estimated coefficient to erase the estimated effect. Second, in robustness exercises (table 6), we conduct a placebo test with a reform placed in a pre-reform year, to test if effects started before the expected date of graduation. In both cases, we find limited effects of pre-trends.

Nevertheless, event study coefficients suggest that among post-reform cohorts, fees for female declined relative to their male counterparts in the same specialty, province and age, and relative to gaps in less exposed units.

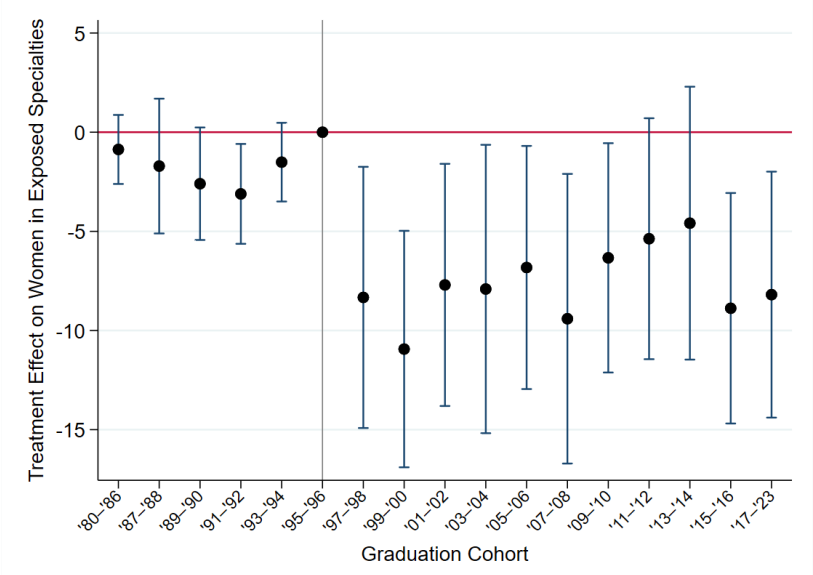


Figure 5: Event Study - Log (Fee)

Note: The graph plots an event study where the outcome is the log of fee. Regressions include specialty, province and year fixed effects, and standard errors are clustered by specialty and province. The omitted period is 1995-1996. Data are from a survey of doctors by the authors.

	Fee (Rs.)	Log(Fee)	Pr(Fee=0)	Log(Fee)	Salary
	Primary Practice			Second Practice	
	(1)	(2)	(3)	(4)	(5)
Exposure x Female x Post	-439.97	-5.93**	0.77*	-448.36	1.89
	(745.81)	(2.62)	(0.40)	(917.13)	(3.97)
Female x Post	-54.92	3.84*	-0.48	43.25	-0.33
	(630.39)	(2.10)	(0.32)	(752.97)	(3.40)
Female	-220.86	-5.44**	0.71**	62.29	-8.70**
	(774.34)	(2.06)	(0.29)	(774.82)	(3.69)
Exposure x Post	-1194.35***	-0.21	-0.18	-183.37	-2.62
	(371.97)	(1.74)	(0.30)	(328.62)	(3.44)
Exposure x Female	988.07	8.27***	-1.07***	384.07	9.31**
	(896.15)	(2.66)	(0.40)	(959.10)	(4.40)
Exposure	1008.80	0.69	0.07	456.55	-4.61
	(851.56)	(2.44)	(0.41)	(757.53)	(5.38)
Observations	506	506	506	511	209
Mean DV	752.89	5.03	0.21	349.90	4.04
R^2	0.3	0.3	0.3	0.2	0.5

Note: The table presents results on consulting fee and salary interval of medical doctors. Fee is recorded in Pakistani Ruppees (Rs.), "Log(Fee)" is the logarithm of fee charged, while "Pr(Fee=0)" measures the probability that no fee is charged. "Salary" is measured in intervals of Pakistani Rs. 50,000. All regressions control for age, age-squared and 2 way interaction terms. Additional fixed effects included are for province, graduation year and medical specialty group. Standard errors are clustered at the province and medical specialty group level. Stars represent statistical significance. Data are from a survey of Pakistani doctors conducted by the authors. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 3: Fee and Salary

4.1 Mechanisms

4.1.1 Cohort Analysis

To begin to understand the mechanisms underlying our results, we examine heterogeneity across graduation cohorts. This allows us to understand if gaps exist at a specific career stage or among all post-reform cohorts. We define a pre-reform cohort as consisting of doctors who graduated before or in 1997 (cohort 0), and three post-reform cohorts: 1998-2006 (cohort 1), 2007-2014 (cohort 2), and 2015-2023 (cohort 3). In the next two sub-sections, we examine two distinct mechanisms which could explain our results, including cohort analyis results.

Columns 1 and 2 of table 4 show that female doctors in cohorts 2 and 3 charge a lower fee relative to their male peers and the comparison groups, with the largest fee gap among cohort 2 doctors. Meanwhile, the triple differences coefficient for cohort 1 is positive and statistically insignificant. This suggests that fee gaps are especially concentrated among mid-career and junior female doctors in cohorts 2 and 3, respectively.

Consistent with earlier results, women in every post-reform cohort in more exposed specialties spend more hours teaching than their male counterparts and comparison. The coefficient for teaching hours is largest among the oldest cohorts (column 4), while the ratio of teaching to practice hours is greatest and significant for cohort 2 (column 5). This pattern suggests that differences in teaching arise early on and are sustained, and that the allocation of teaching to practice hours may be especially high among mid-career post-reform women.

We also find suggestive differences in academic rank.[27] On average, post-reform female doctors in more exposed units are of similar rank in our sample as male doctors and relative to the comparison groups (table A4). However, there appears to be a gap in rank among mid-level post-reform cohorts, which is consistent with the overall pattern of differences being concentrated among mid-career cohorts.

[27] We are underpowered for this analysis relative to other outcomes because of more missing values.

	Fee	Log(Fee)	Practice Hrs	Teaching Hrs	Teaching Hours / Practice Hours
	(1)	(2)	(3)	(4)	(5)
Exposure x Female x Cohort 1	315.61	1.10	-1.10	5.42**	1.24
	(1576.95)	(8.49)	(3.64)	(2.36)	(0.87)
Exposure x Female x Cohort 2	-768.58	-9.46**	-2.08	4.12**	1.51**
	(922.17)	(4.45)	(2.20)	(1.57)	(0.57)
Exposure x Female x Cohort 3	-551.47	-8.63***	-0.28	3.83*	0.99
	(779.88)	(2.56)	(2.43)	(2.25)	(0.78)
Female x Cohort	-865.55	-2.50	1.43	-5.32***	-1.34*
	(1252.97)	(6.55)	(2.90)	(1.86)	(0.68)
Female	-166.89	-5.30**	-1.38	2.43**	0.84*
	(784.36)	(2.19)	(1.56)	(1.20)	(0.47)
Exposure x Cohort	-1719.76***	-6.50***	0.93	-1.32*	-0.44
	(555.53)	(2.24)	(0.73)	(0.67)	(0.27)
Exposure x Female	818.62	7.87***	0.99	-2.15	-0.71
	(901.92)	(2.69)	(1.98)	(1.44)	(0.58)
Exposure	1018.96	0.60	-0.45	0.20	0.00
	(888.92)	(2.57)	(1.24)	(1.19)	(0.38)
Observations	506	506	511	511	511
Mean DV	752.89	5.03	2.60	0.66	0.18
R^2	0.4	0.3	0.3	0.3	0.3

Note: The table presents cohort analysis for central outcomes. The omitted cohort is Cohort 0 which graduated between 1971-1997. Cohort 1 is the first post-reform cohort, graduating in 1998-2006; cohort 2 is the middle post-reform cohort, graduating in 2007-2014; cohort 3 is the most recent post-reform cohort, graduating in 2015-2023. All regressions control for age, age-squared and 2 way interaction terms. Additional fixed effects included are for province, graduation year and medical specialty group. Standard errors are clustered at the province and medical specialty group level. Stars represent statistical significance. Data are from a survey of Pakistani doctors conducted by the authors. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 4: Cohort Analysis

