

TALENT, TRUST, AND HEALTH: THE EFFECTS OF THE FIRST FEMALE PHYSICIANS¹

Tamar Matiashvili

Stanford University

October 14, 2026

[Click here for the latest version.](#)

Abstract

How does removing barriers to entry for previously excluded groups affect the productivity of high-skill professions? This paper studies the entry of female physicians into the Russian Empire's rural public healthcare system after the world's first full-term medical school for women graduated its first cohort in 1878. I digitize historical records to construct a novel panel of the medical labor force and vital statistics covering 336 districts from 1876–1910 and exploit variation in the timing of female physicians' replacement of departing male physicians in a staggered difference-in-differences design. I find that the first female physicians reduced infant and young adult mortality of both sexes, increasing natural population growth and life expectancy. To interpret these findings, I develop a conceptual framework distinguishing supply-side selection from demand-side gender concordance preferences. On the demand side, female physician entry increased clinic use and displaced untrained midwives, consistent with women shifting toward physician care when female providers became available. On the supply side, declines in young adult male mortality, despite historical evidence that men were unlikely to increase care-seeking in response to female physicians solely due to their gender, suggest positive selection among the first female entrants. In the long run, the framework predicts a rise in average quality but an attenuation of the initial quality gap; consistent with this prediction, estimated gender differences in physician value added are smaller among later graduation cohorts. Overall, the findings suggest that opening medicine to women improved health through both greater physician effectiveness and increased use of formal medical care.

¹E-mail: tamrim@stanford.edu. Professional website: www.tamrim.com. I am deeply grateful for the guidance and support of Ran Abramitzky, Petra Persson, Maria Polyakova, and Alessandra Voena. I am very grateful to the staff of the Slavonic Library of the National Library of Finland for their guidance in my search for historical data, and to Timur Natkhov for his assistance in obtaining original scans of the documents from the Russian Federation State Archive. I thank Ketevan Magradze and Tamar Mgebrishvili for excellent research assistance. I am very grateful for helpful comments by Marcella Alsan, Mark Anderson, Liran Einav, Michela Giorcelli, Amanda Gregg, Lauren Harris, Caroline Hoxby, Steve Nafziger, Isaac Sorkin, Janet Stefanov, Helena Roy, Gertjan Verdickt, and Eric Wang. This work greatly benefited from the input by seminar, conference, and workshop audiences at Northwestern University, Yale University, the University of San Francisco, Middlebury College, Ludwig Maximilian University of Munich, Stanford Institute for Theoretical Economics (Gender), WEAI Job Market Workshop, the Cliometric Society conference, the Society of Labor Economists annual conference, the Economic History Association (EHA) annual conference, and the Stanford History Department, as well as numerous internal workshops of the Stanford Economics Department. This work was supported by the EHA Graduate Fellowship, Leonard W. Ely and Shirley R. Ely Graduate Student Fellowship, and Stanford Graduate Fellowships, as well as research grants from the George P. Shultz Dissertation Fellowship, Institute for Humane Studies (grant no. IHS020558), the Stanford King Center on Global Development, EHA, the Stanford Europe Center, and Stanford's Diversity Dissertation Research Opportunity.

1 Introduction

How does the entry of previously excluded groups into high-skill professions affect the productivity of those professions? Barriers to education and employment have historically prevented many workers from entering occupations in which they might have been highly productive. Women and racial, ethnic, and religious minorities were often formally or informally excluded from universities and from high-skill professions such as medicine, law, academia, and public service. A large empirical literature has studied how the expansion of educational and labor market opportunities affected the individuals who gained access. Women provide a prominent example: expanding women’s opportunities changed their schooling, careers, and fertility (e.g., [Goldin and Katz, 2002](#); [Goldin, 2014](#)). Much less is known about how these expansions affected the productivity of occupations from which these groups had previously been excluded.

Theoretical work suggests that these productivity effects could be large. When barriers prevent talented individuals from entering occupations suited to their comparative advantage, aggregate productivity can fall substantially (e.g., [Lagakos and Waugh, 2013](#); [Restuccia and Rogerson, 2017](#); [Hsieh and Klenow, 2009](#); [Hsieh et al., 2019](#)). Yet there is little causal, micro-level evidence on whether removing such barriers raises the productivity of a profession by expanding the pool of talent from which it can draw. Further, the productivity consequences of entry need not only operate through changes in the talent pool. Opening a profession to a previously excluded group can also affect demand through preferences for demographic concordance between providers and consumers, a margin that existing Roy-style occupational choice models do not address. Few settings combine both plausibly exogenous variation in entry timing and the data needed to measure the productivity and concordance effects.

I study this question using the entry of female physicians into the Russian Empire’s public rural healthcare system following the establishment of the world’s first full-term, five-year medical school for women in 1872.² Physicians were employed by district-level governments (*Zemstvos*³) and assigned to geographically defined catchment areas, within which they practiced independently and provided near-free care. Because patients were effectively tied to their assigned physician, and only one physician practiced in a given area, this system limited both workplace discrimination and patient sorting across providers. These institutional features make the setting well suited to studying the effects of female physician entry. In the main analysis, I focus on cases in which a female physician replaced a departing male physician without an accompanying increase in total physician employment, separating changes in physician composition from contemporaneous expansions in

²While some women had previously obtained medical degrees in Western Europe, these were typically granted through exceptions or informal channels. In the United States, women’s medical colleges existed but offered only two-year programs until the 1890s.

³The Russian Empire was also one of the pioneers in this regard, one of the first countries to provide publicly-administered, near-free healthcare to large subsets of population following the establishment of *Zemstvos*, rural elected local governments. For more details on the reforms and the role of *Zemstvos*, see, for example, [Nafziger \[2011\]](#).

physician supply.

Why might the entry of female physicians affect health outcomes in this setting? I consider two mechanisms. On the demand side, patient behavior may respond to physician gender. If patients prefer gender-concordant care, the arrival of female physicians may increase healthcare utilization among women and, by extension, children. This mechanism reduces mortality if receiving care from a physician improves outcomes relative to no treatment or informal care, and therefore predicts mortality declines concentrated among women, infants and children. It also predicts a decrease in reliance on informal providers, such as midwives.

On the supply side, the larger barriers women faced may have positively selected the earliest female physicians on clinical ability, which I define broadly as a bundle of traits that may affect patient mortality.⁴ If so, female physician entry should reduce mortality among groups whose survival contemporary medicine could affect, including male patients, whose care-seeking was unlikely to increase in response to female physician entry. As barriers to entry relaxed over time, this positive selection would weaken, while access to a broader pool of talent could raise average physician quality and improve population health in the long run. These contrasting predictions allow me to distinguish between the two mechanisms empirically.

To measure how female physician entry affected mortality and health outcomes, I digitize novel district-level data on *Zemstvo* physicians by gender for the period 1876 to 1915, covering more than 330 districts of the Russian Empire with *Zemstvo* governments. To examine health outcomes, I digitize a comprehensive panel of district-level vital statistics spanning 1876 to 1910, the last year these records were published. The data include annual births by gender and detailed records of deaths by gender and age, including month-specific mortality within infancy and five-year age groups thereafter. These data enable the study of within-infancy and adult mortality outcomes that have been largely inaccessible in historical contexts. I augment these data with rich information on district-level medical labor and healthcare provision for the first thirteen years of the panel. These highly granular data allow me to investigate the channels through which female physicians affected mortality and to examine changes in reliance on informal care. I also digitize cause-of-death records by age and sex from a *Zemstvo* province, which allow me to assess whether the mortality effects are concentrated among groups for whom contemporary medicine offered greater scope for intervention.

I exploit the staggered timing of first female replacement hires across districts in a difference-in-differences design. Between 1878 and 1915, 92% of studied districts had hired their first female physician, with adoption smoothly staggered over time. For the main analysis, I focus on districts in which the first female physician replaced a departing male physician without an increase in the

⁴This prediction requires ability to meaningfully affect entry probabilities. Medical education and employment may also have depended on family resources, social connections, patronage, or political acceptability, which I do not observe. Such factors would weaken selection on ability if they are unrelated to it. If they systematically differed by gender, or were correlated with ability, their implications would be ambiguous, but they would attenuate the differences in most cases.

total number of physicians, allowing me to isolate changes in physician composition from expansions in physician supply. Therefore, the identifying variation comes from differences across districts in the timing of these first female replacement hires: I compare within-district changes in outcomes before and after female physician entry across districts treated at different times. I estimate both average treatment effects and event studies, which trace the dynamics of the effects and allow me to assess pre-treatment trends.

I find three main mortality results. First, infant mortality declined by approximately 3.8% on average. In the corresponding levels specification, this amounts to 13.7 fewer deaths per 1,000 births, or a 1.4 percentage-point decline. For context, infant mortality in 1876 was 31.4%. The effects were larger and more sustained in middle infancy, particularly between months 1–5. Second, mortality declined significantly among young men and women aged 15–54. Third, I find declines in mortality among children, although negative pre-treatment trends make these estimates less conclusive. I find no statistically significant effects on mortality among older age groups.

This pattern of mortality effects is difficult to reconcile with a purely demand-driven mechanism. Increased care-seeking driven by gender concordance would operate primarily through women, infants, and children, and therefore predict improvements concentrated among these groups. However, I also find substantial mortality declines among young men, whose healthcare utilization is unlikely to increase in response to female physician entry. Therefore, the decline in young male mortality points toward differences in physician effectiveness as an important part of the explanation.

These mortality reductions translated into improvements in population growth and life expectancy. Using district-level population counts from the 1897 Census combined with annual birth and death records, I construct measures of natural population growth. I find that female physician entry increased population growth rate by 0.11 percentage points relative to a pre-treatment mean of 1.65%. This reflects both declining mortality and weakly rising birth rates, consistent with improved survival among young adults and thus increased child-bearing years in the population. I also examine period life expectancy at birth using age- and sex-specific census population data. The entry of the first female physicians increased female life expectancy by 1.1 years (2.8%) and male life expectancy by 1.2 years (3.5%).

To understand why the mortality effects are concentrated among infants, children, and young adults, I examine variation in causes of death and the scope for medical intervention at the time. Newly digitized cause-of-death data show that these age groups accounted for a disproportionate share of deaths from causes that were potentially preventable or treatable through medical care. Infant mortality was largely driven by conditions linked to poor hygiene and feeding practices [Natkho and Vasilenok, 2023], which could be mitigated through improved maternal knowledge and basic care, consistent with contemporary accounts of female physicians’ efforts to instruct mothers [Jakobson, 1958]. Among children, mortality was dominated by infectious disease, especially smallpox, for which the quality and reliability of vaccination could materially affect survival [Veretennikova, 1984;

Chapman, 2026]. Among young men, mortality was often driven by injuries followed by infection, for which recently introduced antiseptic techniques such as sterilizing instruments and cleaning wounds to prevent bacterial infection could substantially improve survival, as emphasized by Mokyr [2005, p.94]. These techniques had become mainstream in the Russian Empire by the late 1870s and early 1880s [Breido, 1956]. Among young women, mortality was often driven by childbirth-related complications, for which the same infection-prevention practices, together with improved obstetric techniques, could reduce maternal deaths. Older adults, on the other hand, mostly died from chronic and infectious diseases for which effective treatment remained limited. This evidence suggests that female physicians were most effective in contexts where existing medical knowledge and practices could meaningfully prevent or treat the underlying causes of death.

I next examine some of the channels through which female physicians may have reduced mortality. Their entry increased clinic use and was associated with a suggestive increase in the share of births taking place in clinics. I also document large and persistent declines in the number of untrained midwives, suggesting that female patients substituted away from informal providers toward female physicians when the latter became available. This substitution provides evidence of demand-side responses, consistent with increased formal care-seeking among women in the presence of female physicians.

I then address several important threats to identification. Female physician entry was not accompanied by increases in other healthcare inputs controlled by *Zemstvo* boards, such as clinic beds or local medical budgets, making coincident improvements in healthcare provision an unlikely explanation for the mortality decline. Nor do differences in physician career stage appear to account for the results. Female physicians had higher estimated value added than male physicians from similar graduation cohorts, and linking entering female physicians to the departing male physicians they replaced yields no evidence that differences in graduation year or the age of the departing physician explain the mortality effect.

The findings are also robust to alternative samples, specifications, and identification strategies. The results are similar when I include all first female physician entries rather than restricting attention to replacements of departing male physicians. As a spatial placebo, I find no mortality effects in the urban populations of treated districts, where the female physicians in my sample did not practice. The results are also stable across alternative estimators and placebo tests, and an interrupted time series design examining within-district mortality changes around female physician entry yields similar patterns. I further implement alternative identification strategies, including a stacked design comparing female and male replacement hires occurring in the same year and comparisons of female and male physician expansions, which provide broadly consistent, though less precise, evidence.

To study the broader implications of these results, I formalize the supply and demand channels in a conceptual framework that predicts the long-run composition and quality of the physician workforce. As barriers to entry relax, the exceptional positive selection of the earliest female entrants diminishes,

so the initial female advantage in average physician quality generally attenuates over time. The long-run composition of the profession depends on underlying ability distributions and gender-concordance preferences. I first consider a benchmark in which men and women have identical latent medical-ability distributions and concordance preferences are symmetric. In this case, the long-run features gender parity in both physician shares and average quality. I then relax the equal-distribution benchmark: if women have a higher underlying distribution of medical ability,⁵ the workforce becomes majority female and average female quality is higher, while marginal quality remains equal across genders. In almost all cases, the initial gender gap in quality is predicted to shrink over time.

I examine these predictions in two ways. First, using the linked physician panel and estimating individual-level physician value added, I find that gender differences in physician value added are smaller among later graduation cohorts. Second, I use modern data to evaluate the longer-run implications of the framework. Today, in all countries formerly part of the Russian Empire, women account for substantially more than half of the physician workforce. One possible explanation is that female patients disproportionately prefer female physicians, increasing relative demand for female physicians. I provide evidence against a purely concordance-driven explanation using experimental evidence from Kissel, Matiashvili, and Roy [2026], where we show that gender concordance preferences are confined to gender-sensitive specialties. However, using data from the Republic of Georgia, I document that women comprise over 60% of practicing physicians across nearly all specialties — including non-gender-sensitive ones. This pattern is consistent with an underlying female advantage in medicine, although other forces affecting occupational sorting cannot be ruled out.

Literature Review This paper contributes to three strands of literature. First, it contributes to the literature on the misallocation of talent (e.g., Hsieh and Klenow [2009]; Lagakos and Waugh [2013]; Gollin, Lagakos, and Waugh [2014]; Böhm and Watzinger [2015]; Restuccia and Rogerson [2017]; Gradstein [2019]; López and Torres [2020]; Celik [2023]). A large body of work, largely theoretical or based on macro-level data, shows that restricting access to labor markets can generate substantial aggregate losses by preventing talented individuals from pursuing professions of their comparative advantage (e.g., Hsieh et al. [2019]; Ashraf et al. [2022]; Lee [2024]; Brouillette [2025]). Descriptive empirical work has also documented that under-representation may lead to inefficiencies (e.g., Abramitzky et al. [2024]; Iaria, Schwarz, and Waldinger [2024]). A smaller causal literature examines how demographic restrictions on access to high-skill positions affect subsequent outcomes: Besley et al. [2017] show that gender quotas in Sweden increased the competence of male politicians by crowding out less capable men. Lubczyk and Moser [2025] study female scientists’ entry into firms

⁵Formally, I assume that the female ability distribution first-order stochastically dominates the male distribution and that this dominance is preserved in upper-tail means: for any common entry threshold t , $E[A_f | A_f \geq t] > E[A_m | A_m \geq t]$. The latter condition ensures that, at a common equilibrium entry threshold, female physicians have higher average ability.

during WWII and find increases in men’s publications in these firms.⁶ This paper contributes to this literature by studying the removal of a formal barrier excluding women from a high-skill profession and measuring its consequences for physician effectiveness and population health through female physicians’ own productivity.