4.1.2 Greedy Jobs

The concentration of fee gaps in specific cohorts suggests that there may be a penalty at certain career stages in highly exposed units. While women in these cohorts are likely to have young children, we cannot directly examine differentials in child penalties because of data limitations.[28] However, to provide evidence consistent with greedy jobs, we examine if fee gaps vary with the median hours worked by using weekly hours data from Center for Studying Health System Change (2024), 1996-1997. In figure A4 fee disparities are highest in units with the highest median hours worked, and also in units which were previously more male-dominated.[29]

In table A5, we examine whether fee pricing increases more with practice hours in exposed units relative to less exposed units, in the spirit of convex returns to hours worked examined in Cortés and Pan (2019) and Goldin (2014). Column 1 shows that the relationship between log of fee and log of hours practicing (in intervals) is positive within specialty.[30] In column 2, the relationship becomes insignificant with province fixed effects, but is still positive. In columns 3 and 4, the fee premium to more practice hours appears to be higher among post-reform men, and non-existing among pre-reform men. Together these pieces of evidence are consistent with more exposed units being more likely to comprise of "greedy" jobs which have penalties for women at specific career stages.

4.1.3 Composition Changes

Survivor Bias and Selective Attrition in the Registration Data We interviewed doctors nearly two decades after the 1992 reform. Since many pre-reform doctors may have exited the labor force in 20 years, our sample is likely to over-represent doctors with longer careers, and potentially those with higher ability, leading to an upward bias in estimates.

We examine selective changes in doctor attrition in two ways. First, using doctor registration

[28] We collected data on first born children only. The mean age of women when they have their first child is 30 years in our sample, and the mean age of women in cohort 2 is 38 years. We also do not have panel data on respondents.

[29] For this exercise we estimate a regression where female and post are interacted with quantiles of hours worked and exposure. In the regression for median weekly hours, the merged specialties do not match exactly, which deceases sample size, and we use quartiles instead of quintiles of median hours.

[30] Since practice hours are measured in intervals, we focus on the direction, and not the magnitude of coefficients.

data, which provide information on the years for which a doctor maintained a valid license. As mentioned earlier, doctors are granted a 1 year temporary license after graduating from medical school and a permanent license on completing residency. The permanent license is renewed ever 5 years (Doctor Medical Lectures, 2024). While there is some bunching at 5 year intervals, PMDC data also show cancellations in between renewals.

In figure 6, we plot coefficients from 15 regressions, where the outcome in each is the probability of a valid license each year after its initial issue for up to 15 years after registration. The greatest decline in the likelihood of an active license is about 5 years after registration, after which licenses are renewed. Among those who remain after this point, the retention gap does not further decrease, and appears to recover by 15 years after the year of license issue. Furthermore, plotted coefficients are generally not statistically significant.

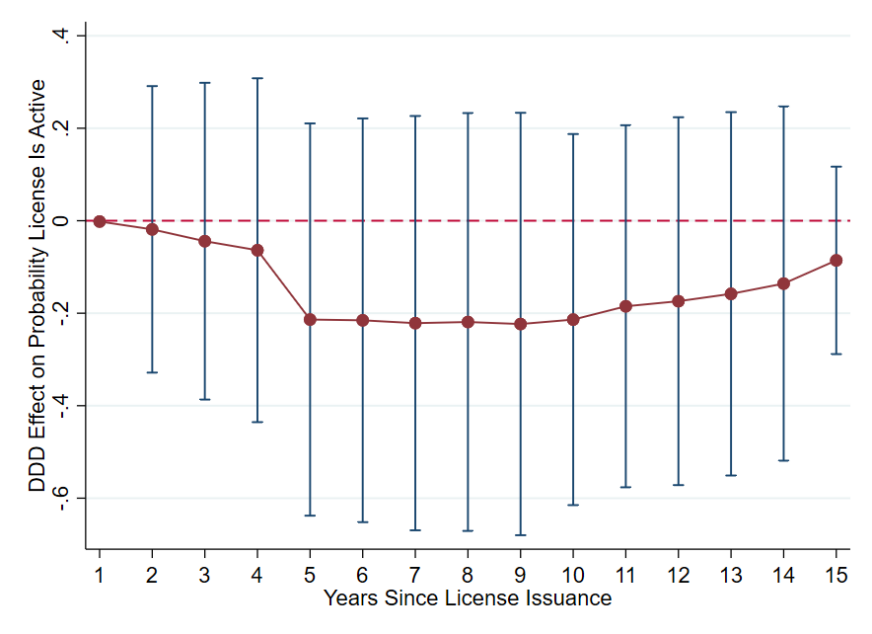


Figure 6: Probability of an Active License, by Year Since License Issue

Note: The figure plots coefficients from 15 regressions using equation 1. Each regression estimates the probability of having an active license each year after its initial issue for upto 15 years after license issue. All regressions include specialty, province and year of license issue fixed effects. Standard errors are clustered by specialty and province. The figure is based on doctor registration data from the PMDC.

Contemporary Gender Gaps Using Labor Force Surveys We also examine survivorship among pre-reform cohorts by examining how contemporary gender gaps for these cohorts

compare with gender gaps in our survey. Contemporary gender gaps are found from the Labor Force Surveys of Pakistan (LFS, 1990-2012). We restrict doctors in this LFS sample to be within 10 years of when they are expected to have graduated,[31] and focus on the probability of working and hours worked, since these exist in both datasets.[32] We merge pre-reform cohorts from these two datasets and calculate gender gaps as gaps between men and women in these outcomes.

In figure A5, gender gaps in the LFS are contemporaneous while gender gaps in our survey are among surviving doctors observed about 20 years after graduation. If survivors are positively selected on labor-force attachment, we might expect women to have lower probability of working and fewer hours relative to men than the same gaps in our survey. Table A9 also reports coefficients from regressions where these outcomes are regressed on an indicator that equals 1 for respondents of our survey, an indicator for female gender and the interaction of the two (along with graduation year fixed effects). Female doctors are less likely to work (column 1) and work fewer hours (column 2). Individuals in our survey are more likely to report working and work a greater number of hours. The gender gap in each outcome is positive but under statistical significance. This is consistent with the positive but under significant attrition patterns we find using registration data (figure 6).

Direct Tests of Selection Changes in Entry into Specialty-Province Units We also directly examine selection changes at entry using our main estimating equation 1 and background characteristics that we collected information on. Table A8 examines if gaps in these characteristics (most of which are pre-determined) changed after the reform. In column 1, post-reform female cohorts are more likely to have completed their house job in a large hospital than comparison groups. Beyond this, there do not appear to be significant differences in the quality of the medical school attended, father's education, age when a doctor had a first child and having ever been

[31]The LFS do not provide a graduation years. Year of graduation is predicted from when an individual is 23 years of age and is based on the assumption that the individual entered medical school at age 18 and completes medical school in 5 years. Further, we use data on individuals with a medical (MBBS) degree instead of an occupation, since data on occupations are more sparse.

[32]The probability of working in the LFS corresponds to working in the previous week or having a job that an individual will return to. In our survey, it corresponds to if the doctor is currently practicing. Hours worked in the LFS correspond to total hours worked across all jobs. In our survey it corresponds to the sum of maximum hours spent practicing, teaching and research, respectively, based on the time interval that an individual selected.

married, though with the caveat that our sample becomes smaller due to fewer responses.