Second, this paper contributes to the health economics literature on how provider characteristics influence patient behavior and outcomes. A substantial body of medical research has examined gender concordance preferences between patients and physicians, with many studies reporting associations with better health outcomes [Schmittiel et al., 2009; McDevitt and Roberts, 2014; Greenwood, Carnahan, and Huang, 2018; Eggermont et al., 2018; Lee et al., 2019] and improved perceptions of care [Bertakis and Azari, 2012; Smith, 2022]. Other studies, by contrast, have found no significant effects on outcomes [Jerant et al., 2011; Krähenmann-Müller et al., 2014; Yang et al., 2016] or perceptions of care [Bertakis, 2009; Scheid and Smith, 2017; Crawford et al., 2019].⁷ These studies document associations between physician-patient gender concordance and outcomes. Causal evidence in economics has shown that racial concordance can increase care-seeking for preventive services [Alsan, Garrick, and Graziani, 2019] and gender concordance can increase care-seeking among female physicians’ female family members [Anderson et al., 2026], but there is little causal evidence of the effect gender concordance among patients on their care-seeking, where preferences may be stronger, particularly in gender-sensitive specialties such as gynecology or urology [Kissel, Matiashvili, and Roy, 2026], and in more patriarchal settings. This paper provides novel causal evidence that increasing the availability of female physicians led to higher healthcare utilization among women, shifting demand towards formal healthcare.

Finally, this paper contributes to the growing literature on the determinants of the large declines in mortality and increases in life expectancy around the turn of the twentieth century. Much of this work has focused on non-medical public health improvements, such as water filtration and sewage infrastructure, and their effects on infant and child mortality [Cutler and Miller, 2005; Ferrie and Troesken, 2008; Kesztenbaum and Rosenthal, 2017; Chapman, 2019; Alsan and Goldin, 2019; Egedesø, Hansen, and Jensen, 2020; Anderson, Charles, and Rees, 2022]. Other studies emphasize the role of medical innovations, including sulfa drugs and smallpox vaccines [Jayachandran, Lleras-Muney, and Smith, 2010; Ager, Worm Hansen, and Sandholt Jensen, 2018], as well as institutional changes such as health insurance expansions [Winegarden and Murray, 2004; Bauernschuster, Driva, and Hornung, 2020] and the creation of public health systems [Hoehn-Velasco, 2018; Marein, 2023]. In contrast, relatively little is known about the productivity of physicians during this period, particularly before the widespread use of antibiotics. Existing evidence [Anderson et al., 2020; Hollingsworth et al., 2024; Saaritsa, Simanainen, and Ristola, 2024; Liebert and Mäder, 2026],

⁶Schaede and Mankki [2026] provide related evidence from a quota favoring male entrants into primary-school teaching in Finland. Although the quota admitted academically lower-scoring men, students exposed to these teachers experienced better long-run outcomes, illustrating that observable admission credentials need not fully capture occupational productivity.

⁷See Lau et al. [2021] for a comprehensive review.

largely focuses on later periods or on changes in medical supply rather than physician productivity. Building on this literature, this paper provides new causal evidence that physicians, and female physicians in particular, influenced mortality across multiple age groups in the late nineteenth century, and shows that individual physician quality mattered for population health.

The rest of the paper is structured as follows. Section 2 recounts the history of the opening of the world’s first full-term medical school for women and the history of rural public medicine. Section 3 describes the novel data digitized for this study and presents descriptive statistics. Section 4 describes the empirical strategy and identifying assumptions. Section 5 presents the mortality effects, interprets the effect sizes in the context of existing literature, relates them to causes of death, examines the channels through which they arise, and reports robustness checks. Section 6 develops the conceptual framework, derives its long-run implications, and evaluates them using historical and modern data. Section 7 concludes.

2 Historical Background

This section describes the institutional and historical context surrounding the entry of the first female physicians in the Russian Empire. The creation of a formal medical education path for women, paired with a decentralized system of rural public healthcare, resulted in substantial variation in when and where female physicians entered the profession. This section also explains how the structure of rural public healthcare limited demand-side and workplace discrimination, which allows me to measure physician performance with relatively little friction.

2.1 World’s First Full-Term Medical School for Women

In 1872, the Russian Empire established the world’s first full-term medical school for women.⁸ In the United States, women’s medical colleges had existed since the 1850s, but they offered limited curricula and did not adopt four-year programs until 1891 [Fee and Brown, 2004].⁹ Pressure to expand access to medical training for women in the Russian Empire stemmed in part from public health concerns. Contemporary discussions emphasized the limited training of existing midwives: “there arose a need for educated midwives who could replace the ignorant traditional ones, through whose ineptitude so many lives perish and so much suffering is caused” [Abramov, 1886, p. 5]. In response, the government’s medical council surveyed the six existing male medical schools on the potential benefits of creating a course for “educated midwives.” All six endorsed the proposal

⁸By 1872, several women were already practicing medicine in the Russian Empire, having obtained degrees as exceptions at men’s universities in Zurich and Paris [Pigeard-Micault, 2007; Gosteli, 2019]. The first two women officially admitted to a medical university were from the Russian Empire, matriculating at the University of Zurich in 1867. In 1849, Elizabeth Blackwell became the first woman to earn a medical degree in the United States, graduating from Geneva Medical College in New York. However, her admission was reportedly intended as a practical joke and did not establish a broader precedent [Nimura, 2021; Blackwell, 2005, p. 27]. These early cases of women earning medical degrees from traditionally male institutions were rare and lacked systematic institutional support.

⁹By the 1890s, medical education in the United States remained largely unregulated, and very few institutions offered four-year programs at all [Fee and Brown, 2004]. Moehling, Niemesh, and Thomasson [2026] discuss how the regulation of medical education in fact led to a reduction in the share of female physicians in the US.

and recommended that women receive not only midwifery instruction but also a general medical education.

As a trial, a four-year course was launched in St. Petersburg in 1872. Yet from the outset, the original goal of producing “learned midwives” quickly gave way to a broader vision: women were instead offered a full medical education. Before the first cohort graduated, a fifth year was added, and the program was formally renamed the *Courses for Female Doctors*. The curriculum closely matched that of men’s medical schools, with the exception of courses in forensic and police medicine, fields female doctors could not legally practice in, which were replaced with more advanced midwifery instruction (p. 7). The courses were taught by the same professors who taught in the all-male Imperial Medical-Surgical Academy in Saint Petersburg.

The students were subject to entry and annual qualifying exams, with the final licensing exams at the end of the fifth year. The final exams were identical to those of the men’s schools. Members of the medical committee who attended the final exams for the first graduating cohort concluded that the women possessed all the necessary knowledge for medical practice [Abramov, 1886, pp. 9-10]. Thus, the first women physicians of the Russian Empire entered the medical workforce. The use of entry, annual qualifying, and final licensing examinations indicates that demonstrated academic ability played an important role in the trainees’ progression through education and into medical practice. This supports the premise that selection into the physician workforce was at least partly ability-based.

The courses quickly gained popularity. By 1885, nearly 400 women held the right to practice medicine in the Russian Empire [Ministry of Internal Affairs, 1885]. Between 1887 and 1897, however, the courses were shut down under a new Interior Minister who opposed them, citing the adequate existing supply of male physicians [Tarnovskaya, 1903, p. 509]. A subsequent policy shift fueled by renewed activism led to their reinstatement and expansion in 1897. This decision was driven by mounting evidence of the critical role female doctors played in treating female patients. Contemporary reports emphasized that many women, due to “shyness” or cultural norms, avoided seeking care from male physicians until their conditions had progressed significantly [Tarnovskaya, 1903].¹⁰ Further pressure came from the specific needs of Muslim women, who were often entirely unable to access medical care in the absence of female doctors.¹¹

By 1910, the number of female doctors had grown more than fivefold, reaching 2,075 [Ministry of Internal Affairs, 1911]. Their share of the medical profession rose from under 2.5% in 1885 to over

¹⁰This sentiment was not specific to the Russian Empire. A famous American obstetrician Charles D. Meigs noted in his work *Females and Their Diseases*: “All these evils of medical practice spring not, in the main, from any want of competency in medicines or medical men, but from the delicacy of the relations existing between the sexes [...] There are women who prefer to suffer the extremity of danger and pain rather than waive those scruples of delicacy which prevent their maladies from being fully explored.” (Meigs [1848], as quoted in [Reeder, 2024, pp.13]).

¹¹While an interesting case to study, as Muslim women were prohibited from seeing male doctors, very few districts with *Zemstvos* that administered public rural healthcare had sizable Muslim populations, so I am unable to pursue this interesting heterogeneity in this setting.

10% by 1910. Women thus became an increasingly vital part of the medical workforce, particularly in rural public medicine, where they held full rights to practice independently.

Contemporaries had much to say about female doctors. Abramov [1886], a contemporary publicist, in his defense of keeping the women’s medical courses open, cited an 1879 *Zemstvo* survey reporting that “many *Zemstvos* proclaimed the advantages of hiring specifically female doctors to serve rural populations.” Newspapers also frequently reported on the successes of female physicians, some acquiring near-mythic reputations: “[A female *Zemstvo* physician in Kursk province] became so popular that the peasants called her “the Merciful One”, sent by the Lord himself” [Suschinsky, 1883, p. 21]. Abramov also argued that female doctors were generally superior to their male counterparts: “the female doctor turns out to be more effective, more hard-working, more attentive to the patients than the male physicians, and so her practice is wider than that of male physicians” (pp. 21-22). A newspaper article wrote about a female *Zemstvo* doctor in Tula province: “[she] obtained a reputation of an excellent surgeon, performing surgeries [...] wholly unheard of in [her] district.” [Suschinsky, 1883, p. 21]

Female medical students came from a wide range of social backgrounds. Using data from the first eight years of admissions, Suschinsky [1883] reports that approximately 75% were wives or daughters of men in civil service, such as bureaucrats, educators, and physicians, as well as guild and small merchants, military personnel, and honorary citizens. Children of peasants were also represented, though in much smaller numbers (Appendix Table A.1.1). These distributions are broadly comparable to the social origins of male *Zemstvo* physicians, as shown in Frieden [1981, Table IV.C.3].

2.2 *Zemstvo* Medicine: Rural Public Medicine of the Russian Empire

In 1864, the imperial government introduced the *Zemstvo* system of local self-government as part of a broader set of progressive administrative reforms following the emancipation of the serfs in 1861. *Zemstvos* were elected district and provincial councils responsible for providing crucial local public goods, including primary education, roads, and rural healthcare [Imperial Chancellery, 1867]. Because the central government did not directly manage most local services, the development and organization of medical care in rural areas was largely determined by these local institutions. The healthcare provision programs, which were virtually free, aimed to bring modern medical care to the peasantry.¹² Districts were divided into catchment areas, each typically served by one doctor responsible for a clinic and the entire local population (see Appendix Figure A.3.1 [Frenkel, 1913]). These doctors provided both inpatient and outpatient care at the clinic and often traveled long distances to treat patients in remote villages. Their responsibilities ranged from clinical treatment to health education and preventive care, particularly in areas such as hygiene and infant feeding.

¹²Individual *Zemstvos* differed in their payment structures, with some asking for small payments for clinic stays. Almost all charged residents of other catchment areas or districts. For examples of patient payment structures, see Appendix Table A.2.1.

They were usually assisted by several nurses, whose role was to carry out doctors’ instructions and administer medicines during follow-up home visits.

Three features of the *Zemstvo* medical system make it a compelling setting to study the impact of female physician entry. First, patients were effectively tied to their assigned catchment-area physician: in the vast majority of cases, access to near-free public care required visiting that physician, while private alternatives were prohibitively expensive for most peasants. Even if the district allowed going to a physician in a different catchment area, that would often require traveling long distances. This limited the scope for patient self-selection across physicians and instead shifted variation to the margin of whether patients sought care at all or relied on informal providers. Second, because each catchment area was typically served by a single physician, practitioners, both male and female, operated with substantial autonomy, limiting workplace discrimination. Third, districts were divided into a small number of catchment areas (median of four), so each physician was responsible for a large share of the local population, making changes in physician composition consequential for aggregate district-level health outcomes.¹³

While serving as a *Zemstvo* doctor was demanding, it came with notable benefits. Chief among them was financial: these positions paid over twice as much, about 1,000 rubles a year on average,¹⁴ compared to typical government jobs, which paid under 440 rubles.¹⁵ They paid several-fold more than midwife and nurse postings [Zhbankov, 1891]. They were also highly prestigious. From the start, *Zemstvo* practice was seen as a form of “service to the people,” a calling embraced by many men and, later, women. Doctors prepared themselves “to fight the ignorance of the popular masses and to change the entire aspect of their world view” [Frieden, 1981].

Zemstvo practice also served as an important career stepping stone. Among the 1903-1909 *MD* cohorts¹⁶ living in *Zemstvo* provinces, more than 47% of men and almost 52% of women entered *Zemstvo* service in the first three years after their graduation.¹⁷ Meanwhile, half of all physicians newly hired by *Zemstvos* between 1889-1910 had 5 or fewer years of experience. These comparable

¹³The empirical analysis is conducted at the district rather than catchment-area level because vital statistics are available only at the district level. The estimated effects therefore capture the district-level consequences of introducing a first female physician even though she directly served one catchment area. Patients could seek care outside their assigned catchment area, but almost all *Zemstvos* charged residents of other catchment areas or districts, limiting this margin of patient sorting; see Appendix Table A.2.1.

¹⁴Available wage data suggest that gender-based pay differences were uncommon. Zhbankov [1891] reports physician salaries by district in 1888. Among the 50 districts employing a female physician for which wages are reported, 38 did not report different salaries by gender, and 12 reported male-favoring wages.

¹⁵Calculations based on Table 2.4 of Frieden [1981].

¹⁶I start in 1903, because physician-level data are available starting in 1889, and the first graduation of the newly reopened female medical school took place in 1903. The previous most recent graduation year was 1887.

¹⁷Among all physicians practicing in these provinces in 1889-1910, 24% of female and 20% of male physicians were *Zemstvo* physicians, suggesting that these jobs likely drew similar types of male and female physicians. A gap existed instead in the shares of physicians pursuing military and police-type jobs (~ 13% men vs 0% women, largely due to institutional limitations) and private practices (26% men vs 44% women). These calculations are based on the data processed for Matiashvili [2026], which exploits the physician-level data in a movers’ design to estimate physician value added and explore the characteristics of high and low-value-added physicians of this era.

shares suggest that both men and women actively pursued *Zemstvo* positions upon graduation, easing concerns about negative selection into the job type.

Hiring for *Zemstvo* physician positions typically followed a formal process of vacancy postings, applications, and candidate evaluation. Appendix Figure A.4.1 shows an example of a 1910 vacancy advertisement published in the weekly newspaper *Russian Doctor* [Vladislavlev and Podvisotskii, 1910]. Applicants responded to such postings by submitting their credentials. Application packets typically included educational and professional records and, occasionally, reference letters, suggesting that previously observed ability played a role in hiring decisions. These were submitted to *Zemstvo* boards, which deliberated on candidates before making a final decision. Because vacancies arose from physician turnover and were filled through this decentralized process, the timing of female physician entry was often driven by local staffing needs. This makes it unlikely that systematic differences in underlying health trends were correlated with the timing of the first female physician hire.