To quantify the extent to which selection at entry into occupations may change our estimated coefficients, we could also control for medical school quality, restricted to a common sample in table A10. Focusing on our two main findings, accounting for medical school quality does not substantially change the magnitude of the coefficients. Fee becomes more negative by 0.03 log points, while the coefficient for teaching hours is marginally larger, consistent with teaching hours being correlated with lower fee. However, these are not consistent with female selection worsening relative to male selection in more exposed units, which is what is might be expected when entry becomes less selective after the reform. Given our statistically insignificant results in table A8, we cannot read into these results too much.

4.1.4 Other Mechanisms

Demand-side factors, such as patient demand, could also contribute to our findings. Patients in more exposed units may be less comfortable consulting with female doctors or spending an extended amount of time with them. Lower demand would then be reflected in lower fee charged by female doctors and an increase in teaching time. Conversely, pre-reform female doctors are fewer and therefore more scarce. If patients have distinct demands for doctors of a specific gender, an increase in the supply of female doctors could lower the price they charge. While we cannot rule out these mechanisms, note that patient volume did not change after the reform, even as fee and time use did. Moreover, we find that the gaps are highest among mid-career cohorts, whereas a patient-preference mechanism might be expected to affect female doctors across all cohorts.

4.2 Effects for the Overall Unit

In table 5, we use exposure-based difference-in-differences to examine if specialty units changed overall due to exposure to female entry post-reform. We do not find significant changes in any of our main outcomes. For example, despite larger sample sizes, log of fee and salary do not change significantly in more exposed units relative to less exposed units after the reform. Time use outcomes and patient volume are also stable. These results suggest that units did not change

overall, units did not change differently based on exposure in the outcomes we examine; instead, the differences are in how male and female doctors practice within these units.

	Log(Fee)	Salary	Teaching Hours	Practice Hours	Number OP	Time OP	Scheduled Breaks
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Exposure x Post	-1.50	-10.19	1.02	1.50	16.85	-0.85	-0.34
	(1.73)	(13.68)	(1.22)	(1.14)	(14.43)	(0.77)	(0.23)
Exposure	4.10*	24.55	-0.57	-0.16	3.14	1.42*	0.35
	(2.34)	(32.42)	(1.11)	(1.21)	(32.03)	(0.85)	(0.35)
Observations	507	215	512	512	512	512	512
Mean DV	5.02	6.69	0.66	2.60	30.08	2.47	0.23
R^2	0.3	0.4	0.2	0.3	0.2	0.2	0.3

Note: The table presents ITT difference-in-difference analysis for overall specialty-province units. All regressions control for age and age-squared. Additional fixed effects included are for province, graduation year and medical specialty group. Standard errors are clustered at the province and medical specialty group level. Stars represent statistical significance. Data are from a survey of Pakistani doctors conducted by the authors. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 5: Changes in the Overall Field

5 Robustness Checks and Extensions

Controlling for medical shool quality and background characteristics: Selection differences between specialties and between pre-reform and post-reform medical doctors can bias our estimates. For example, female doctors who are less motivated may sort into specialties with a low pre-existing female share and also charge a low fee. Medical college entry score or secondary school exit exam scores are not publicly available for the early 1990s and 1980s to the best of our knowledge. Similarly, parental education data are rarely collected for medical doctors. Therefore, we collect this data in our survey and control for it as pre-determined characteristics in our regressions on current practice outcomes.

Specifically, we control for secondary school exit exam scores ("Intermediate" or "Inter") and father's education level (table 6, panel II).[33] For individuals who did not remember their exact

[33]The majority of respondents (96%) had completed Intermediate exams; we converted scores to

	Log(Fee)	Zero Fee	Teaching Hours
	(1)	(2)	(3)
I. Control for Medical College Quality			
Exposure x Female x Post	-6.25*	0.29	6.78***
	(3.68)	(0.57)	(2.04)
Observations	427	427	431
R^2	0.32	0.30	0.34
II. Control for Inter Grade and Father's Education			
Exposure x Female x Post	-15.72*	2.03*	6.33**
	(7.76)	(1.03)	(2.82)
Observations	190	190	194
R^2	0.61	0.58	0.51
III. Trends in Specialty Characteristics			
Exposure x Female x Post	-8.99*	1.09*	5.44***
	(4.91)	(0.62)	(1.91)
Observations	372	372	374
R^2	0.34	0.33	0.28
IV. Survey Round Control			
Exposure x Female x Post	-5.84**	0.50	4.82***
	(2.89)	(0.48)	(1.28)
Observations	489	489	493
R^2	0.32	0.29	0.29
V. Excluding Female Shares Between 1992-1997			
Exposure x Female x Post	-6.16*	0.87	3.83*
	(3.42)	(0.53)	(2.14)
Observations	461	461	465
R^2	0.34	0.31	0.29

Note: The table presents robustness checks discussed for our main results in section 6. Data are collected by the authors.

Table 6: Robustness Checks

score but only remembered their division (which is a score range, similar to a grade), we assigned the mean score of that division. While the sample size decreases for these regressions, our results remain consistent with the main findings.

Relatedly, we also control for the quality of the medical college attended. In panel I. of table 6, we control for medical college quality. The PMDC grants each medical college a grade, ranging from A+ to F, based on college inspections conducted by the PMDC. These include inspections of the infrastructure and employment standards.[34] Public colleges have two rounds of published grades from two inspections (in 2013 and 2019), while private medical colleges have these from 2019. We include both of these grades in our regressions, and the results remain similar to our main findings.[35]

Controlling for Trends in Specialty Characteristics: Trends across units may be varying in a way that biases our results. For example, more exposed units may be experience salary decreases or a change in time use structures that is unrelated to the reform. To mitigate this concern, in panel III. of table 6, we account for the evolution of specific specialty characteristics by allowing for different time trends in characteristics. We focus on two traits: median income and median hours worked. Specialty traits are obtained from an out-of-sample population of doctors from the 1996-1997 round of Center for Studying Health System Change (2024), under the assumption that the relative ordering of specialty traits, such as hours worked or pay, is similar to that in Pakistani specialties. The benefit of using an out-of-sample survey is that the traits in this survey are plausibly exogenous to our data. We control for trends in these two traits in our regressions after dropping any specialties that do not match between the data sets. Our results are generally consistent with our main results.

Benchmarking Fee Results Using Online Portal Data: Since some of our outcomes are self-reported, we benchmark our fee findings against data from Oladoc.pk and Marham.pk, two of the

equivalent Intermediate scores for individuals who completed O'Level exams using Inter Boards Coordination Commission (2025). Equivalency scales for conversion are available for more recent years; however, this exercise is robust to applying hypothetical stricter equivalence scales, available upon request.

[34] Available from the PMDC website. See here:https://pmdc.pk/Publication/Standards

[35] Missing 2013 grades are replaced with 2019 grades.

largest online healthcare platforms in Pakistan. Doctors on these portals are also self-selected, but unlike our survey, the fees and timings are patient-facing and used to book real consultations, so they are likely accurate and verified by clinics. This makes measurement error in the key outcome less of a concern. We match portal doctors to our survey doctors on gender, specialty, and district of practice.

In table A6, the online portal estimates are noisier, but the coefficient signs mostly line up with our main results. In column 1, consultation fees for female doctors are lower than for male doctors and relative to comparison groups for both hospital and tele-health appointments, though not significantly. In column 5, female doctors hold more junior positions than male doctors, relative to the pre-reform gap and to less exposed units, and in column 6 are less likely to report patient procedures as available services.

The number of observations on the time spent with each patient (reported by patients) is much fewer in this subsample of doctors on portals. In figure A7 we plot coefficients from a simpler difference-in-differences regression of time per patient on indicators for female and post interacted. There is some suggestive evidence that less time with each patient may decrease patient satisfaction. Using data from online portals, we plot binscatter diagrams of time spent with each patient against satisfaction with the diagnosis, and overall satisfaction (figure A6). In both cases, time per out patient are few in our subsample of doctors, but indicate a mildly positive relationship, especially for satisfaction with the diagnosis. On their own estimates based on online portals are weak, but they agree in sign with the survey, and the negative fee gap is internally consistent with post-reform female doctors sitting in more junior positions than their male counterparts.

Specification Checks: We run two additional specification tests. First, in panel IV. of table 6, we include fixed effects for survey round to account for any features specific to the survey round being conducted affecting our results. The second is to define the pre- and post-reform female shares excluding registrations between 1992-1996 (panel V. of table 6). In this interim time period, the reform had passed, and the female share of doctors may have started to grow as some doctors graduated earlier than the expected graduation year of 1997. Therefore, attributing these years to

either pre-reform or post-reform years could contaminate the female shares calculated. In both panels, our results for log of fee and teaching hours remain similar in direction.[36]

Regressions with IPW Weights: We repeat the analysis for our main findings using IPW weights in table A7. IPW weights are calculated by combining survey with registration data, after which we conduct a logistic regression for being in our survey on indicators for province and specialty fixed effects. Inverse probability weights are then constructed using estimated regression coefficients. Most results remain similar to the main findings, though significance decreases for time spent with out-patients.

6 Conclusion

We examine how the expansion of female doctors in medicine affects medical practice in Pakistan in the long run. Our analysis draws on original survey data from over 700 doctors across all provinces, as well as administrative records on medical registrations. A key empirical challenge is that women often enter occupations only after occupations have already undergone structural changes. To address this, we leverage variation in exposure to a plausibly exogenous change in admissions policy.

While more exposed units do not show broad aggregate changes as more women enter, female doctors in these units charge lower fees than their male counterparts, and relative to gaps in less exposed units and pre-reform cohorts. They also spend their time differently, and distinctly spend more time in teaching than the comparison groups. Although these doctors do not differ from their male peers in practice hours and patient volume, female doctors spend less time with each patient and are more likely to take scheduled breaks, which may be adaptation strategies to time-intensive specialty requirements.

Taken together, our findings suggest that units highly exposed to female entry are more segmented by gender in the long-run, 20 years after the reform. These results imply that while

[36] In panel V, note that the sample size decreases because we are excluding years in calculating the pre-reform and the change in female share of doctors. If certain units first appeared in that interim period, they would no longer have a pre-reform female share defined.

gender diversity policies improve representation, they are unlikely to generate equal outcomes without complementary measures that promote advancement and reduce structural barriers for women. Future work may look at how the reform changed the quality of patient care, contributing to literature on how health policies affect patients themselves (Alexander, 2020; Das et al., 2016; Das and Hammer, 2005; Das et al., 2008).

References

AAMC (2018). Women are changing the face of medicine in America. https://www.aamc.org/news/women-are-changing-face-medicine-america.

Adams, A., K. Hara, K. Milland, and C. Callison-Burch (2025). The gender wage gap in an online labor market: The cost of interruptions. *Review of Economics and Statistics 107*(1), 55–64.

Aga Khan University Hospital (AKUH) (2026). Frequently Asked Questions. https://www.aku.edu/mcpk/residency/Pages/faqs.aspx.

Alexander, D. (2020). How do doctors respond to incentives? unintended consequences of paying doctors to reduce costs. *Journal of Political Economy 128*(11), 4046–4096.

Aqeel, F. (2024). Removing barriers to entry in medicine: Evidence from pakistan. *The World Bank Economic Review*, lhad039.

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

Black, S. E. and C. Juhn (2000). The rise of female professionals: Are women responding to skill demand? *American Economic Review 90*(2), 450–455.

Blau, F. D. and A. E. Winkler (2017). Women, work, and family. Technical report, National Bureau of Economic Research.

Bolotnyy, V. and N. Emanuel (2022). Why do women earn less than men? evidence from bus and train operators. *Journal of Labor Economics 40*(2), 283–323.

Cain, G. G. (1976). The challenge of segmented labor market theories to orthodox theory: A survey. *Journal of economic Literature 14*(4), 1215–1257.

Calkins, A., A. J. Binder, D. Shaat, and B. Timpe (2023). When sarah meets lawrence: The effects of coeducation on women's college major choices. *American Economic Journal: Applied Economics 15*(3), 1–34.

Callaway, B., A. Goodman-Bacon, and P. H. Sant'Anna (2021). Difference-in-differences with a continuous treatment. *arXiv preprint arXiv:2107.02637*.

Carlana, M., E. Miglino, and M. M. Tincani (2024). How far can inclusion go? the long-term impacts of preferential college admissions. Technical report, National Bureau of Economic Research.

Center for Studying Health System Change (2024). Community tracking study physician survey, 1996-1997: [united states]. Accessed February 14, 2024.

CIEC (2025). Pakistan average monthly wages.

Collis, M. R. and C. Van Effenterre (2025). *Workplace Hostility*. IZA-Institute of Labor Economics.

Cortés, P. and J. Pan (2019). When time binds: Substitutes for household production, returns to working long hours, and the skilled gender wage gap. *Journal of Labor Economics 37*(2), 351–398.

Cortés, P. and J. Pan (2023). Children and the remaining gender gaps in the labor market. *Journal of Economic Literature 61*(4), 1359–1409.

Das, J., A. Chowdhury, R. Hussam, and A. V. Banerjee (2016). The impact of training informal health care providers in india: A randomized controlled trial. *Science 354*(6308), aaf7384.

Das, J. and J. Hammer (2005). Which doctor? combining vignettes and item response to measure clinical competence. *Journal of Development Economics 78*(2), 348–383.

Das, J., J. Hammer, and K. Leonard (2008). The quality of medical advice in low-income countries. *Journal of Economic perspectives 22*(2), 93–114.

Dawn Newspaper (2011). "Government Helpless to Regulate Doctors Fee Structure". https://www.dawn.com/news/623181/govt-helpless-to-regulate-doctors-fee-structure.

Deakin, S. F. et al. (2013). Addressing labour market segmentation: The role of labour law.

Deschacht, N., S. Vujić, and O. Frison (2025). The greedy jobs phenomenon as a driving force behind the gender pay gap: A systematic review. *De Economist 173*(1), 177–204.

Doctor Medical Lectures (2024). "PMDC Renewal Process: Complete Guide for Doctors in Pakistan". https://dmlec.com/pmdc-renewal-process-everything-you-should-know/.

Feldstein, M. S. (1970). The rising price of physician's services. *The Review of Economics and Statistics*, 121–133.

Folke, O. and J. Rickne (2022). Sexual harassment and gender inequality in the labor market. *The Quarterly Journal of Economics 137*(4), 2163–2212.

Gicheva, D. (2013). Working long hours and early career outcomes in the high-end labor market. *Journal of Labor Economics 31*(4), 785–824.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review 104(4)*, 1091–1119.

Greenberg, K., M. Wasserman, and E. A. Weber (2026). The effects of gender integration on men: Evidence from the us military. *The Quarterly Journal of Economics*, qjag016.

Hampole, M., F. Truffa, and A. Wong (2026). Peer effects and the gender gap in corporate leadership: Evidence from mba students. *The Quarterly Journal of Economics*, qjag015.

Harju, J., S. Jäger, and B. Schoefer (2021). Voice at work. Technical report, National Bureau of Economic Research.

Harrington, E. and M. E. Kahn (2026). Can remote work narrow gender gaps in greedy jobs? In *AEA Papers and Proceedings*, Volume 116, pp. 218–222. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203-2425.

Harris, J. (2022). Do wages fall when women enter an occupation? *Labour Economics 74*, 102102.

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2019). The allocation of talent and us economic growth. *Econometrica 87*(5), 1439–1474.

Hussain, A., M. S. Sial, S. M. Usman, J. Hwang, Y. Jiang, and A. Shafiq (2019). What factors affect patient satisfaction in public sector hospitals: Evidence from an emerging economy. *International journal of environmental research and public health 16*(6), 994.

Inter Boards Coordination Commission (2025). Conversion for equivalence. https://ibcc.edu.pk/conversion-for-equivalence/.