Finally, understanding the scope of medical knowledge at the time is important for understanding how physicians could have affected population health in this historical setting. While medicine had not yet reached the modern scientific model, important advances had been made in infection prevention through antiseptic and aseptic methods, which were important for preventing wound infections, especially among young adults. Obstetric practices taught in medical schools aided childbirth. High infant mortality, driven by poor feeding and hygiene practices [Natkhov and Vasilenok, 2023; Gilyarovskiy, 1866], made the educational role of physicians particularly important. A large share of child mortality was accounted for by infectious diseases. While vaccines did not exist for most diseases, a powerful smallpox vaccine did exist, if administered correctly. By contrast, most older adults perished from chronic or poorly understood diseases. These differences imply that physician quality could plausibly affect mortality outcomes in some domains but not others. Further discussion of causes of death is provided in Section 5.3.

3 Data and Descriptive Statistics

To study how the entry of female physicians affected health outcomes, I collected, digitized, and assembled a novel district-by-year panel that combines detailed information on the physician labor force with rich vital statistics and measures of healthcare provision, covering 336 *Zemstvo* districts in 35 provinces.¹⁸ The main outcome panel spans 35 years, from 1876 to 1910: physician and vital statistics data are available throughout this period, while most other healthcare provision outcomes are available from 1876 to 1888. I additionally digitize physician data through 1915, which I use to distinguish districts that had not yet hired a female physician by 1910 but did so by the end of the observable period, from districts for which no female physician hire is observed. All outcome datasets begin in 1876, two years before the first female physicians graduated in 1878. In this

¹⁸There were a total of 373 districts in these provinces in 1876. However, 37 districts did not have tabular data in the *Zemstvo Medical Time Series* due to a lack of record-keeping. These districts are not part of the analysis.

section, I describe these data sources in detail and explain how they allow me to identify both the timing and location of first female physician entry.

3.1 Medical Personnel

Female Physician Entry into *Zemstvo* Practice The central variable of interest is the treatment indicator for the entry of the first female physician into a district *Zemstvo*’s employ. To construct this measure, I draw on two newly digitized historical sources. The first is the *Zemstvo Medical Collection*, a seven-volume compendium compiled in 1890 by prominent *Zemstvo* physician Dmitrii Nikolaevich Zhbakov [Zhbakov, 1891]. Prepared for the annual Pirogov Doctors’ Conference to commemorate 35 years of *Zemstvo* medicine, the Collection documents the medical histories of provincial and district *Zemstvos* from 1865 to 1888, covering nearly the entire universe of *Zemstvo* districts. The compendium includes district-level time-series tables reporting annual medical statistics, such as the number of physicians (with female physicians recorded separately when present), nurses, midwives, clinic beds, inpatient and outpatient visits, smallpox vaccinations, and other metrics (see Appendix Figure B.1.1). While many district-years lack information on specific indicators, such as clinic or home births, physician counts are nearly complete, and fully complete from 1875 onward. The series ends in 1888.

The second data source used to construct the treatment indicator for the entry of the first female physician into a district is the annual *Medical Directories of the Russian Empire*, which included all registered practicing physicians. I use volumes from 1890 to 1916 (e.g., Ministry of Internal Affairs [1890]), which, beginning in 1890, reported physicians’ practice locations. Licensed practicing physicians were required to report their name, birth year, academic title, graduation year, job title, and residence location each year. While some physicians may not have lived in the same village in which they practiced, the analysis is conducted at the district level, and given the large geographic distances involved, residence provides a reliable measure of district-level practice location.¹⁹ Because the directories were typically published in the first quarter of the year, they effectively describe the medical workforce from 1889 to 1915.²⁰ Each edition was divided into three sections: Male Doctors, Female Doctors, and Late-Reporting Doctors.²¹ I digitize all three sections (see Appendix Figure B.2.1 for illustration), harmonize location names, and assign villages to districts where relevant. I treat the three sections as a single physician roster for each directory year: physicians listed as late-reporting are assigned to that directory year and included in district physician counts, same as physicians listed in the main Male Doctors and Female Doctors sections. I then construct a district-by-year panel of female physician counts and append it to the panel derived from the *Zemstvo*

¹⁹In addition, housing was often included as an in-kind benefit in district *Zemstvo* contracts, given the relative scarcity of rental property. See an example job advertisement in Appendix Figure A.4.1. Because housing was provided by the specific district *Zemstvo*, it would have been located within that district.

²⁰The 1915 volume was never published, so I do not separately observe the 1914 medical workforce. Physicians who first appear in the 1916 volume are therefore assigned an entry year of 1915.

²¹Physicians who did not submit their information by the deadline were listed in the appendix “Late-Reporting Doctors.” As discussed in subsection 5.4, female physicians were considerably less likely to report late than male physicians.

Medical Collection. In Appendix B.3, I assess the continuity of the resulting physician series around the transition from the *Zemstvo Medical Collection* to the *Medical Directories*.

I define first female physician entry as the first year a female physician appears in a district’s *Zemstvo* employ. Between 1878 and 1915, 310 out of 336 districts (92%) employed at least one female physician. During the 1878–1910 outcome period, 158 districts (47%) employed multiple female physicians simultaneously at some point. However, second and subsequent female physician hires were concentrated toward the end of the outcome period, and the average cumulative duration of simultaneous multiple-female-physician presence was only 2.7 years. This provides limited variation with which to separately estimate the effects of additional female physicians. I therefore focus on the timing of the first female physician hire as the primary treatment variable.

Male Doctors Accurately measuring the number of male physicians is central to this study. It allows me to identify first female physician hires through 1915 in which a woman replaced a departing male physician. The empirical strategy is explained in greater detail in Section 4. This variable also offers a fuller picture of healthcare provision at the time, supports additional robustness checks, and constitutes an important part of the total physician number control included in the baseline specifications. I digitize male physician counts from the *Zemstvo Medical Collection* for 1876–1888 and supplement these with geolocated data from the *Medical Directories of the Russian Empire* for 1889–1915, following the procedure described above for female physicians. In 1876, the median district had three male *Zemstvo* physicians; by 1910, this number had doubled to six. Female physicians were always in a minority in any district.

Midwives Midwives, both semi-trained and untrained, including traditional village midwives, were commonly employed by local *Zemstvos* to assist with childbirth. The vast majority of *Zemstvos* hired such women. Data from the *Zemstvo Medical Collection* (1876–1888) provide systematic information on their employment. Of the 336 districts in the sample, 324 reported midwife counts. Among these, the median number of midwives per district was two, and 81% of district-years recorded the presence of at least one midwife.

As noted above, concerns about limited training among existing midwives and their ability to manage difficult deliveries was one of the motivating beliefs behind the establishment of the female medical school.²² At the same time, contemporary accounts suggest that many women who were reluctant to seek care from male physicians instead relied on the midwives. One contemporary observer wrote that “women [...] do not seek help in time [...] owing to a natural sense of modesty before a male doctor, and instead endure [...] with a midwife, whose knowledge is insufficient to bring real benefit [...]” [Suschinsky, 1883]. Female physicians therefore offered an alternative source of care that could induce women to shift from midwives toward physicians, a substitution margin I

²²Historical evidence from early twentieth-century Sweden suggests that the expansion of professional midwifery reduced maternal mortality [Lorentzon and Pettersson-Lidbom, 2021]. However, contemporary complaints about midwives in the Russian Empire frequently emphasized their limited training, suggesting that this group differed substantially from the trained professionals studied in other settings.

examine directly in Section 5.4.

Nurses and Other Lower-Level Personnel In addition to physicians and midwives, the *Zemstvos* employed several types of nurses: male, female, and, sometimes, nurse-midwives, who had more training than the midwives described above but primarily served in nursing roles. Most nurses were locally trained and primarily implemented physicians’ instructions.

Hiring and firing decisions for these lower-level personnel were either directly administered by the physicians in charge or were made with physicians’ recommendations [Zhbankov, 1891]. 333 districts reported male nurse counts, with the median number of male nurses equal to ten, and 98.9% of the district-years reported having at least one male nurse. Female nurses were less common; only 5% of the district-years reported having any female nurses. 323 districts reported nurse-midwife counts, but the median number of these workers was 0, and only 26% of district-years employed at least one nurse-midwife.

Zemstvos also employed smallpox vaccinators who helped the doctors and nurses vaccinate the population. 224 districts reported hiring these workers, and those that did had a median number of 4 vaccinators. 68% of these district-years reported at least one vaccinator.

3.2 Vital Statistics

All data on the number of births and deaths by age group and sex come from the Interior Ministry’s annual publication of the *Statistics Chronicles of the Russian Empire*, which published *The Movement of Population in European Russia* annually in 1876–1910, covering 50 provinces (with ca. 500 districts) of the Russian Empire, including all *Zemstvo* provinces (e.g. [Central Statistical Committee of the Ministry of Internal Affairs \[1890\]](#)). I digitize these data (Appendix Figure B.4.1 for illustration). To my knowledge, this is the first paper to digitize the complete panel of these vital statistics. I additionally digitize district-level age- and sex-specific population counts from the 1897 Population Census, which I use to benchmark age-specific mortality and to construct population-growth and life-expectancy measures described below.

An important limitation of the vital statistics data is that they are based on clergy records of vital events. Clergy of all denominations were required to maintain metric books following Peter the Great’s 1722 edict [[Imperial Chancellery, 1830](#)]. Although the clergy were officially instructed to record events such as stillbirths and deaths before christening, registration of such cases ultimately depended on parental discretion. As a result, death rates within the first month of life may be undercounted. Importantly, because these data were compiled independently of the medical system, they are not mechanically affected by changes in physician staffing. While some measurement error is likely, particularly in volatile years, it should be uncorrelated with treatment. These statistics were highly regarded by contemporary statisticians and formed the basis for all official analyses of the period.

Infant Deaths Infant deaths are defined as deaths between the ages of 0 and 1 in a given year, and the infant mortality rate is measured as deaths before age one per 1,000 births in the same calendar year. Because births and infant deaths are observed annually rather than by birth cohort, this measure is an approximation to the cohort infant mortality rate. Infant mortality was extraordinarily high: 29.4%, on average, across the sample, compared with about 0.55% in the United States today. Within the time frame, I also observe deaths in the first month of life, 1–2 months, 3–5 months, and 6–12 months.

Childhood Deaths I define childhood deaths as those occurring between ages 1 and 14.²³ Due to the lack of yearly population counts by age group and sex, I focus on trends in death counts rather than mortality rates. I use the 1897 Census population to benchmark the magnitude of these outcomes around the middle of the sample period. Mortality rate among children ages 1–14 in the analysis districts in the census year was approximately 2.9%. Importantly, estimates of female physician entry on early childhood deaths may be biased upward due to survival bias: if female physicians improve infant survival among less healthy newborns, this may raise the number of deaths in later age groups in nearby years. As a result, any mortality-reducing effects of female physician entry, especially on childhood mortality, may be understated in this framework.

Youth and Adult Deaths I define youth and early adulthood deaths as those occurring between ages 15 and 54, and older adulthood deaths as those aged 55 and above. Age-group distinctions were motivated by both the structure of the data — for instance, in some years, all deaths between ages 15 and 34 were reported together — as well as historical patterns of life expectancy. Conditional on surviving to age 5, life expectancy was approximately 55 years, according to [Novoselskii \[1916, p.120\]](#). I therefore treat those aged 55 and above as older adults in this setting. Using the 1897 Census population and corresponding death counts, the annual mortality rate was approximately 1.0% among those ages 15–54 and 5.1% among those ages 55 and above.

Causes of Death To interpret heterogeneity in the mortality effects across age groups and sex, I additionally digitize cause-of-death data from an 1880 report for Kostroma Province, one of the *Zemstvo* provinces in the analysis sample [[Ivanov, 1883](#)]. The report provides deaths by cause, age group, and sex. To match the setting studied in the main analysis, I restrict these data to rural areas and exclude cities. Unlike the annual vital statistics described above, these data are available for a single province and year and are therefore not used as outcomes in the difference-in-differences analysis. Instead, I use them descriptively to characterize the causes of death faced by different demographic groups and the extent to which those causes were potentially responsive to contemporary medical care. Appendix Table [A.5.1](#) reports the resulting distributions.

Population and Life Expectancy The 1897 Population Census, the only population census conducted in the Russian Empire, was carried out in January 1897 and published in separate province-level volumes. I digitize district-level total and age-specific current population counts by

²³Separating childhood into ages 1–4 and 5–14 yields very similar results for each group, so I combine them for parsimony.

gender.²⁴ I use these data together with the annual vital statistics to reconstruct natural population growth and an approximate annual period life expectancy at birth for districts in my sample.²⁵

3.3 Healthcare Provision

Using the *Zemstvo Medical Collection*, I construct a district-by-year panel of healthcare provision, utilization, and innovation measures. These data allow me to examine the impact of female physicians on both the intensive and extensive margins of healthcare delivery.

First, I digitize data on childbirths in rural *Zemstvo* clinics. While childbirth traditionally occurred at home, institutional deliveries were a relatively recent innovation aimed at improving outcomes for both mothers and infants. I then examine the share of all births, with this denominator taken from the vital statistics, that took place in clinics. In 1876, no district reported any childbirths in clinics. By 1888, 25% of districts reported some, albeit a small share, of births in clinics.

Second, I digitize the number of patients who were treated at least overnight in the clinics (inpatient visits). These were usually pre- and post-surgery patients, but also included laboring mothers. Highly infectious patients were sometimes treated in the same clinics, but would often have separate wards to avoid the spread of infections.

I also digitize data on clinic beds and *Zemstvo* medical budgets over this period. Clinic capacity, budget allocations, and physician hiring were all determined by *Zemstvo* boards. These data therefore allow me to test whether other dimensions of medical provision changed concurrently with female physician entry, which could otherwise cause me to misattribute their effects to the physician entry.

Together, these newly digitized data enable a rare examination of the effects of physician entry on age- and sex-specific mortality, as well as on healthcare provision and take-up. To my knowledge, no other historical setting from this period offers data of comparable detail on mortality, healthcare provision, and the physician labor force.

3.4 Descriptive Statistics

When and where did the female physicians enter? Figure 1 maps the spatial distribution of the first female physician hires over time. Figure 1a shows that female physicians were first hired by four district *Zemstvos* in 1878, the first year female physicians graduated. Following early successful adopters, female physicians began to be hired across the country. Iulia Kviatkovskaia, a female physician who graduated in 1886, wrote in her memoirs (Kviatkovskaia [1937, p. 82]):

"We had to overcome great difficulties to earn our place among the male doctors, who were, by and large, our primary opponents. Great service in this regard was rendered

²⁴I use exact population counts as reported on the day of the Census, rather than permanent population counts. For example, for Vladimir province, I digitize the first column of Table 1 (starting on page 2) of [Central Statistical Committee of the Ministry of Internal Affairs \[1904\]](#).

²⁵See Appendix C.3 for more detail on the construction of this measure.

to us - female doctors as a whole - by the first women, [...] who were accepted into Zemstvo service [...]. Thanks to the fact that these women doctors upheld high standards of knowledge and in no way fell short of the men, other Zemstvos gradually began to admit women into service without hesitation."

As Figure 1 shows, the hiring of the first female physicians was geographically dispersed. By 1910, when my outcome data end, 85% of districts had hired at least one female physician; by 1915, this share had reached 92%. For the empirical analysis, I restrict the sample to districts whose first female physician hire by 1915 replaced a departing male physician, as described in Section 4.2. Districts with no observed female physician hire through 1915 and districts whose first female physician hire coincided with an expansion in the physician workforce are excluded from the main analysis sample. Figures 1d–1f show the geographic diffusion of entry within the analysis sample through the end of the mortality outcome period. As Appendix Figure B.5.1 illustrates, most years during this period saw at least one new district hire its first female physician.

Summary Statistics Table 1 reports summary statistics for the analysis sample. Panel A focuses on physicians and vital statistics, which are available for the entire 1876–1910 period. A median district in this sample hired its first female physician in 1893 and had four physicians. The table highlights important features of vital events in the late Russian Empire. Between a quarter and a third of all births resulted in death within the first year of life, with roughly one-quarter of these infant deaths occurring in the first month of life. Infants and children aged 0–14 together accounted for two-thirds of all deaths in a typical year. Male deaths were more concentrated in infancy and childhood than female deaths.²⁶ Panel B reports healthcare provision metrics available for 1876–1888. A median district had two midwives, no nurse-midwives, ten male nurses, no female nurses, and four smallpox vaccinators. A median district had 42 clinic beds and treated 525 inpatients annually, 5% of whom died during their stay. It recorded no clinic births and had an annual medical budget of 18,000 rubles (equivalent to $\sim$ \$300,000 in 2026 US dollars).

4 Empirical Strategy

I begin by developing a conceptual framework that distinguishes two mechanisms through which replacing a male physician with a female physician may affect health outcomes: changes in patients' demand for care and differences in physician effectiveness. The framework generates testable empirical predictions that guide the analysis. I present an empirical strategy to estimate the effects of first female physician entry using a staggered difference-in-differences design that exploits variation across districts in when a departing male physician was first replaced by a first female physician. Finally, I discuss the assumptions required for a causal interpretation.

²⁶The slightly changing sample sizes reflect oddities of the original data source. In some cases, book binding covered columns, which are recorded as missing data. All deaths were reported in a separate column that was never missing and thus has the maximum number of observations.

4.1 Conceptual Framework and Empirical Predictions

Why might the entry of female physicians affect population health outcomes, and how can we distinguish between competing mechanisms? I focus on two channels that motivate both the empirical design and the organization of the results, summarized in Table 2. While other mechanisms may also have contributed to the documented mortality effects, historical accounts suggest that these two channels were particularly important. Moreover, they generate distinct empirical predictions that I can test relatively directly, including in settings where alternative mechanisms are unlikely to operate.

Demand-side concordance. Modern medical literature (e.g., [Varadarajulu, Petruff, and Ramsey \[2002\]](#); [Nolen et al. \[2016\]](#)) and contemporary historical documents (e.g., [Suschinsky \[1883\]](#); [Veretenikova \[1984\]](#)) emphasize that patients may prefer gender-concordant care. If female patients are more willing to seek care from a female physician than from a male physician, whether due to modesty, cultural norms, or trust, then the arrival of female physicians increases female healthcare utilization. As long as the treatment effect of physicians is greater than that of midwives or no providers, this channel predicts mortality improvements concentrated among women and, through maternal behavior, among infants and young children. It also predicts a decline in reliance on informal providers such as untrained midwives, who served as a gender-concordant but lower-quality alternative to male physicians.

Supply-side selection. If entry and persistence in medicine were increasing in latent medical ability, the greater barriers faced by women imply that the earliest female physicians would be selected from farther into the upper tail of the ability distribution than incumbent male physicians. Throughout, I use broad physician ability to mean latent clinical productivity, i.e., the set of physician attributes and behaviors that translate into better patient outcomes. These attributes include medical knowledge and technical skill, as well as effort, attentiveness, adherence to effective practices, and other dimensions of physician behavior that may positively affect patient outcomes. To the extent that physician ability translated into improved patient survival, this channel predicts mortality declines across all patient groups whose survival contemporary medicine could affect, regardless of gender.

These two mechanisms generate distinct empirical predictions, summarized in Table 2. The displacement of untrained midwives tests the demand-side channel. Because midwives offered women a gender-concordant, but lower-quality, alternative to male physicians, their displacement would suggest substitution toward formal care once female physicians became available. This displacement could reflect either realized declines in midwife use, or changes anticipated by local governments at the time of hiring. Supporting evidence also comes from changes in the share of births taking place in clinics. I therefore test the demand-side channel using changes in patterns of care rather than female, infant or child mortality, which may also reflect supply-side selection, supply-side concordance,²⁷ or female physicians’ more intensive midwifery training.

²⁷Female physicians more effectively treating female and pediatric patients.

For the supply-side channel, the primary evidence comes from changes in *young adult male mortality*. Contemporary evidence from female physicians working in urban hospitals, where patients could choose their providers, suggests that men were unlikely to increase care-seeking in response to female physician entry: “men did not seek care from us” [Kviatkovskaia, 1937, p.81]. Similarly, data from St. Petersburg city public healthcare show that female physicians treated fewer adult male ambulatory patients per physician than did male physicians, despite having similar total ambulatory caseloads [Abramov, 1886].²⁸ A decline in young adult male mortality following female physician entry would therefore be difficult to explain through increased male care-seeking, supply-side concordance, or differences in midwifery training, and would instead point toward greater clinical effectiveness of the early female cohorts.

I formalize these channels in a simple model in Section 6, where I also derive long-run predictions for the equilibrium composition and quality of the physician workforce.

4.2 Primary Identification Strategy

Using a staggered difference-in-differences design, I estimate the effects of female physician entry on mortality rates, as well as on healthcare utilization and provision outcomes, such as the share of births in clinics and the number of midwives. I compare within-district changes in outcomes before and after female physician entry across districts that experienced entry at different times.

I focus on districts whose first female physician hire is observed through 1915, excluding districts for which no female physician hire is observed over this period. I further restrict the sample to cases in which a female physician replaced a departing male physician.²⁹ The identifying variation thus comes from district-years in which the total number of *Zemstvo* physicians does not increase between $t - 1$ and t , but a female physician appears for the first time in year t .³⁰ I interpret these events as cases in which a female physician filled a vacancy left by a departing male physician, rather than expansions in medical staffing. Because vacancies are not directly observed in the earlier data, I infer replacements from non-increasing physician counts and the appearance of a female physician.

An ideal experiment would compare districts hiring male versus female replacement physicians in the same year. This comparison is not feasible for the full sample because data on male replacement vacancies are unavailable for the 1876–1888 period. Because 46% of first female physician entries had occurred by 1888, restricting the analysis to districts with first female entry after 1888 would substantially reduce an already limited sample. Nevertheless, as a robustness check, I implement a stacked design comparing female and male replacement events in this restricted sample and

²⁸Female and male physicians treated 3,327 and 3,448 ambulatory patients per physician, respectively. Across ambulatory and home care combined, the corresponding numbers of adult male patients were 549 per female physician and 724 per male physician.

²⁹In practice, I classify an entry as a replacement when a female physician appears for the first time in year t and the total number of physicians in the district does not increase between $t - 1$ and t .

³⁰Because the 1915 *Medical Directory*, which would have reported the 1914 physician workforce, was never published, I classify a first female physician entry observed over this interval as a replacement if the total number of physicians does not increase between 1913 and 1915, and assign the entry year to 1915.

re-estimate my primary specification using the corresponding districts, female-entry cohorts, and observation period. The two approaches yield similar estimates that are also consistent with the full-sample results, although considerably less precise. This comparison provides suggests that the main findings are not driven by the choice between these identification strategies.

I define treatment as an absorbing state beginning in the first year a district hires a female physician. Therefore, the treatment indicator remains equal to one in subsequent years, even if the initial female physician leaves and the district temporarily employs no female physician. This definition corresponds to the causal effect of the first female physician hire, and allows the initial entry to have persistent effects on medical practice and patient behavior after the entering physician departs. In some districts, the first female physician left and another female physician was hired only after an intervening period without one. As a robustness check, I re-estimate the results after excluding post-treatment district-years in which no female physician is present.

The analysis sample therefore consists of 107 districts whose first female physician hire by 1915 was a replacement hire. For mortality outcomes, which are observed from 1876 to 1910, 101 of these districts were treated by 1910. The remaining six districts first hired a female physician between 1911 and 1915 and are untreated throughout the mortality outcome period. For healthcare provision outcomes, which are observed from 1876 to 1888, 49 of the 107 districts were treated by 1888. The remaining 58 had not yet been treated by 1888 and are untreated throughout this shorter outcome period.

For count outcomes, such as deaths in post-infancy age groups³¹ or the number of midwives, I estimate treatment effects using a Poisson pseudo-maximum likelihood estimator with high-dimensional fixed effects (Correia, Guimarães, and Zylkin, 2020). This method accommodates count outcomes that include zeros and permits the inclusion of multiple fixed effects. The coefficients have a semi-elasticity interpretation.

For outcomes measured in rates, I estimate the following equation:

$$Y_{dt} = \alpha + \beta \mathbf{1}(t \geq \text{first female doctor entry year})_{dt} + \delta_d + \mu_t + \Psi X_{dt} + \epsilon_{dt} \quad (1)$$

Similarly, for outcomes measured in counts, I estimate:

$$E[Y_{dt} \mid X_{dt}, \delta_d, \mu_t] = \exp(\alpha + \beta \mathbf{1}(t \geq \text{first female doctor entry year})_{dt} + \delta_d + \mu_t + \Psi X_{dt}) \quad (2)$$

where d indexes district and t indexes year. X_{dt} includes the total number of physicians. By construction, districts whose first female physician hire coincided with an expansion in the physician workforce are excluded from the analysis sample.³² Expansions in subsequent years and contractions

³¹For these age groups, meaningful mortality *rates* cannot be constructed due to a lack of age- and gender-specific annual population counts.

³²The identification strategy requires the total number of physicians not to increase between $t - 1$ and t , while a female physician is recorded in t but not in $t - 1$.

in all years remain in the sample. δ_d and μ_t represent district and year fixed effects, respectively. Standard errors are clustered at the district level.

The treatment variable is $\mathbf{1}(t \geq \text{first female doctor entry year})_{dt}$, which equals one from the first year a district employs a female physician onward. For outcomes estimated in levels β gives the change in the outcome associated with female physician entry. For log-transformed outcomes, β represents a log-point change. For count outcomes estimated using PPML, the proportional treatment effect is $\exp(\beta) - 1$.

I construct event studies to examine pre-treatment trends and estimate dynamic treatment effects. Because treatment timing varies across districts, the panel is unbalanced in relative time. For instance, 15-year effects can only be estimated for districts that hired a female physician by 1895, while 10-year pre-trends can only be estimated for those hiring after 1885. Relative time is defined using first-hire dates observed through 1915, so districts whose first female physician hire occurs after the end of the relevant outcome panel contribute only pre-treatment observations for that outcome. The equation takes the following form for rates regressions:

$$\begin{aligned} Y_{dt} = & \alpha + \sum_{j=2}^{J-1} \gamma_j (\text{Lead } j)_{dt} + \gamma_J (\text{Lead } \leq -J)_{dt} \\ & + \sum_{i=0}^{I-1} \tau_i (\text{Lag } i)_{dt} + \tau_I (\text{Lag } \geq I)_{dt} \\ & + \delta_d + \mu_t + \Psi X_{dt} + \epsilon_{dt} \end{aligned} \quad (3)$$

and, similarly, for count outcomes:

$$\begin{aligned} E[Y_{dt} \mid X_{dt}, \delta_d, \mu_t] = & \exp \left(\alpha + \sum_{j=2}^{J-1} \gamma_j (\text{Lead } j)_{dt} + \gamma_J (\text{Lead } \leq -J)_{dt} \right. \\ & + \sum_{i=0}^{I-1} \tau_i (\text{Lag } i)_{dt} + \tau_I (\text{Lag } \geq I)_{dt} \\ & \left. + \delta_d + \mu_t + \Psi X_{dt} \right) \end{aligned} \quad (4)$$

where j indexes years prior to first female physician entry and i indexes years following entry. The coefficients γ_j capture differences in outcomes prior to female physician entry, while τ_i capture dynamic post-entry effects. Due to limited observations in the tails of the event-time distribution, I bin extreme event times at J leads and I lags.³³ The omitted category is the year immediately

³³For the main mortality outcomes, I set $J = 10$ and $I = 15$, so that the endpoints contain event times less than or equal to -10 and greater than or equal to 15 , respectively. The values of J and I differ for healthcare provision outcomes because of their shorter temporal coverage.

prior to the district’s first female physician hire. I include time-varying controls for the number of physicians, though results are unchanged when these controls are excluded. Standard errors are clustered at the district level. All estimates are unweighted, but results are robust to weighting by the 1897 Population Census district population, the number of annual births, and the inverse of the number of physicians.

Causal interpretation of the difference-in-differences estimates requires that, absent female physician entry, districts experiencing entry at different times would have followed similar outcome trends. I examine this parallel-trends assumption using event-study estimates throughout Section 5, as well as a series of associated robustness checks in Section 5.5. Because the baseline two-way fixed-effects estimator may also be sensitive to treatment-effect heterogeneity under staggered adoption, I verify that the results are similar using alternative estimators designed for such settings.

A related concern is that male physicians may have disproportionately exited districts following temporary adverse health shocks. This would have resulted in female physician entry to coincide with subsequent mean reversion in mortality. If such a shock were concentrated in the year preceding entry, the earlier pre-treatment event-study coefficients would be negative relative to the omitted $t = -1$ period. I find no such pattern. To examine this concern more directly, I additionally compare mortality in $t = -1$ with the average of the earlier pre-treatment periods. Section 5.5 shows that mortality immediately before entry is statistically indistinguishable from this earlier pre-treatment average.

I then focus on two additional threats to interpretation that are particularly relevant in this context. The first is that the estimated effects may reflect differences in training between entering female physicians and the male physicians they replaced. Female physicians were, on average, more recently trained, raising the possibility that the estimated effects capture differences in medical training correlated with time.

Several facts mitigate this concern. Physician turnover in *Zemstvo* districts was high. On average, a new physician was hired approximately every two years, and most new hires had graduated within the previous five years. Therefore, female physician entry occurred alongside substantial inflows of recently trained male physicians, limiting actual differences in training recency between entering female physicians and the male physicians they replaced. I assess this concern directly in Sections 5.5 and 6.3 by comparing the characteristics of entering female and departing male physicians, and by estimating physician value added within graduation cohorts.

Male and female physicians also received largely comparable formal training. Female physicians were taught by the same faculty who taught male medical students in St. Petersburg and faced identical graduation requirements. The main curricular differences were that female physicians received more extensive training in midwifery and no training in forensic medicine. To the extent that these differences affected physician effectiveness, they would be most relevant for outcomes

related to childbirth. Section 5, however, also examines mortality among groups for whom this training difference is unlikely to be relevant.

A second threat is that female physician entry may have coincided with other health-improving investments. For instance, more progressive districts may have been both more willing to hire female physicians and more likely to implement other healthcare reforms at the same time. By construction, first female physician hires that coincide with an expansion in the total physician workforce are excluded from the analysis sample. A remaining concern is that districts could have bundled female physician hiring with other changes in healthcare provision. I assess this possibility in Section 5.5 by examining whether female physician entry coincided with changes in other inputs controlled by *Zemstvo* boards, including medical budgets and clinic beds. I also report estimates from a complementary shift-share instrument for female physician entry based on district-level physician vacancies and the regional supply of newly graduated female physicians.