JUSTIA US Supreme Court (1996). United states v. virginia, 518 u.s. 515 (1996). https://supreme.justia.com/cases/federal/us/518/515/.

Karpowitz, C. F., S. D. O'Connell, J. Preece, and O. Stoddard (2024). Strength in numbers? gender composition, leadership, and women's influence in teams. *Journal of Political Economy 132*(9), 3077–3114.

Khalid, F., W. Raza, D. R. Hotchkiss, and R. H. Soelaeman (2021). Health services utilization and out-of-pocket (oop) expenditures in public and private facilities in pakistan: an empirical analysis of the 2013–14 oop health expenditure survey. *BMC health services research 21*, 1–14.

Kim, S. and P. Moser (2025). Women in science. lessons from the baby boom. *Econometrica 93*(5), 1521–1560.

Kumar, N., U. Lee, M. Lowe, and O. Ogunnote (2026). Internal versus institutional barriers to gender equality: Evidence from british politics. *American Economic Review 116*(5), 1914–1953.

La Brooy, C., H. Sabanovic, S. J. Méndez, J. Yong, A. Scott, A. G. Elshaug, and K.-H. Prang (2025). 'charge what you think you're worth': a qualitative study exploring the gender pay gap in medicine and the role of price transparency. *Internal Medicine Journal 55*(3), 384–392.

Le Barbanchon, T., R. Rathelot, and A. Roulet (2021). Gender differences in job search: Trading off commute against wage. *The Quarterly Journal of Economics 136*(1), 381–426.

Lee, L. M. and J. C. Thompson (2019). Gender wage differentials: A preference based approach. *Manuscript Under Review*.

Li, M. (2024). Devaluation for whom? feminization and wages in an economically polarized labor market, 2003–2019. *Social Forces 103*(2), 495–519.

Mas, A. and A. Pallais (2017). Valuing alternative work arrangements. *American Economic Review 107*(12), 3722–3759.

Masood, A. (2019). The influence of marriage on women's participation in medicine: The case of doctor brides of pakistan. *Sex Roles 80, no. 1*, 1–18.

Medley, S. (2016). " gender balancing" as sex discrimination in college admissions. *Harv. CR-CLL Rev. 51*, 537.

Merali, Z., A. K. Malhotra, M. Balas, G. R. Lorello, A. Flexman, T. Kiran, and C. D. Witiw (2021). Gender-based differences in physician payments within the fee-for-service system in ontario: a retrospective, cross-sectional study. *CMAJ 193*(41), E1584–E1591.

Oldenziel, R. (2001, 02). Making technology masculine: Men, women, and modern machines in america, 1870-1945. *Bibliovault OAI Repository, the University of Chicago Press 106*.

Painter, R. (2017). Comparing care at teaching hospitals to non-teaching hospitals. Tech report, Painter Law Firm Medical Malpractice Attorneys.

Pan, J. (2015). Gender segregation in occupations: The role of tipping and social interactions. *Journal of Labor Economics 33*(2), 365–408.

Pasha, H. A. and A. Nasar (2003). Public private partnership in the health sector: Evidence from a developing country.

Rambachan, A. and J. Roth (2019). An honest approach to parallel trends. *Unpublished manuscript, Harvard University*.

Richardson, J. R., S. J. Peacock, and D. Mortimer (2006). Does an increase in the doctor supply reduce medical fees? an econometric analysis of medical fees across australia. *Applied Economics 38*(3), 253–266.

The Express Tribune (2021). New admissions policy: 50:50 ratio for men and women in medical colleges, says PMDC. https://tribune.com.pk/story/767721/new-admissions-policy-5050-ratio-for-men-and-women-in-medical-colleges-says-pmdc.

The News (2024). "Private Practice of Doctors not Banned". https://www.thenews.com.pk/print/1216686-private-practice-of-doctors-not-banned.

Topalova, P. and A. Khandelwal (2011). Trade liberalization and firm productivity: The case of india. *Review of economics and statistics 93*(3), 995–1009.

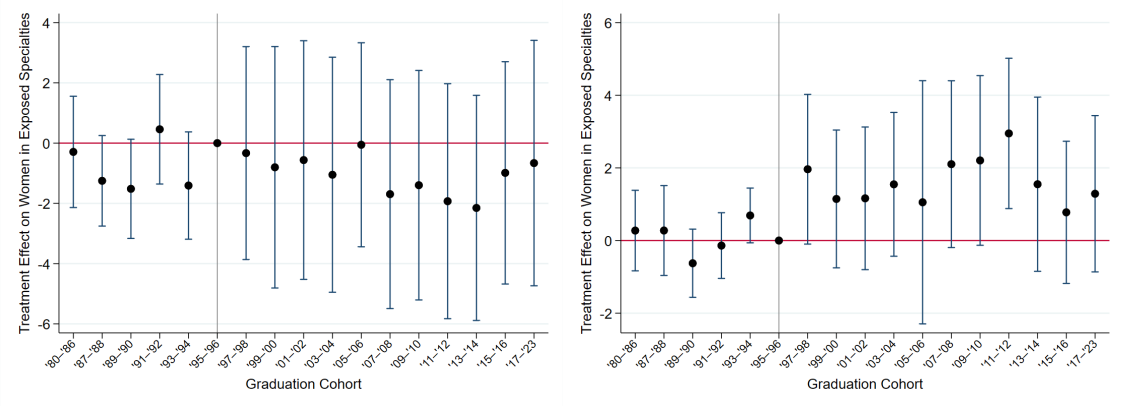
Wasserman, M. (2017). Hours constraints, occupational choice, and gender: Evidence from medical residents.

Wiswall, M. and B. Zafar (2018). Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics 133*(1), 457–507.

Wooldridge, J. M. (2007). Inverse probability weighted estimation for general missing data problems. *Journal of econometrics 141*(2), 1281–1301.

Zhao, L., Y. Luo, and L. Li (2025). Place-based affirmative action in college admission and educational attainment. *Journal of Development Economics*, 103624.