4.3 Balance on Observables and Raw Trends

The hiring of female physicians was not random: it depended on the availability of vacancies, female physicians applying to posted positions, and *Zemstvo* boards choosing to hire them. I therefore assess whether baseline trends differed across three groups: (1) early adopters, treated between 1878 and 1888, when both mortality and healthcare provision outcomes are available, comprising 45.8% of the sample; (2) late adopters, treated between 1889 and 1910, when mortality but not healthcare provision outcomes are available, comprising 48.6% of the sample; and (3) future adopters, treated between 1911 and 1915, after the outcome data end, comprising 5.6% of the sample.

In addition to examining event-study pre-trends for each outcome, Table 3 compares baseline rates of change in key demographic and healthcare variables across these three groups. None of the joint F -tests reject equality of baseline trends at the 5% level, with only one (change in births) comparison marginally significant at the 10% level. Appendix Table B.5.1 similarly compares baseline (1876) levels across adoption groups: none differ significantly at the 5% level, and only two differences (mortality at months 1-2 and the number of midwives) are marginally significant at the 10% level.

I also examine raw trends in the primary outcome of infant mortality, which can be meaningfully expressed as a rate rather than a raw count. For other mortality outcomes, I observe counts of deaths, which mechanically increase with population growth. Figure 2 plots infant mortality in calendar time for early adopters (1878–1888), late adopters (1889–1910), and future adopters (1911–1915). Although the differences in baseline levels are not statistically significant in Appendix Table B.5.1, the figure shows higher initial infant mortality among early adopters. Differences in outcome levels do not threaten identification in a difference-in-differences design so long as trends would have evolved similarly in the absence of treatment. Consistent with this requirement, the formal baseline comparisons above show no statistically significant differences across adoption groups. As an additional robustness check, I allow for differential trends by baseline mortality by interacting

baseline mortality with year fixed effects.

The patterns of infant mortality in Figure 2 are consistent with female physician entry reducing infant mortality. All three groups followed similar trends between 1876 and 1878. During 1878–1888, when treatment was concentrated among early adopters, infant mortality declined more rapidly in these districts, narrowing their initial gap with late and future adopters. Late adopters tracked future adopters closely during most of the 1876–1888 period, with somewhat higher mortality levels, but subsequently caught up with and generally surpassed future adopters during 1889–1910, when they began to receive treatment. Appendix Figure B.6.1 shows this stark picture of convergence in space.³⁴

5 Results

I find that the entry of the first female physicians led to large and persistent reductions in infant mortality and mortality among both young men and women, alongside increases in population growth and life expectancy. I also find suggestive evidence of reductions in child mortality, but no effects on older adult mortality. I benchmark these estimates against modern and historical health-improving interventions and further show that the observed mortality patterns are consistent with the scope of contemporary medical knowledge. Using detailed data on healthcare provision, I then examine potential channels, documenting increased clinic use and reduced dependence on midwives, consistent with a demand-side response.

5.1 Mortality

Infant Mortality I begin by examining the effects of the first female physician entry on infant mortality. Figure 3 plots event-study coefficients estimated from equation 3, with the outcome measured as the logarithm of infant mortality. Coefficients can therefore be interpreted as semi-elasticities.

Pre-trends are visually flat in the ten years prior to female physician entry, and a joint F-test fails to reject that pre-treatment coefficients are jointly zero ($p = 0.94$).³⁵ Coefficients turn negative upon physician entry and reach statistical significance by year two. Though estimates in years three and four are noisier and lose statistical significance, point estimates remain negative throughout and are statistically significant in subsequent years, with effects increasing slightly over the first five years, before stabilizing at a lower level. Because deaths in the first day(s) of life, including stillbirths and very early neonatal deaths, are inconsistently recorded in these clergy-based records (Section 3.2), the infant mortality measure used here understates mortality concentrated at birth itself. The observed decline, therefore, reflects improved survival over the course of infancy, dominated by

³⁴Appendix Figure B.7.1 plots trends in infant mortality in relative time, from five years before to ten years after the first female physician entry in a district. To avoid compositional changes affecting the means, I construct a fully balanced panel of districts observed in all relative years, which leaves 41 treated districts. Infant mortality rates were generally declining over this period, but the decline appears to accelerate following the first female physician entry.

³⁵While conventional pre-trend tests may be underpowered, the HonestDiD sensitivity analysis by Rambachan and Roth [2023] in Appendix D.3 shows that the results remain robust under modest deviations from parallel trends.

causes of death related to hygiene and feeding practices.

The gradual improvement in the initial years after entry is consistent with mechanisms that take time to diffuse, including the development of patient trust and the dissemination of medical knowledge. As female *Zemstvo* physician Alexandra Arkhangelskaya described, medical and maternity stations played an important educational role in rural communities:

“Giving birth in a hospital, [the peasant woman] sees simplicity and cleanliness in how she and her child are cared for. [...] Every medical and maternity station [...] serves as a school for the peasant woman. She not only learns for herself, but also spreads it to the women around her.” Alexandra Arkhangelskaya, *Zemstvo physician 1885–1905* in [Jakobson, 1958, p.23].

This contemporaneous account emphasizes the role of medical care as a channel for the diffusion of better hygiene and infant-care practices. As discussed further in Section 5.4, women were more likely to give birth in these clinics when a female physician was in charge.

I present the average treatment effect estimates in Table 4. Column (1) reports estimates without controlling for the total number of physicians; Column (2) adds this control, with the coefficient remaining unchanged. This preferred estimate suggests that infant mortality declined by 13.7 deaths per 1,000 births following female physician entry. In a median treated district, which recorded 7,349 births per year, this effect translates into approximately 100 fewer infant deaths annually. This drop likely resulted from a combination of improved direct medical information and greater dissemination of education.

Column (3) examines heterogeneity by baseline medical provision. In districts where the first female physician represented a larger share of the medical workforce, the potential effect on district-level mortality should be greater. Therefore, I divide districts according to whether the first female physician represented an above-median share of physicians at entry and interact treatment with this indicator. The interaction is negative, but only marginally statistically significant.

Column (4) tests whether the effects of female physician entry were larger in districts with higher initial infant mortality. I measure baseline mortality in 1876, before the first female physicians graduated, and mean-center it across districts. The interaction is negative and statistically significant, indicating larger mortality reductions in districts with higher initial mortality. The coefficient on female physician entry also remains negative and statistically significant, corresponding to the estimated effect at the mean level of infant mortality in 1876.³⁶ Columns (5)–(8) repeat the analyses

³⁶Because early adopters had somewhat higher infant mortality than later adopters (Figure 2), a related concern is that this pattern could reflect differences across treatment cohorts. Appendix Figure D.3.5 reports cohort-specific treatment effects following Callaway and Sant’Anna [2021]. The estimated effects are negative for 21 of 27 treatment cohorts, including 10 post-1888 cohorts, indicating that the mortality reductions are not concentrated among the earliest-treated districts. As an additional robustness check, Section 5.5 allows mortality trends to vary flexibly with initial mortality by interacting 1876 infant mortality with year fixed effects.

of columns (1)–(4) using log infant mortality.

One alternative explanation for the decline in infant mortality is that female physicians affected fertility decisions, including induced abortions, thereby changing the composition of live births. However, I find no significant effect of female physician entry on the number of births (Appendix Figure C.2.1); if anything, the estimates suggest a slight increase in live births. Although this does not rule out changes in birth composition, the results provide no evidence of a decline in live births that might be expected if access to abortion increased.

Infant Mortality by Month of Life The *Vital Statistics* provide detailed data on infant deaths by age groups: first month of life, 1–2 months, 3–5 months, and 6–12 months, allowing for a more detailed analysis of the origins of declines in infant mortality. Examining heterogeneity within infancy helps identify which causes of death female physicians were most effective in addressing.

Figure 4 presents event-study estimates for death rates per 1,000 births for each of these categories, adding estimates from each month group one at a time for visual comparison. Two patterns stand out. First, infant mortality reductions accrued throughout infancy. Second, point estimates are smallest and least precise in the first month of life and largest in the middle months of infancy, although the estimates are not statistically distinguishable across age intervals.

Relatively small and noisy estimates for the first month of life likely reflect two factors. First, the recording of stillbirths and deaths prior to christening was at the discretion of parents. While some provinces reported stillbirths in the *Movement of Population* data, many reported none - a pattern that is implausible and suggests inconsistent reporting rather than the absence of such events. Second, a large fraction of deaths in the first month of life seemed to be poorly understood at the time and therefore likely regarded as untreatable, as further discussed in Section 5.3.

Large declines occurred in later infancy, particularly between one and five months of age, when infants were most vulnerable to poor feeding and hygiene practices. The effects decline in magnitude after six months. Notably, nearly three-quarters of infant deaths occurred within the first six months of life, accounting for more than 220 deaths per 1,000 births.

Child and Adult Mortality While infant mortality accounted for over 40% of annual deaths in the sample, female physicians treated patients across all ages and genders. The rich vital statistics data allow me to examine effects beyond infancy. Figure 5 presents estimates separately for childhood (ages 1-14), early adulthood (ages 15-54), and older adulthood (ages 55 and above), as described in Section 3.2.

I estimate Equation 2 separately for each age-sex group. Because annual age-specific population counts are unavailable, the dependent variables are death counts and the PPML estimates are reported as proportional changes in deaths, $100[\exp(\beta) - 1]$. To provide an absolute benchmark for their magnitude, I use age- and sex-specific population counts from the 1897 Census together with 1896 deaths to calculate mortality rates near the middle of the sample period. Figure 5 reports

in parentheses the implied percentage-point change in annual mortality obtained by applying each estimated proportional effect to the corresponding 1896 mortality rate.

The estimated effects on infant deaths are negative and similar to those in Table 4. Female physician entry reduces infant deaths by approximately 5.3-5.5%, corresponding to an implied decline in infant mortality of approximately 1.52-1.69 percentage points at the 1896 mortality rate. For children ages 1–14, deaths fall by approximately 5.7-6.2% between sexes, corresponding to implied declines in annual mortality of approximately 0.16-0.18 percentage points, respectively. However, the corresponding event studies for child deaths in Appendix Figure C.1.1 display negative pre-trends, so I interpret these estimates cautiously.

I find large and statistically significant declines in deaths among both men and women ages 15–54. Female physician entry reduces male deaths by approximately 9.2% and female deaths by 7.5%. The confidence intervals are relatively wide for these age groups, so the estimates should be interpreted primarily as evidence that mortality declined rather than as precise estimates of the magnitude of the effect. At the corresponding 1896 mortality rates, these estimates imply reductions in annual mortality of approximately 0.09 and 0.07 percentage points, respectively. Thus, although the proportional declines in deaths are substantial, their absolute magnitudes are considerably smaller when expressed relative to the population at risk. I find no statistically detectable effects among individuals ages 55 and above.

Population Growth As a demographic implication of the observed changes in deaths, I reconstruct annual district-level natural population growth rates by combining the birth and death series with population counts from the January 1897 Census.³⁷ Figure 6 shows that the entry of female physicians is associated with a significant acceleration in natural population growth. On average, natural growth increases by 0.11 percentage points relative to a pre-treatment mean of 1.65%. Appendix Figure C.2.1 decomposes this effect and shows that the increase is initially driven by mortality reductions and, in later years, by suggestive increases in births. The latter effect is likely coming from both decreases in deaths among men and women in young adulthood, and lower prevalence of stillbirths, which I cannot capture but would appear in this pattern.

Life Expectancy To translate the observed mortality reductions into an aggregate metric, I construct approximate annual period life expectancy at birth using an indirect standardization of the 1897 Census. Figure 7 presents the corresponding event-study estimates. During this period, in the studied districts, mean female life expectancy at birth was about 39 years, while male life expectancy was about 35 years. The entry of a female physician increased female life expectancy by roughly 1.1 years (2.8%) and male life expectancy by about 1.2 years (3.5%). Appendix C.3 provides details on the construction of these estimates.

³⁷Migration is unlikely to pose a major concern: according to the 1897 Census, 93% of the rural population in the studied provinces were born in the same district in which they were enumerated.

5.2 Interpreting Effect Sizes

The mortality reductions documented above are large for the entry of a single physician into a district. These point estimates should not be interpreted as exact magnitudes: confidence intervals are fairly wide for several age-specific outcomes, particularly for young-adult mortality. Moreover, because the outcomes are observed at the district level, the estimates capture district-wide effects and cannot be mechanically rescaled by the number of physicians or catchment areas in the district. Such a calculation would require assuming that catchment areas were equally sized and that the effects of female physician entry were confined to patients directly served by the entering physician. The latter assumption is unlikely to hold for outcomes affected by the diffusion of medical knowledge and health practices - contemporary accounts explicitly describe women spreading hygiene and infant-care practices learned at medical and maternity stations to others in their communities [Jakobson, 1958]. Existing estimates from other settings are not directly comparable to the replacement-type variation studied here, but they provide context for the magnitude of health effects associated with changes in provider quality in low-resource environments.

Okeke [2023] implements a randomized experiment in Nigeria, where communities were assigned either an additional mid-level provider, identical in quality to existing ones, or a physician. Those assigned an additional mid-level provider experienced no detectable mortality reductions, meaning that on the margin, an additional provider of similar quality may not improve health outcomes. On the other hand, communities assigned a physician saw seven-day infant mortality fall by approximately 20%, with each physician representing roughly one-sixth of the total provider workforce. In settings of severe medical access scarcity, closely analogous to the historical environment studied here, a single high-quality physician can thus produce large reductions in mortality.

These findings are also broadly consistent with existing literature on historical healthcare interventions. Anderson et al. [2020] find that introducing occupational licensing requirements for midwives, improving the quality of an average midwife, reduced maternal mortality by 7–8% across US states between 1900 and 1940. Liebert and Mäder [2026] find that one additional physician per 1,000 population reduced infant mortality by 23% in German municipalities in the 1930s, where the average municipality had roughly 96 physicians.

At the same time, large-scale public health interventions have been documented to have much larger effects. Alsan and Goldin [2019] find that the joint introduction of safe water and sewerage reduced child mortality in Massachusetts municipalities by roughly one-third over four decades, and Cutler and Miller [2005] attribute approximately half of the decline in US urban mortality between 1900 and 1936 to water chlorination and filtration.

Further, several features of the setting justify expecting large per-physician effects of the type identified here. First, the baseline environment was one of very high preventable mortality and very low medical quality: most childbirth occurred outside formal care, hygiene practices were poor, and

the untrained midwives who dominated childbirth assistance were widely regarded as ineffective or harmful (Section 2). Among young men, mortality was frequently driven by injuries followed by infection, conditions for which timely antiseptic wound care could substantially improve survival (Section 5.3). In this context, even modest improvements in care quality, such as antiseptic technique during delivery or wound care, and basic hygiene instruction, could translate into substantial proportional reductions in mortality across multiple demographic groups.

Finally, these estimates should be interpreted as capturing the impact of a highly selected group of early entrants. Evidence in Section 6 shows that the female advantage in estimated physician value added is concentrated among earlier graduation cohorts, consistent with stronger positive selection among pioneers. As discussed in Section 6.2, liberalization would, in the long run, exert upward pressure on the average quality of male physicians by raising the competitive bar and gradually shrinking the quality gap between male and female physicians. If positive selection was strongest among the earliest entrants, the effects estimated here may exceed the long-run gains once entry barriers are fully relaxed and selection thresholds adjust.

5.3 Causes of Death by Sex and Age Group

What explains the heterogeneity in effects across age groups and gender? To shed light on this question, I use historical evidence on causes of death by age and sex. I digitize cause-of-death data from an 1880 report for Kostroma Province, a *Zemstvo* province in my sample [Ivanov, 1883]. To maintain consistency with the study setting, I focus on rural areas and omit cities. I first describe cause-of-death profiles for each demographic group and then relate the estimated effects to these underlying mortality patterns. Appendix Table A.5.1 summarizes these distributions.

Infants Most deaths among infants under six months were attributed to abdominal diseases linked to poor hygiene and feeding practices, and were, in principle, preventable through physician-provided education and, in some cases, even treatable if caught on time. Natkhov and Vasilenok [2023] report that over 90% of ethnic Russian infants, who make up the vast majority of this study’s sample,³⁸ were introduced to solid food before six months of age, significantly increasing their risk of gastrointestinal illness. Contemporary medical observers expressed concern over such practices. Gilyarovsky [1866], who studied childhood mortality in Novgorod Province, described infants left unattended for over twenty-four hours, fed spoiled dairy, kept warm in ovens, or placed in worm-infested cradles. He concluded that he “found many child-rearing customs that are incompatible with the safety of children” (p. XLV).

The second most common cause was classified as “inherent weakness,” with most such deaths, by definition of the data, occurring in the first month of life [Ivanov, 1883, p. IV]. This category likely captured complications from childbirth, undiagnosed conditions, and what would now be classified

³⁸A median district in my sample had 93% ethnic Russians, making heterogeneity analysis along these lines difficult. *Zemstvos* were generally established in places that would pose a lesser risk of secession and therefore, by definition, did not include many districts with non-Russian majorities.

as sudden unexpected infant deaths.³⁹ The vagueness of this category suggests that contemporary physicians neither understood nor knew how to treat such cases.

Childhood In childhood (ages one to fifteen), mortality was driven primarily by infectious diseases, which accounted for 70–72% of deaths. Smallpox was an especially important cause ($\sim 47\%$) of mortality among children, and contemporary vaccination technology left substantial scope for improvements in quality. [Chapman \[2026\]](#) shows that in nineteenth-century Britain, large declines in smallpox mortality occurred only after improvements in the quality and reliability of vaccination. Contemporary evidence suggests that similar variation in vaccine quality existed in the Russian Empire. Describing her response to a local smallpox epidemic, the *Zemstvo* physician Anna Veretennikova wrote: “Instead of the disgusting substance that I found in the possession of the nurse and the vaccinators, I ordered [new materials]. The success of vaccination [...] exceeded my expectations: for every hundred successful cases, there were only three or four unsuccessful ones” [[Veretennikova, 1984](#)].⁴⁰ Higher-quality physicians could therefore have reduced childhood mortality not only through treatment, but also through more effective preventive care.

Abdominal diseases remained an important cause of death, accounting for 15% of deaths. Fever of unknown etiology also accounted for $\sim 8\%$ of all deaths, some of which may have reflected wound infections, as discussed below. Because these latter causes accounted for only a small share of deaths in this age group, however, proportionally similar reductions in abdominal-disease mortality and wound infections would translate into much smaller effects on overall childhood mortality than in infancy or early adulthood.

Adults In early adulthood (ages sixteen to forty-nine⁴¹), causes of death diversified and diverged by gender. Apoplexia and violent or accidental causes, which would have included infections resulting from such causes, accounted for 17.1% of male deaths, compared with 3.8% of female deaths in the Kostroma data.⁴² In addition, 18.7% of deaths, the largest share of all male age groups, were attributed to unspecified fever. Unspecified fever may also have included infections following injury. When injuries were followed by infection, deaths could be prevented with timely and appropriate wound care.⁴³ Access to more competent wound care, including using the antiseptic method which has been deemed “one of the simplest and cheapest life-saving ideas in history” [[Mokyr, 2005](#), p. 94], had become widespread in the Russian Empire, followed by more rigorous aseptic techniques in the

³⁹In 2022, 18% of all infant deaths in the United States were attributed to unexplained causes [[Centers for Disease Control and Prevention, 2024a,b](#)]. Of these, 41% were due to Sudden Infant Death Syndrome (SIDS), 31% to unknown causes, and 28% to accidental suffocation or strangulation.

⁴⁰“Success” of smallpox vaccination could be evaluated shortly after, based on the reaction at the injection site [[Chapman, 2026](#)].

⁴¹The age group cutoffs were decadal, so I could not cut them off at 54, as in the rest of the paper.

⁴²While the original table only mentions Apoplexia and Accidental Causes, the detailed breakdown of Accidental Causes includes events such as murder, which I classify as Violent.

⁴³A famous case of a death resulting from lack of antiseptic care is that of the US President James Garfield, who received a non-life-threatening gunshot wound but passed away from ensuing infections likely caused by his doctors’ refusal to use antiseptic methods, such as probing the President’s wounds after dismounting their horses without washing their hands [[Ehrhardt, O’Leary, and Nakayama, 2018](#)].

1890s [Breido, 1956]. However, their effectiveness depended on strict adherence to protocols, creating scope for variation in physician performance. It is important to note here that I do not interpret the approximately 9.2% point estimate literally as the share of male deaths prevented through this channel. The cause-of-death evidence helps explain why young men may have been particularly responsive to differences in physician quality. Violent and accidental causes represented a significant share of their mortality, and some of these injuries created substantial scope for mortality reductions due to antiseptic wound care.

Among women, childbirth was a major cause of death, accounting for 14.7% of deaths in the 16–49 age group. Non-specific fever deaths, which could include cases of maternal sepsis, which was often undiagnosed (e.g., Ministry of Internal Affairs [1889]), accounted for another 19.5% of deaths, the largest share of all female age groups. Many childbirth-related deaths were potentially preventable with contemporary medical knowledge: antiseptic and aseptic practices could reduce infections associated with delivery,⁴⁴ while safer obstetric techniques could reduce mortality from other complications of childbirth. The scope for medical intervention was therefore substantial.

At older ages, mortality was dominated by chronic and infectious diseases for which effective medical interventions were limited at the time. In addition, over the longer time horizons, deaths prevented in earlier ages would have shifted to this age group.

5.4 Channels of Mortality Reductions

Having shown that the entry of female physicians reduced mortality and increased population growth and life expectancy, I next examine how healthcare provision and take-up changed following female physician entry. These outcomes provide evidence on the *channels* through which the mortality effects may have operated. Section 6 then turns to *why* female physicians differed from the male physicians they replaced.

Table 5, Panel A, shows how healthcare provision changed following female physician entry. The number of in-clinic patients increased, while the estimated increase in the share of births occurring in clinics is positive but imprecisely estimated. These patterns could reflect increased demand or more efficient clinic operations.

As discussed in Section 2, midwives, often semi-trained or untrained, provided an important alternative source of childbirth care. Column (3) of Table 5, Panel A⁴⁵ shows a sharp decline in the number of midwives following female physician entry, consistent with substitution away from these providers once female physicians became available. Because midwife employment was determined by local *Zemstvos*, this response may reflect either realized changes in patient demand or boards'

⁴⁴A well-known example of the consequences of inadequate antiseptic practices comes from 19th-century Viennese hospitals, where maternal mortality from sepsis was significantly higher among patients treated by doctors than by midwives. Physicians often assisted with childbirth immediately after performing autopsies, while midwives did not [Loudon, 2013].

⁴⁵See also Appendix Figure D.5.1 for the corresponding event study.

anticipation of such changes. Contemporary accounts directly describe this substitution. Reflecting on her own practice after arriving in a rural district, Anna Veretennikova noted that women turned to her rather than the local midwife in pathological deliveries, leaving the midwife with little work: “her presence in the village was unnecessary, and [she] had absolutely nothing to do” [Veretennikova, 1984, p. 53]. I find little evidence that other lower-level medical personnel adjusted following female physician entry.

The increase in clinic use, a suggestive increase in the share of births taking place in clinics, and a sharp decline in midwife employment provide a plausible utilization channel through which female physician entry may have contributed to the mortality reductions.

Additional suggestive evidence points to another channel: administrative diligence. A gender gap in time spent on electronic records today has been well-documented [Ganguli, Rivara, and Inouye, 2022; Bessen et al., 2025]. I document a related descriptive pattern in this historical setting. The *Medical Directories*, upon which this study draws for physician labor force counts, reported three lists of physicians every year: male, female, and late submitters. Out of 608 *Zemstvo* doctors on these lists of late-submitting physicians between 1889-1910, only 2 (0.3%) were women. This is roughly one-twentieth of women’s 6.8% share of *Zemstvo* physician-years. While not a direct measure of clinical effectiveness, this pattern suggests that female physicians were more consistent in administrative tasks, potentially improving continuity and quality of care.

5.5 Robustness

Alternative Identification Strategies, Estimators and Specifications. Appendix D.1 reports results from three alternative identification strategies. First, I compare mortality changes around female and male replacement hires using a stacked design. Each stack groups districts hiring female replacement physicians in a given year with districts hiring male replacement physicians in that same year, using observations from five years before through five years after the replacement and including stack-specific district and calendar-year fixed effects. To distinguish differences in identification strategy from differences in sample composition, I also re-estimate my primary specification over 1885–1910 using the same districts and female-entry cohorts. The estimates from the two approaches are similar in magnitude and exhibit a broadly comparable pattern across age and sex groups, although both are imprecise. Both approaches produce negative point estimates for infants and adults ages 15–54, consistent with the mortality declines in the main analysis. This comparison supports the use of the main identification strategy - the variation in the timing of female replacement entry - as an alternative to directly comparing female and male replacement events.

Second, I examine changes in mortality around expansions in the physician workforce, comparing expansions that introduce a district’s first female physician with expansions that add a male physician. I construct non-overlapping event windows around one-physician expansions. Male physician

expansions are considerably more frequent than first-female expansions and are therefore estimated relatively precisely. I find little evidence that male physician expansions reduce mortality: with the exception of mortality among children ages 1–14, the estimates are close to zero across age groups. The female-versus-male estimates are substantially less precise because there are fewer first-female expansion events, most fall at the tail-end of the observation period, and, after constructing non-overlapping windows, these events have fewer post-entry years per event. For infant mortality, the point estimates are negative and suggest that expansions introducing the first female physician reduced mortality relative to male physician expansions, although the estimates are too imprecise to draw strong conclusions. For other age groups, the differential estimates do not exhibit a clear pattern. I interpret this exercise as primarily providing evidence that physician-supply expansions alone did not generate mortality declines comparable to those following female physician entry, consistent with findings in a more contemporary setting by [Okeke \[2023\]](#), who finds that expansions in access to providers of similar quality have little effect on mortality.

Third, I estimate an interrupted time-series specification that allows both the level and the district-specific mortality trend to change following the first female physician’s entry. This approach uses within-district variation to compare mortality after entry with the trajectory predicted by each district’s own pre-entry trend. The results provide additional evidence of mortality improvements at younger ages. Infant mortality falls discretely at entry and subsequently declines relative to its pre-entry trend. I find a similar pattern for children ages 1–14 of both sexes, with negative shifts in both mortality levels and post-entry trends. For ages 15–54, the estimated level shifts are negative but imprecise, while the subsequent trend increases for men and does not significantly change for women. I find no evidence of either level or trend changes at ages 55 and older.

I next implement several checks on the event-study specification. First, to examine whether female physician entry followed a temporary adverse mortality shock, I compare mortality in the year immediately preceding entry with the earlier pre-treatment period. Appendix Figure [D.3.1](#) plots all pre-treatment event-time coefficients after re-centering them so that the average through $t = -2$ equals zero. Mortality at $t = -1$ is statistically indistinguishable from this earlier pre-treatment average and does not stand out relative to the surrounding pre-treatment coefficients, providing no evidence of a mortality spike immediately before female physician entry.

For outcomes expressed in rates. I re-estimate the effects using the imputation estimator of [Borusyak, Jaravel, and Spiess \[2024\]](#) and the heterogeneous-treatment-effect robust estimators of [Callaway and Sant’Anna \[2021\]](#) and [Sun and Abraham \[2021\]](#). Results are consistent across approaches (Appendix Figures [D.3.2](#) and [D.3.3](#)).

I also estimate a specification that departs from the absorbing-treatment definition by excluding post-entry district-years in which no female physician was present, which yields very similar results (Appendix Figure [D.3.6](#)). A Goodman-Bacon decomposition confirms that no pairwise comparisons receive negative weights [[Goodman-Bacon, 2021](#)]. A permutation test that randomly reassigns the

observed female physician entry years across districts, while preserving the empirical distribution of treatment timing, yields a two-sided permutation p -value of 0.016 (Appendix Figure D.3.7), indicating that an estimate at least as extreme as the observed coefficient on infant mortality occurs infrequently under random reassignment of treatment timing.

Given the different levels of baseline mortality among earlier versus later-adopting cohorts (Figure 2), I interact baseline mortality levels with year fixed effects to allow for differential trends by baseline mortality. The results are very similar to the main estimation (Appendix Table D.3.1). Results are also robust to alternative weighting schemes: importantly, results are negative, statistically significant, and larger in magnitude when weighting districts by their 1897 Census population, although I retain the unweighted specification as the baseline because, as I show, population growth itself is affected by female physician entry (Appendix Table D.3.1). The results are also robust to spatial and serial correlation adjustments using Conley standard errors (Appendix Table D.3.2). I allow residual correlation within radii of 200, 500, and 1,000 kilometers, ranging from relatively local spatial dependence to a deliberately conservative specification permitting correlation across much broader geographic areas.

In Appendix D.3, I apply the Honest Difference-in-Differences framework [Rambachan and Roth, 2023] to assess sensitivity to violations of the parallel trends assumption. The estimated effects remain similar under small departures from parallel trends, allowing post-treatment deviations up to 10% as large as the largest deviation observed in the pre-treatment period. Under larger departures, corresponding to 20–30% of the largest pre-treatment deviation, the robust confidence intervals widen substantially and generally include zero, due to a relatively small sample size.

For the non-linear models using Poisson Pseudo-Maximum Likelihood (PPML) estimation, in Appendix Figure D.3.4 I verify robustness using logged outcomes and the standard TWFE estimator as well as logged outcomes and the Callaway and Sant’Anna [2021] estimator, although the event studies using the latter estimator are substantially noisier.

Departing Physicians: Training Recency & Physician Quality. Next, I address the identification concern that the results reflect training recency rather than physician gender, and a related concern that the results reflect characteristics of the departing male physician rather than the quality of the entering female physician. I provide direct evidence using individual-level data from the *Medical Directories*, available for the 1889–1915 period, which I link to identify both the entering female and departing male physician in each replacement district. I find no statistically detectable difference between the value added⁴⁶ of departing male physicians and that of other male Zemstvo physicians (Appendix Figure E.1.4). This provides no evidence that female physicians systematically replaced unusually low-value-added men.

The most direct test of training recency compares the graduation years of the entering female

⁴⁶More on value added estimation in Section 6.3.

physician and the departing male physician. Appendix Table D.4.1, column (1), interacts female physician entry with the difference between their graduation years, where positive values indicate that the entering female physician graduated more recently. The interaction is small and statistically indistinguishable from zero. Column (2) enters the graduation years of the entering female and departing male physicians separately; neither interaction is statistically distinguishable from zero. Column (3) examines the age at exit of the departing male physician. This estimate is also imprecise and provides no evidence that the observable age of the departing physician accounts for the estimated mortality effect.

Bundled Treatment. A remaining concern is that districts bundled female physician hiring with other changes in healthcare provision that could independently reduce mortality. I address this threat to identification in several ways.