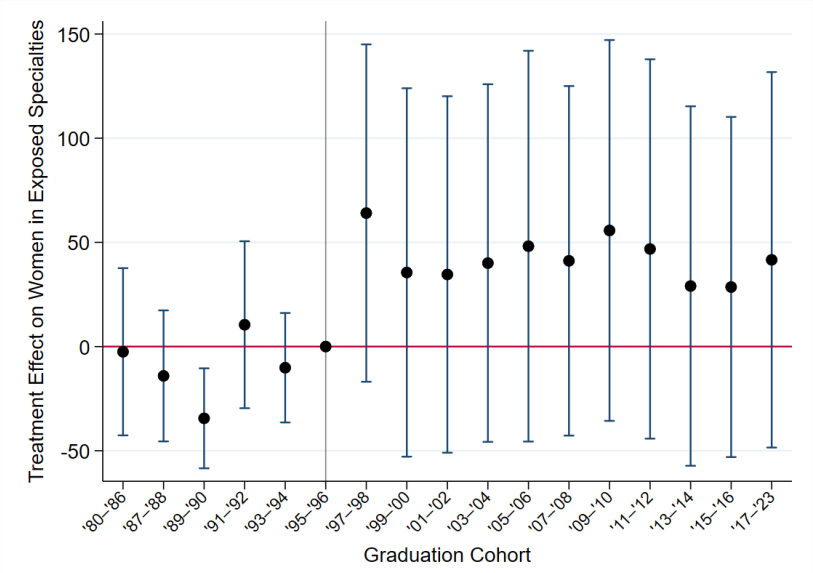
A Appendix



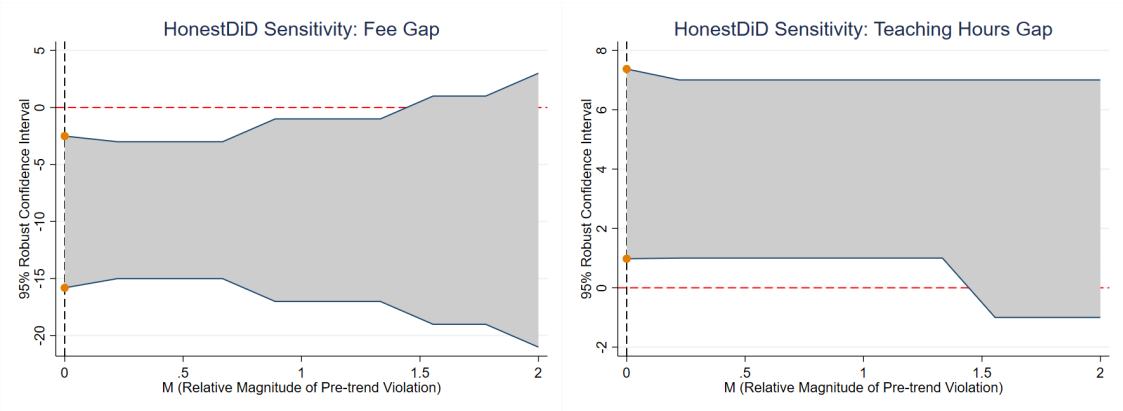
(a) Event Study - Hours Spent Practicing (b) Event Study - Hours Spent Researching

Figure A1: Event studies for hours spent practicing and hours spent on research each week, both in intervals of hours. Regressions include specialty, province and year fixed effects, and standard errors are clustered by specialty and province. The omitted period is 1995-1996. Data are from a survey of doctors by the authors.

Figure A2: Event Study - Maximum Hours Working

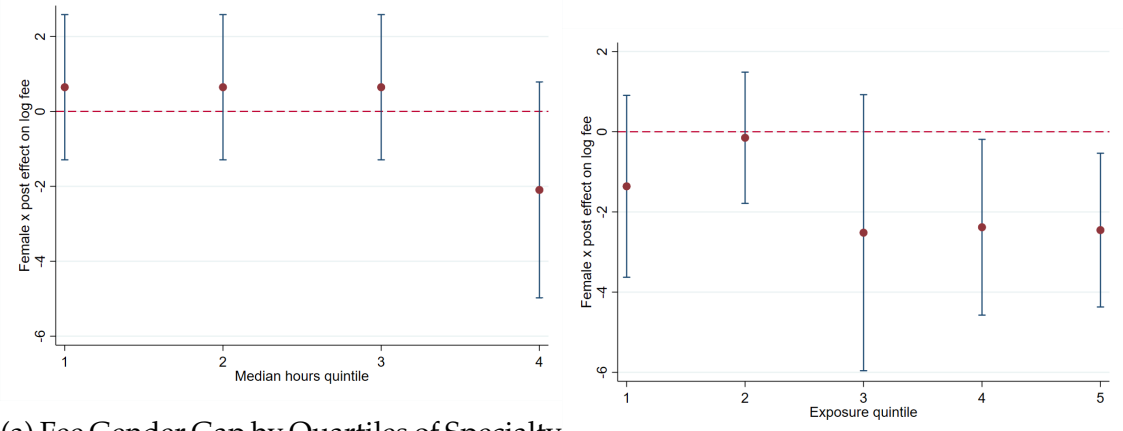


Event studies for the maximum number of hours practicing, teaching and researching per week, calculated with the upper bounds of time intervals. Regressions include specialty, province and year fixed effects, and standard errors are clustered by specialty and province. The omitted period is 1995-1996. Data are from a survey of doctors by the authors.



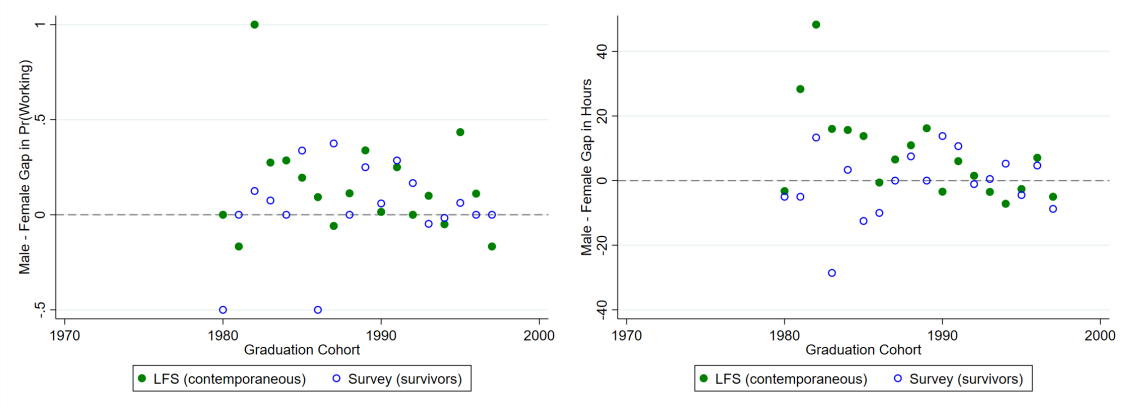
(a) Log (Fee) (b) Teaching Hours

Figure A3: Honest DID (Rambachan and Roth, 2019) event studies for the log of fee and for hours spent teaching each week. Data are from a survey of doctors by the authors.



(a) Fee Gender Gap by Quartiles of Specialty Median Hours (b) Fee Gender Gap by Quintile of Exposure

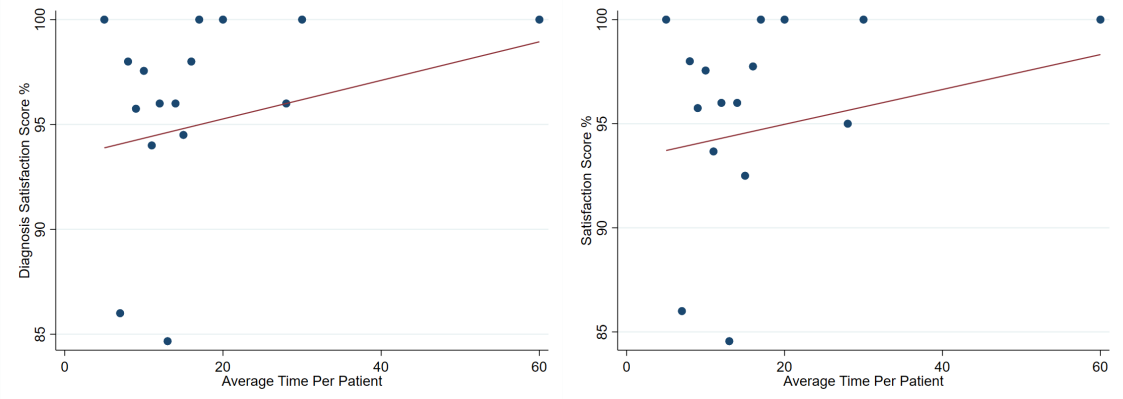
Figure A4: The figure provides estimates of fee gender gaps by (a) the median number of hours worked in the specialty, where median hours are obtained from an external survey of physicians, and (b) the quintile of exposure to the reform, measured as the pre-reform female share of doctors.



(a) Probability of Working (b) Hours Worked

Figure A5: Contemporaneous vs. Survivor Gender Gaps: LFS and Survey

Note: The figure plots gender gaps in the probability of working (panel a) and in total hours worked last week (panel b), from the LFS and from our survey. The LFS gaps are contemporaneous, between doctors within 10 years of their predicted graduation. The survey gaps are among "survivors", observed about 20 years after graduation.



(a) Diagnostic Satisfaction (b) Overall Satisfaction

Figure A6: Correlation Between Satisfaction and Time Spent with Each Patient: Online Portals

Note: The figure uses data on doctors from two online portals: Oladoc.pk and Marham.pk, and plots a binscatter of patient satisfaction against time spent per patient. In panel (a) satisfaction is in terms of diagnosis, and in panel (b) it is overall satisfaction.