Appendix Figure D.4.1 examines two margins determined by *Zemstvo* boards: clinic beds and medical budgets. If female physician entry coincided with broader improvements, these outcomes might increase. Especially, medical budgets would be expected to increase to provide funds for the life-saving changes. I find no evidence of such increases. A remaining concern is that lower salaries for female physicians could have freed resources for other mortality-reducing inputs even if total medical budgets remained unchanged. This possibility appears quantitatively limited. In 1888, 38 of the 50 districts employing female physicians for which wages are observed paid male and female physicians equally. Among the remaining districts, the female-male wage differential typically amounted to approximately 0.2% of an average district medical budget and was at most 0.8%, making it unlikely that the resulting savings financed substantial changes in other medical inputs.

Table D.3.1 also shows comparable effects for early and late adopters. This suggests that results are not driven by a small set of early, reform-oriented districts.

These checks cannot fully rule out unobserved bundled changes. This is particularly relevant given the frequent turnover of physicians in districts, which makes vacancies relatively common and therefore weakens the interpretation that the timing of female hires was driven by rare, idiosyncratic physician exits.⁴⁷ As a complementary exercise, Appendix D.2 reports a shift-share IV that interacts district vacancies with the regional supply of newly trained female physicians, defined as graduates within three years of completing training. This design requires the additional assumption that, conditional on fixed effects and physician employment, local turnover does not affect infant mortality through channels other than female physician entry. The 2SLS estimate is negative and similar in magnitude to the baseline estimate, but is not statistically significant.

⁴⁷In the 1889-1910 sample, there was some physician turnover reflected in changes in rosters about every other year, not counting expansions and contractions. This rate would likely be smaller for the full sample, as the 1889-1910 period experienced a great famine, a war, and a revolution, which affected physician employment.

Additional Validity Checks. As an additional placebo test, I examine mortality in large cities located within the same districts as the rural populations studied in the main analysis.⁴⁸ These cities were administratively separate from the *Zemstvo* healthcare system and were not served by *Zemstvo* physicians. Because restricting this exercise to the replacement-hire analysis sample leaves only 12 cities, I instead use all cities located in districts that eventually hired a female physician. Treatment timing is defined by the first female physician entry in the surrounding *Zemstvo* district. I find no corresponding decline in urban mortality (Appendix Figure D.4.2a).

Appendix Figure D.4.2b extends the main analysis of infant mortality by including all districts: both districts in which the first female physician replaced a departing male physician and districts in which her entry coincided with an expansion.⁴⁹ In the baseline specification, I exclude expansion events because they often coincided with broader institutional changes or parallel investments in healthcare, making it difficult to isolate the effect of female physician entry alone. Results using the full sample remain similar in magnitude to the main analysis. However, pre-treatment trends are less flat, although never statistically significantly different from zero. This likely reflects the fact that many expansion entries occurred in the final decade of the sample, coincident with rapid growth in the medical workforce and potentially other unobserved improvements in healthcare infrastructure that likely accompanied medical care expansions.

Measurement of Entry Timing. An important assumption underlying the main estimation sample is that the timing of the first female physician entry is measured accurately. For 1889–1910, when I rely on the *Medical Directories*, some imprecision is possible. These directories were typically published in the first quarter of year t , and I assign each roster to year $t - 1$, assuming it reflects the prior year’s workforce. This could introduce error if, for example, a district hired its first female physician in January of year t , but I assign her entry to $t - 1$ and classify the hiring as a replacement based on physician counts between $t - 2$ and $t - 1$. To address this concern, Appendix Figure D.4.2c restricts the sample to districts that experienced no increase in the number of physicians between $t - 2$ and t . Results remain very similar to the main specification.

6 Mechanisms

Section 4.1 introduced the supply and demand channels informally. I now formalize them to clarify the conditions under which each mechanism generates the predicted patterns, and to set up the long-run analysis in Section 6.2. I then evaluate the long-run predictions using physician-level value-added estimates for the short and medium run, and modern evidence for the long run.

6.1 Model: Short-Run Effects

Let there be two types of adult patients, indexed by sex, $i \in \{f, m\}$, who each face a fixed health shock and must decide whether to seek medical care or not. In this setting, I assume that patients

⁴⁸Data from Malein and Matiashvili [2024].

⁴⁹In Table D.3.1, I present the average treatment effect for this unrestricted sample.

are randomly assigned to a doctor of fixed sex and cannot choose between physician genders - a reasonable approximation given the institutional context of the rural Russian Empire, and consistent with constraints in many rural healthcare systems today.

In any district, for each patient type i , they either have access to a gender-concordant or a gender-discordant physician j :

1. **Gender-concordant physician:** Normalize the probability of seeking care from a gender-concordant physician to one.⁵⁰ Thus, the share of type- i patients willing to seek care from a gender-concordant physician is given by

$$p_{i,j} = 1, \forall i = j$$

2. **Gender-discordant physician:** Patient i seeks care with some probability that could be less than one, if perceived cost (such as shame, or discomfort) is smaller than perceived benefit of treatment.

$$p_{i,j} \leq 1, \forall i \neq j$$

Then, the survival likelihood within a patient type can be expressed as: $s_{i,j} = s_{i,0} + p_{i,j}\tau_j$, where $s_{i,0}$ is the probability of survival without treatment for male or female patients, and τ_j is the treatment effect, or value added, of receiving care from a doctor of type j .

To evaluate the net effect of assigning a population of type i to a female versus a male doctor, I define the survival difference:

$$\Delta s_i = s_{i,f} - s_{i,m} = p_{i,f}\tau_f - p_{i,m}\tau_m$$

This difference reflects two potential sources of variation: differences in care-seeking behavior ($p_{i,i} > p_{i,j}$) due to beliefs or preferences; and differences in physician value added by physician gender (τ_f vs. τ_m).⁵¹ The sign and magnitude of Δs_i will vary by patient type i , depending on which margin dominates.

Below, I consider each adult patient type in turn. The primary distinction between patient groups is that male-patient outcomes are relatively more informative about the supply-side channel, while effects for female patients may reflect both supply- and demand-side responses.

Adult Male Patients For adult male patients, we observe:

$$\Delta s_m = p_{m,f}\tau_f - \tau_m$$

This difference is positive - that is, an adult male patient benefits from being assigned a female physician if and only if:

⁵⁰Here, the framework abstracts away from generalized discrimination (e.g., a scenario where all patients inherently prefer male physicians, regardless of the patient's own gender). The key predictions for adult male patients hold as long as men are less likely to seek care from a female physician than a male physician ($p_{m,f} < p_{m,m}$).

⁵¹This framework assumes away any supply-side concordance effects. For example, if physicians are in fact more effective at treating gender-concordant patients, then we would expect larger treatment effect differences for female physicians treating female patients, and attenuated effects for female physicians treating male patients. The current data do not allow me to comment on the presence of these concordance effects.

$$\frac{\tau_f}{\tau_m} > \frac{p_{m,m}}{p_{m,f}} = \frac{1}{p_{m,f}} \geq 1$$

Consistent with the contemporary evidence presented in Section 4.1, we have assumed that $\frac{1}{p_{m,f}} \geq 1$ that is, men experience weak discordance disutility when treated by a female physician, so on net, weakly fewer male patients seek care from female physicians than male physicians. Therefore, male patients benefit from the assignment to female doctors only if entering female physicians have sufficiently larger value added than male physicians to offset the weakly reduced likelihood that male patients seek care from them. In practice, this inequality may be even harder to satisfy if increased care-seeking by female patients (who prefer gender-concordant providers) crowds out marginal male patients.

Adult Female Patients The survival gain from a female physician replacing a male physician can be expressed as:

$$\Delta s_f = \tau_f - p_{f,m}\tau_m$$

This difference is positive whenever:

$$\frac{\tau_f}{\tau_m} > p_{f,m}, \quad p_{f,m} < 1$$

That is, this difference is positive whenever the relative ability advantage of female physicians over male physicians exceeds the probability that female patients seek care from male doctors. Given the assumption that female patients prefer gender-concordant providers, $p_{f,m} < p_{f,f} = 1$, female physician entry can improve female population-level outcomes even when female physicians are *less* effective than male physicians because they bring more women into medical care. Therefore, estimates for female (and infant) mortality cannot conclusively distinguish between supply- and demand-side mechanisms. The availability of data on midwives, however, provides a pre-existing gender-concordant alternative against which changes in women’s care-seeking can be interpreted. The decline in midwife employment following female physician entry is consistent with women shifting toward female physicians once they became available.

6.2 Model: Long-Run

The results of this paper thus far capture a short-run, local effect of the *first* female entrants, when the profession was still adjusting to the removal of entry barriers. An important question is how these effects evolve in the long-run equilibrium once medical education and hiring are fully liberalized. The long-run equilibrium depends on three forces: selection into medicine, underlying ability distributions across genders, and demand-side concordance preferences.

To frame this, consider how entry barriers influence selection into the profession. Throughout, I abstract from gender-differentiated occupational sorting and hold fixed the relative attractiveness of medicine. I assume that, conditional on other determinants of entry, the probability of entering and

remaining in medicine is increasing in latent medical ability, and that the pre-existing additional barriers faced by women raise their effective ability threshold for entry. Consider three periods: $T - 1$, T , and $T + 1$. Period $T - 1$ is the pre-entry equilibrium with only male physicians. T captures the short-run impact of the first female entrants following liberalization. $T + 1$ represents the long-run equilibrium without barriers. The difference-in-differences estimates therefore identify the *local* treatment effect of the initial female cohorts relative to the prevailing male ability threshold at time T .

Initial Selection Under the ability-monotone entry assumption, when barriers to women’s entry are removed, the earliest female physicians are drawn from the extreme upper tail of the female ability distribution. Because these entrants come from a very high ability threshold, their average productivity initially exceeds that of incumbent male physicians (Figure 8a).

Equal Ability Distributions A useful benchmark case assumes that the underlying ability distributions of men and women are identical and that entrants into medicine are drawn from the upper tail of these distributions. As female entry expands, the selection thresholds governing entry into medicine must adjust for both genders. Women begin to enter from progressively lower portions of the female ability distribution. The male threshold rises as higher-ability women replace lower-ability men, until both genders are selected from a common equilibrium threshold t_1 (Figure 8b). In this equilibrium, the marginal entrant is identical across genders, implying gender parity in both physician shares and average ability. Because the new equilibrium threshold exceeds the pre-reform male threshold, entry liberalization raises the profession’s overall ability threshold and increases medical productivity. Appendix E formalizes this result.

Shifted Ability Distributions The long-run equilibrium need not exhibit gender parity if the underlying ability distributions differ across genders. Suppose, for example, that the female ability distribution first-order stochastically dominates the male distribution and satisfies upper-tail mean dominance:

$$E[A_f \mid A_f \geq t] > E[A_m \mid A_m \geq t]$$

for any relevant common entry threshold t . The first condition implies that a larger share of women exceeds any common threshold, while the second implies that, conditional on exceeding that threshold, female physicians have higher average ability.

As illustrated in Figure 8c, the equilibrium threshold t_2 will exceed the equal-distribution benchmark ($t_{f_0} > t_2 > t_1 > t_{m_0}$). In this scenario, women constitute a majority of physicians and have higher average quality, while the marginal quality of male and female physicians remains equal because both groups are selected from the same marginal threshold. This means that removing barriers to entry can raise the overall quality of the profession even if the resulting workforce is not gender balanced. Importantly, the initial gender gap in average quality narrows, although the extent of convergence depends on the underlying talent distributions.

Role of Gender Concordance Preferences Finally, demand-side preferences may also affect the equilibrium composition of the physician workforce. If patients exhibit gender-concordance preferences, the demand for male and female physicians may differ even when underlying ability distributions are identical. For example, if female patients have stronger preferences for female physicians, the demand for female physicians will increase relative to that for male physicians.

In equilibrium, stronger demand for female physicians relaxes their entry threshold, producing a female-majority workforce but higher average and marginal quality among male physicians (Figure 8d). If both forces coexist, female comparative advantage in medicine and stronger demand for female physicians, the female share of the profession expands further. In that case, the effect on average quality differences becomes ambiguous: higher underlying female ability increases average quality, while relaxed selection due to higher demand reduces it. However, stronger demand for female physicians still implies a lower female entry threshold and therefore higher marginal quality among male physicians.

Across all cases, the exceptional selection advantage of the earliest female entrants attenuates as entry thresholds adjust. Average gender differences in physician quality may nevertheless persist or even reverse when underlying ability distributions or demand differ across genders.

6.3 Empirical Evidence: Short and Medium Run

I now turn to evaluating the predictions of the framework. First, I estimate physician-level value added and compare its distribution across male and female physicians. I then examine how gender differences in estimated value added evolved across graduation cohorts to assess whether they exhibit the convergence predicted by the conceptual framework.

Estimating Value Added Differences I estimate value added using the full linked physician panel for 1889–1910, restricted to physicians employed in *Zemstvo* positions from [Matiashvili \[2026\]](#). The value-added construction follows the logic of estimators used in the teacher value-added literature, in particular [Biasi \[2021\]](#), which relies on group-level outcomes rather than direct physician–patient linkage: I residualize district-year mortality on district and year fixed effects and allocate the resulting residual equally across all physicians active in that district-year. I identify physician value added from changes in district-level mortality as physicians enter, leave, and move across *Zemstvo* districts. I use spell-level leave-one-out estimates when I observe more than one spell per physician, and I apply empirical Bayes shrinkage to the estimates. The procedure is described in greater detail in Appendix E.1. Appendix E.1 also reports a calibrated simulation showing that estimated value added is positively correlated with true physician quality despite substantial signal attenuation. I measure value added for 8,430 unique physicians (7,602 male and 828 female) in the 1889–1910 period.

The resulting distributions of estimated physician value added, separately for male and female physicians, are shown in Appendix Figures E.1.1a–E.1.1c, for infant and young adult (female and male)

mortality outcomes. Here, differences in young male mortality most cleanly map to true physician value added, as this group is least likely to reflect demand-side changes in health-seeking behavior. Female physicians have higher estimated value added on average across all three outcomes. The probability that a randomly drawn female physician outranks a randomly drawn male physician ranges from 0.52 to 0.53, relative to 0.5 under the null of no difference, and is statistically significant for two of the three outcomes. The share of female physicians whose estimated value added exceeds the male median ranges from 51-55%. While these average differences are modest, they vary considerably across graduation cohorts, which I examine next. As a descriptive benchmark, the decline in young adult male mortality following female physician entry is similar in magnitude to that following the entry of the first physician in the top quintile of estimated value added (Appendix Figure E.1.3). By contrast, the entry of physicians in the third quintile produced no measurable effect, while the entry of physicians in the bottom quintile increased mortality, and all these effects persisted for at least five years after entry.

These value-added estimates provide an opportunity to examine whether the data exhibit the convergence pattern described in Section 6.2. In particular, given that we observe a third of a century of female physician entry, do we observe convergence in value added due to weaker selection? Figure 9 illustrates how value-added distributions of male and female physicians evolve across graduation cohorts. Because empirical Bayes shrinkage depends on the precision of physicians' estimated value added, it can differentially compress estimated gender gaps across cohorts when the underlying estimates differ in precision. I therefore use raw value added for this comparison. Separately, to address the concern that early-cohort physicians have more years of panel exposure and therefore potentially more identifying variation, I restrict the early, 1878–1887⁵² cohort to their first six years in the panel,⁵³ approximately matching the average observation window of the late, 1903–1910 cohort.⁵⁴

Among physicians graduating in the 1878–1887 period, female physicians have substantially higher estimated value added than male physicians: the probability that a randomly drawn female physician outranks a randomly drawn male is 0.58, and 59% of female physicians exceed the male median value added. Among physicians graduating in the 1903–1910 period, this gap shrinks: the probability that a randomly drawn female outranks a randomly drawn male is 0.5, and 52% of female physicians exceed the male median. This pattern is consistent with the adjustment predicted by the selection framework. Under the framework, such a pattern would arise as positive selection among female entrants declines and the male selection threshold rises as barriers to female entry relax.