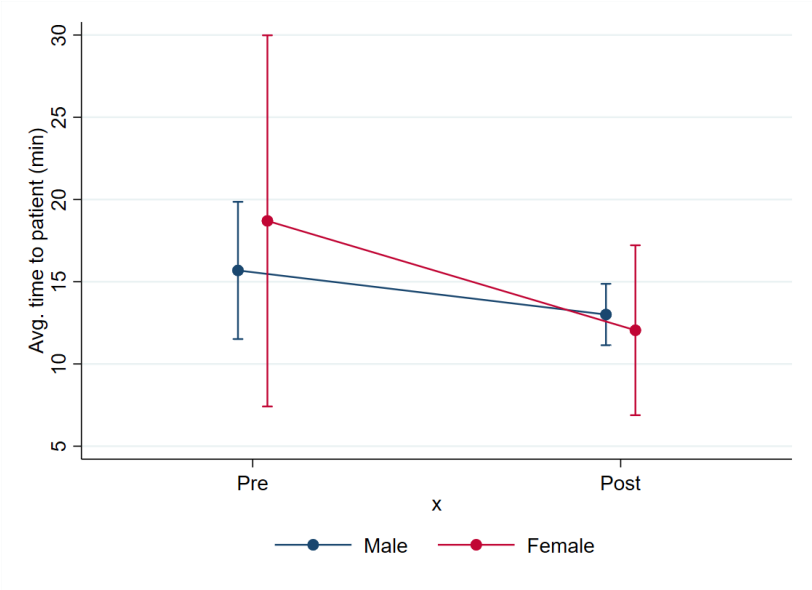


Figure A7: Time per Patient - Online Portals

Note: The graph plots coefficients from a regression of doctors' time spent with patients on the interaction of female and post, with controls for specialty and province. Standard errors are robust.

	Mean	SD	Min	Max	N
Female	0.51	0.50	0.00	1.00	734
Age	45.03	11.95	23.00	76.00	735
Urban Birthplace	0.79	0.40	0.00	1.00	331
Secondary School Score (%)	86.61	7.04	64.00	97.00	300
Mother's Education	11.80	2.97	1.00	15.00	329
Father's Education	13.22	2.01	1.00	16.00	328
No. Other Doctors	2.35	3.00	0.00	20.00	331
Married	0.40	0.49	0.00	1.00	740
Currently Practice	0.94	0.24	0.00	1.00	734
Urban Practice	0.93	0.25	0.00	1.00	689
Public Practice	0.45	0.50	0.00	1.00	689
Have a Second Practice	0.28	0.45	0.00	1.00	688
Log(Fee, Primary)	5.08	2.79	0.00	8.52	682
Log(Fee, Secondary)	6.79	0.93	0.00	8.41	191
Salary	4.05	1.96	1.00	11.00	299
Total Time per Patient, Primary	2.31	1.17	0.00	6.00	740
Time per Patient, Secondary	0.74	1.36	0.00	6.00	740
Total Number Patients, Primary (daily)	29.19	23.36	0.00	100.00	740
Number Patients, Secondary (daily)	4.65	9.29	0.00	46.00	740
Practice Hours (wk)	2.53	1.02	1.00	6.00	687
Teaching Hours (wk)	0.70	0.93	0.00	5.00	735
Research Hours (wk)	0.82	0.88	0.00	5.00	735
Maximum Hours (wk)	60.89	20.30	20.00	170.00	687

Note: The table presents summary statistics for our regression sample of Pakistani doctors. "Primary" refers to a primary practice and "secondary" to a secondary practice. Data are collected by the authors.

Table A1: Summary Statistics: Regression Sample

	Mean (PMDC)	Mean (Survey)	Difference	SE	N
Graduation Year	2001.98	2002.14	-0.16	0.44	241884
Male	0.52	0.50	0.02	0.02	241879
Rural Practice	0.26	0.07	0.18***	0.01	214269
Khyber-Pakktunkhua	0.11	0.12	-0.02	0.01	241836
Punjab	0.46	0.61	-0.15***	0.02	241836
Sindh	0.30	0.12	0.18***	0.01	241836
Surgery	0.17	0.10	0.07***	0.01	41002

Note: The table presents differences in means between groups: doctors in our survey and doctors in the PMDC registration database. The statistics for surgical specialties are calculated relative to all physicians in medicine and surgical specialties and exclude non-specialists. Survey data are collected by the authors.

Table A2: Survey Compared to Doctor Registrations

	Probability Place of Work is:			
	Public	Urban	Large	Academic
	(1)	(2)	(3)	(4)
Exposure x Female x Post	-0.73	-0.22	0.50	0.79
	(0.64)	(0.43)	(0.50)	(0.82)
Female x Post	0.42	0.14	-0.46	-1.07
	(0.53)	(0.38)	(0.41)	(0.69)
Female	0.54	0.50*	0.20	1.53**
	(0.54)	(0.25)	(0.38)	(0.66)
Exposure x Post	0.05	-0.09	-0.50	-1.09***
	(0.44)	(0.29)	(0.52)	(0.40)
Exposure x Female	-0.59	-0.49	-0.09	-1.31*
	(0.65)	(0.29)	(0.46)	(0.77)
Exposure	-0.06	0.30	0.31	0.82
	(0.64)	(0.38)	(0.53)	(0.71)
Observations	511	511	511	446
Mean DV	0.45	0.92	0.73	0.45
R^2	0.2	0.2	0.3	0.4

Note: The table presents results on place of work: public practice, urban practice, large hospital practice, and academic hospital practice. All regressions control for age, age-squared, and survey round fixed effects. Additional fixed effects included are for province, graduation year and medical specialty. Standard errors are clustered at the province and medical specialty. Stars represent statistical significance. Data are from a survey of Pakistani doctors conducted by the authors. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A3: Hospital and Clinic

	Position	At Least Associate	Full Professor	Position
	(1)	(2)	(3)	(4)
Exposure x Female x Post	-0.19	-0.22	0.20	
	(2.53)	(1.10)	(0.97)	
Female x Post	-0.99	-0.21	-0.45	
	(2.02)	(0.90)	(0.77)	
Female	0.57	-0.11	-0.52	0.62
	(2.01)	(1.06)	(0.74)	(2.02)
Exposure x Post	-2.52*	-0.88	-0.62	
	(1.43)	(0.57)	(0.72)	
Exposure x Female	0.61	0.60	0.93	0.54
	(2.40)	(1.27)	(0.88)	(2.47)
Exposure	5.22**	1.14	1.27	5.30***
	(2.05)	(1.03)	(0.76)	(1.94)
Exposure x Female x Cohort 1				-3.99
				(3.10)
Exposure x Female x Cohort 2				-5.69**
				(2.78)
Exposure x Female x Cohort 3				2.70
				(1.99)
Observations	247	247	247	247
Mean DV	1.00	0.79	0.79	0.79
R^2	0.6	0.6	0.5	0.7

Note: The table presents estimates on doctor academic position. The outcome in the first column is academic position, which equals 0 if no Assistant Professor-or-higher academic rank was reported, 1 for Assistant Professor, 2 for Associate Professor and 3 for Full Professor or Consultant. Stars represent statistical significance. In column 2, the outcome is an indicator for having an Associate Professor or higher academic rank, in column 3 it is for being a Full Professor. Column 4 presents a cohort analysis. Data are from a survey of Pakistani doctors conducted by the authors. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A4: Position

	Log(Fee)			
	All Doctors	All Doctors	Pre-Reform	Post-Reform
	(1)	(2)	(3)	(4)
Exposure x Log(Hrs Practice)	0.86**	0.63	-0.23	1.00*
	(0.35)	(0.38)	(0.48)	(0.52)
Observations	324	324	130	194
Mean DV	4.81	4.81	5.18	4.56
R^2	0.3	0.3	0.4	0.3
Province FE	N	Y	Y	Y

Note: The table presents examines the returns to working more practice hours. The sample is of male doctors. The regression estimated in column 1 is

$$Log(Fee)_{ispt} = \alpha_0 + \alpha_1 Exposure_s \times Log(HoursPracticed)_{ispt} + \alpha_3 Age + \alpha_4 Age^2 + \alpha_s + \alpha_t + \varepsilon_{ispt} \quad (3)$$

, where the subscripts denote individual i in specialty s, province p and with a graduation year t. Log(Hrs Practice) is the log of intervals of hours practiced. Column 2 further adds province fixed effect. In column 3 the regression in column 2 is re-estimated on pre-reform cohorts, and in column 4 on post-reform cohorts. Stars represent statistical significance. Data are from a survey of Pakistani doctors conducted by the authors. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A5: Fee Charged per Out-Patient and Practice Hours

	Fee (Rs.) Hospital	Fee (Rs.) Telehealth	Log(Fee) Hospital	Log(Fee) Telehealth	Position	Invasive Procedure
	(1)	(2)	(3)	(4)	(5)	(6)
Exposure x Female x Post	-2600.05	-20598.79	0.72	-2.34	-8.09**	-2.68**
	(5580.85)	(13725.56)	(2.66)	(1.51)	(3.07)	(1.03)
Female x Post	2966.06	17770.14	-0.41	1.40	7.01**	2.00**
	(4762.15)	(13060.11)	(2.29)	(1.34)	(2.81)	(0.85)
Female	-4016.71	-17021.46	-0.13	-1.26	-6.08**	-0.81
	(4548.54)	(12980.48)	(2.02)	(1.28)	(2.70)	(0.84)
Exposure x Post	-970.88	11210.72	-1.49	-1.13	3.28	-0.13
	(2073.07)	(12822.20)	(1.39)	(0.90)	(2.14)	(0.52)
Exposure x Female	4431.08	19943.67	0.21	2.22	6.72**	1.12
	(5386.32)	(13584.76)	(2.35)	(1.44)	(2.95)	(1.04)
Exposure	-1738.00	-11871.68	-0.65	0.63	-3.08	-0.57
	(2275.87)	(13040.87)	(1.50)	(1.06)	(2.26)	(0.88)
Observations	459	257	459	151	151	492
Mean DV	2024.52	1750.00	7.48	7.29	3.60	0.52
R^2	0.4	0.4	0.3	0.6	0.7	0.5

Note: The table presents results using data from doctors on Oladoc.pk and Marham.pk who are similar to doctors in our survey on gender, specialty and district of practice. All regressions include specialty, province and year fixed effects. Standard errors are clustered by specialty and province. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A6: Outcomes using Online Doctors

	Log(Fee)	Salary Hours	Teaching Hours	Practice per OP	Time per OP
	(1)	(2)	(3)	(4)	(5)
Exposure x Female x Post	-14.47**	6.98	4.84**	0.97	-3.01
	(7.15)	(5.27)	(2.35)	(2.34)	(3.33)
Female x Post	10.87*	-5.42	-4.75***	-0.34	1.45
	(5.43)	(4.13)	(1.67)	(1.62)	(2.35)
Female	-8.84*	-8.02*	2.75**	0.33	2.34
	(4.80)	(4.27)	(1.29)	(1.55)	(1.90)
Exposure x Post	3.96	-8.88**	-0.27	-0.34	1.02
	(4.60)	(4.25)	(1.40)	(2.39)	(2.18)
Observations	462	205	467	467	467
Mean DV	5.04	4.02	0.65	2.61	2.44
R^2	0.5	0.5	0.5	0.6	0.5

Note: The table presents results using inverse probability weights (Wooldridge, 2007). All regressions include specialty, province and year fixed effects. Standard errors are clustered by specialty and province. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A7: Inverse Probability Weights

Table A8: Selection Changes by Exposure

	House Job Large	House Job Public	Medical College Quality	Father Education	Intermediate Score (%)	Age at 1st Child	Ever Married
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Exposure x Female x Post	0.67**	0.66	-0.27	5.06	-10.62	-14.05	1.27
	(0.31)	(0.71)	(1.54)	(5.89)	(23.64)	(25.08)	(0.84)
Female x Post	-0.51*	-0.72	0.27	-4.10	9.59	13.76	-0.98
	(0.26)	(0.57)	(1.31)	(5.53)	(19.93)	(21.41)	(0.72)
Female	0.24	0.51	0.28	2.30	-34.87*	-7.90	0.65
	(0.23)	(0.51)	(1.32)	(4.32)	(19.51)	(22.67)	(0.43)
Exposure x Post	-0.52*	-0.38	-0.21	-4.35	-13.55	3.56	0.01
	(0.28)	(0.28)	(0.84)	(4.10)	(10.47)	(5.79)	(0.29)
Exposure x Female	-0.24	-0.51	-0.38	-1.97	36.26	7.26	-0.69
	(0.26)	(0.63)	(1.56)	(4.62)	(22.80)	(26.45)	(0.49)
Exposure	-0.23	-0.58	-2.02	-0.37	-16.38	-6.62	-0.41
	(0.38)	(0.77)	(1.22)	(3.93)	(22.07)	(13.22)	(0.46)
Observations	510	510	448	213	201	190	215
Mean DV	0.86	0.71	3.48	0.94	85.98	29.65	0.87
R^2	0.4	0.3	0.2	0.5	0.5	0.2	0.8

Note: The table presents ITT triple-differences that directly examine changes in selection. The outcome in columns 1-2 are if an individual completed a housejob at a public hospital or a large hospital, respectively. In column 3 the outcome is the medical college "grade" assigned by the PMDC to designate quality. In column 4 the outcome is father's education level, in column 5, it is the individual's secondary exit exam score, in column 6 it is the age at which they had their first child, and in column 7 if the individual was ever married. All regressions control for age and age-squared. Additional fixed effects included are for province, graduation year and medical specialty group. Standard errors are clustered at the province and medical specialty group level. Stars represent statistical significance. Data are from a survey of Pakistani doctors conducted by the authors. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

	Probability(Work)	Hours
	(1)	(2)
Female	-0.16***	-7.01***
	(0.05)	(2.51)
Survey	0.09**	17.63***
	(0.04)	(2.24)
Female x Survey	0.09	4.32
	(0.06)	(3.60)
Observations	647	590
R^2	0.1	0.3

Note: The table presents estimates of gender gaps in the probability of working (column 1) and in total hours worked last week (column 2) in our survey relative to in the Labor Force Surveys of Pakistan (LFS). The LFS gaps are contemporaneous, between doctors within 10 years of their predicted graduation. The survey gaps are among "survivors", observed about 20 years after graduation. All regressions include specialty, province and year fixed effects. Standard errors are clustered by specialty and province. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A9: Comparing Contemporaneous and Long-Term Gaps Among Pre-Reform Cohorts

	Log(Fee)		Hours Teaching	
	(1)	(2)	(3)	(4)
Exposure x Female x Post	-5.79*	-5.82*	6.34***	6.36***
	(3.15)	(3.20)	(1.82)	(1.82)
Observations	443	443	448	448
Mean DV	5.02	5.02	0.64	0.64
R^2	0.3	0.3	0.4	0.4
Added Control:				
Medical School Grade	No	Yes	No	Yes

Note: The table includes baseline regressions for log of fee and hours spent teaching in columns 1 and 3, and regressions which include a control for medical school "grade" in columns 2 and 4. Regressions are restricted to a common sample. Medical college grade is provided by the PMDC and is used as a proxy for medical quality. All regressions include specialty, province and year fixed effects. Standard errors are clustered by specialty and province. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A10: Controlling for Medical College Quality