⁵²The first period while the female medical school was open.

⁵³The matched window equalizes the number of observed panel years but not the career stages. For example, the 1878-1887 cohort had already been practicing for 2-11 years before the panel begins in 1889, while the 1903-1910 cohort enters the panel near graduation. To the extent that value added rises with experience, this comparison favors the early cohort in terms of the levels of value added; however, any such bias would affect male and female physicians within that cohort symmetrically and therefore cannot account for the within-cohort gender gap.

⁵⁴The period after the reopening of the female medical school; the first graduating cohort was 1903.

6.4 Empirical Evidence: Long-Run

How did the physician workforce evolve over time? In Figure 10a, I first document the evolution of the share of female physicians in the Russian Empire and, subsequently, in the Soviet Union and some of the former USSR republics. This share rose steadily throughout the late imperial and Soviet periods, ultimately stabilizing at over 70%, a level that persists in Russia and many Eastern European countries today. Notably, the share of female physicians in the US has also been rising steadily. In 2019-2020, female medical school students outnumbered men for the first time, reaching nearly 55% of the student body by 2023-2024 [Boyle, 2024].

What can these modern workforce patterns tell us through the lens of this paper’s conceptual framework? To interpret them, I combine two pieces of evidence. The first comes from Kissel, Matiashvili, and Roy [2026], where we use an experimental design among respondents in the modern United States to measure willingness-to-wait for gender-concordant care. We find that women exhibit a preference for female physicians only in gender-sensitive specialties: for example, women are willing to wait two extra weeks to see a female urologist or psychiatrist, but show no preference for a female orthopedist, while men exhibit no preference in any specialty. This suggests that while concordance preferences may affect gender composition in some subspecialties, they cannot account for female-majority workforces across the board.

The second piece of evidence comes from the Republic of Georgia, formerly occupied by the Russian Empire and the Soviet Union, whose medical institutions evolved under similar historical and cultural conditions. I compile novel data on the near-universe of practicing physicians in 2025, including their names, specialties and patient reviews, from a public online platform, and assign gender using a first-name dictionary. I find that 66.7% of physicians are female. Figure 10b shows that women constitute the majority in all specialties except anesthesiology (48%), urology/andrology (10%)⁵⁵, and surgery (9%).⁵⁶ Crucially, while the high female shares in pediatrics (88%) and OB/Gyn (81%) could reflect concordance preferences in gender-sensitive specialties, the female majority extends to non-gender-sensitive specialties: neurology (59%) and cardiology (58%) are two examples.

These patterns are consistent with a female comparative advantage in medicine, although they cannot distinguish this mechanism from other forces affecting occupational sorting. Work-life balance considerations, institutional incentives, and other factors that differentially attract women to medicine may also contribute to workforce composition and cannot be ruled out.

⁵⁵As mentioned, in Kissel, Matiashvili, and Roy [2026], we find no gender concordance preferences among men for urologists. In the light of that evidence, such a low share of female urologists may reflect physicians’ own preferences or physicians’ expectations of patient preferences, even if inaccurate.

⁵⁶See Goldin [2015] for discussion of gender gaps in time-inflexible professions such as surgery and their relation to cultural norms surrounding women’s caregiving responsibilities.

7 Conclusion

What happened to health and healthcare when women were allowed to become doctors? The findings in this paper show that first female physician entry was followed by substantial improvements in population health. Mortality declined among infants and young adults of both sexes, with suggestive declines among children. Natural population growth accelerated, and estimated period life expectancy increased. These changes were accompanied by greater clinic use and a shift away from undertrained midwives toward formal medical care. Together with the physician-level value-added evidence, these results indicate that differences in physician effectiveness mattered even in a low-technology medical environment before antibiotics and many modern therapies.

The evidence is consistent with a combination of supply- and demand-side forces. On the supply side, the mortality effects among young men and the physician-level value-added results point toward positive selection among the earliest female physicians. On the demand side, increases in clinic use, suggestive increases in the share of births in clinics, and the displacement of midwives are consistent with female patients making greater use of formal medical care when female physicians became available.

A hundred and fifty years later, what lessons can we learn from these findings? The large productivity effects documented here reflect the removal of formal barriers to entry in a setting where informal barriers were also muted: autonomous practice limited workplace discrimination, and restricted patient sorting limited selective avoidance of female physicians. The results therefore provide evidence that relaxing a formal barrier to entry can generate substantial gains when it expands the pool of talent available to a high-skill profession, as predicted by the talent misallocation literature for which causal micro-level evidence remains limited [[Hsieh et al., 2019](#)].

Formal and informal barriers, however, have not disappeared. In high-income countries, women in medicine and STEM continue to face disparities in promotion and compensation unexplained by differences in qualifications [[Lautenberger and Dandar, 2024](#)]; in the Middle East, North Africa, and South Asia, formal legal restrictions persist, and female labor force participation remains below 40 percent [[Klasen, 2019](#); [Elborgh-Woytek et al., 2013](#)]. The potential gains from expanding the pool of physician talent and increasing women’s use of formal care may be particularly important in lower-income settings where gender norms limit women’s access to male providers. A parallel lesson extends to high-skill immigration in the developed world: one in four physicians in the United States is an international medical graduate, disproportionately concentrated in the most underserved areas [[Braga, Khanna, and Turner, 2024](#)], and policies restricting their entry would function as direct barriers to entry in the framework of this paper. The broader implication is that institutional restrictions that exclude potentially productive entrants can impose costs on the populations those professionals serve.

References

- Abramitzky, Ran, Lena Greska, Santiago Pérez, Joseph Price, Carlo Schwarz, and Fabian Waldinger**, “Climbing the Ivory Tower: How Socio-Economic Background Shapes Academia,” Working Paper 33289, National Bureau of Economic Research December 2024.
- Abramov, Yakov**, *Women’s Medical Courses [Zhenskie Vrachebnie Kursi]*, St. Petersburg: S.P. Yakovlev Publishing, 1886.
- Ager, Philipp, Casper Worm Hansen, and Peter Sandholt Jensen**, “Fertility and Early-Life Mortality: Evidence from Smallpox Vaccination in Sweden,” *Journal of the European Economic Association*, April 2018, *16* (2), 487–521.
- Alsan, Marcella and Claudia Goldin**, “Watersheds in Child Mortality: The Role of Effective Water and Sewerage Infrastructure, 1880 to 1920,” *Journal of Political Economy*, 2019, *127* (2), 586–638.
- , **Owen Garrick, and Grant Graziani**, “Does Diversity Matter for Health? Experimental Evidence from Oakland,” *American Economic Review*, December 2019, *109* (12), 4071–4111.
- Anderson, D. Mark, Kerwin Kofi Charles, and Daniel I. Rees**, “Reexamining the Contribution of Public Health Efforts to the Decline in Urban Mortality,” *American Economic Journal: Applied Economics*, April 2022, *14* (2), 126–157.
- , **Ron Diris, Raymond Montizaan, and Daniel I. Rees**, “Can Personal Access to Medical Expertise Overcome Vaccine Hesitancy?,” Working Paper 35019, National Bureau of Economic Research April 2026.
- , **Ryan Brown, Kerwin Kofi Charles, and Daniel I. Rees**, “Occupational Licensing and Maternal Health: Evidence from Early Midwifery Laws,” *Journal of Political Economy*, November 2020, *128* (11), 4337–4383.
- Ashraf, Nava, Oriana Bandiera, Virginia Minni, and Víctor Quintas-Martínez**, “Gender Roles and the Misallocation of Labour Across Countries,” Working Paper 2022.
- Bauernschuster, Stefan, Anastasia Driva, and Erik Hornung**, “Bismarck’s Health Insurance and the Mortality Decline,” *Journal of the European Economic Association*, October 2020, *18* (5), 2561–2607.
- Bertakis, Klea D.**, “The influence of gender on the doctor-patient interaction,” *Patient Education and Counseling*, September 2009, *76* (3), 356–360.
- **and Rahman Azari**, “Patient-centered care: the influence of patient and resident physician gender and gender concordance in primary care,” *Journal of Women’s Health (2002)*, March 2012, *21* (3), 326–333.
- Besley, Timothy, Olle Folke, Torsten Persson, and Johanna Rickne**, “Gender Quotas and the Crisis of the Mediocre Man: Theory and Evidence from Sweden,” *American Economic Review*, August 2017, *107* (8), 2204–2242.
- Bessen, Sarah Y., Sean Tackett, Carolyn M. Jenks, C. Matthew Stewart, Maria Oliva-Hemker, and Jennifer K. Lee**, “The gender gap in academic surgery: the electronic health

- record,” *BMC Surgery*, February 2025, 25 (1), 84.
- Biasi, Barbara**, “The Labor Market for Teachers under Different Pay Schemes,” *American Economic Journal: Economic Policy*, August 2021, 13 (3), 63–102.
- Blackwell, Elizabeth**, *Pioneer work in opening the medical profession to women* Classics in women’s studies, Amherst, N.Y: Humanity Books, 2005.
- Borusyak, Kirill, Xavier Jaravel, and Jann Spiess**, “Revisiting Event-Study Designs: Robust and Efficient Estimation,” *The Review of Economic Studies*, February 2024, p. rdae007.
- Boyle, Patrick**, “Women in medicine make gains, but obstacles remain,” Association of American Medical Colleges July 2024. <https://www.aamc.org/news/women-medicine-make-gains-obstacles-remain>.
- Braga, Breno, Gaurav Khanna, and Sarah Turner**, “Migration Policy and the Supply of Foreign Physicians: Evidence from the Conrad 30 Waiver Program,” *Journal of Economic Behavior & Organization*, 2024, 226, 106682.
- Breido, I. S.**, *History of Antiseptics and Aseptics in Russia [Istoriia Antiseptiki i Aseptiki v Rossii]*, Leningrad: State Publisher of Medical Literature, 1956.
- Brouillette, Jean-Félix**, “The Lost Marie Curies and Foregone Economic Growth,” Working Paper 2025.
- Böhm, Michael J. and Martin Watzinger**, “The Allocation of Talent over the Business Cycle and its Long-term Effect on Sectoral Productivity,” *Economica*, 2015, 82 (328), 892–911.
- Callaway, Brantly and Pedro H. C. Sant’Anna**, “Difference-in-Differences with multiple time periods,” *Journal of Econometrics*, December 2021, 225 (2), 200–230.
- Celik, Murat Alp**, “Does the Cream Always Rise to the Top? The Misallocation of Talent in Innovation,” *Journal of Monetary Economics*, January 2023, 133, 105–128.
- Centers for Disease Control and Prevention**, “Data and Statistics for SUID and SIDS,” September 2024. https://www.cdc.gov/sudden-infant-death/data-research/data/index.html#cdc_data_surveillance_section_2-data-resources-for-suid-and-sids.
- , “Infant Mortality,” September 2024. <https://www.cdc.gov/maternal-infant-health/infant-mortality/>.
- Central Statistical Committee of the Ministry of Internal Affairs**, *Statistics of the Russian Empire XII: Movement of the Population of European Russia in 1886 [Statistika Rossiiskoi Imperii XII: Dvizhenie Naselenia v Evropeiskoi Rossii za 1886 god]*, Saint Petersburg: Ministry of Internal Affairs, Tipographia i Fototipia V.I. Shtein, 1890.
- , *Statistics of the Russian Empire XXXIII: Movement of the Population of European Russia in 1890 [Statistika Rossiiskoi Imperii XXIII: Dvizhenie Naselenia v Evropeiskoi Rossii za 1890 god]*, Saint Petersburg: Tipo-litographia P. P. Soikina, 1895.
- , *Vladimir Governorate. [Vladimirskaya Guberniia]*, Vol. II of *The First General Population Census of the Russian Empire of 1897.*, Saint Petersburg: Parovaya Skoropechatnya P.O. Yablonskogo, 1904.

- Chapman, Jonathan**, “The contribution of infrastructure investment to Britain’s urban mortality decline, 1861–1900,” *The Economic History Review*, 2019, 72 (1), 233–259.
- , “State Capacity and the (In)Effectiveness of Medical Technology: The Difficult Path to Safe and Reliable Vaccination in Victorian Britain,” 2026. Preliminary working paper, available on request.
- Conley, T. G.**, “GMM estimation with cross sectional dependence,” *Journal of Econometrics*, September 1999, 92 (1), 1–45.
- Correia, Sergio, Paulo Guimarães, and Tom Zylkin**, “Fast Poisson estimation with high-dimensional fixed effects,” *The Stata Journal*, March 2020, 20 (1), 95–115. Publisher: SAGE Publications.
- Crawford, Damian, Suchitra Paranj, Shalini Chandra, Scott Wright, and Flora Kisule**, “The effect of racial and gender concordance between physicians and patients on the assessment of hospitalist performance: a pilot study,” *BMC Health Services Research*, April 2019, 19, 247.
- Curtin, L. R. and R. J. Klein**, “Direct standardization (age-adjusted death rates),” *Healthy People 2000 Statistical Notes*, March 1995, (6), 1–10.
- Cutler, David and Grant Miller**, “The role of public health improvements in health advances: the twentieth-century United States,” *Demography*, February 2005, 42 (1), 1–22.
- Egedesø, Peter Juul, Casper Worm Hansen, and Peter Sandholt Jensen**, “Preventing the White Death: Tuberculosis Dispensaries,” *The Economic Journal*, July 2020, 130 (629), 1288–1316.
- Eggermont, D., M. a. M. Smit, G. A. Kwestroo, R. A. Verheij, K. Hek, and A. E. Kunst**, “The influence of gender concordance between general practitioner and patient on antibiotic prescribing for sore throat symptoms: a retrospective study,” *BMC family practice*, November 2018, 19 (1), 175.
- Ehrhardt, John D., J. Patrick O’Leary, and Don K. Nakayama**, “"Yes, I Shot the President, but His Physicians Killed Him." The Assassination of President James A. Garfield,” *The American Surgeon*, November 2018, 84 (11), 1711–1716.
- Elborgh-Woytek, Katrin, Monique Newiak, Kalpana Kochhar, Stefania Fabrizio, Kangni Kpodar, Philippe Wingender, Benedict Clements, and Gerd Schwartz**, “Women, Work, and the Economy: Macroeconomic Gains from Gender Equity,” Staff Discussion Note SDN/13/10, International Monetary Fund September 2013.
- Fee, Elizabeth and Theodore M. Brown**, “An Eventful Epoch in the History of Your Lives,” *American Journal of Public Health*, March 2004, 94 (3), 367.
- Ferrie, Joseph P. and Werner Troesken**, “Water and Chicago’s mortality transition, 1850–1925,” *Explorations in Economic History*, January 2008, 45 (1), 1–16.
- Frenkel, Zachary**, *Outlines of the Zemstvo Medical-Sanitary Work [oчерki zemskogo vrachebno-sanitarnogo dela]*, St. Petersburg: Typography of corporation "Slovo", 1913.

- Frieden, Nancy Mandelker**, *Russian physicians in an era of reform and revolution, 1856-1905*, Princeton, N.J.: Princeton University Press, 1981.
- Ganguli, Ishani, Frederick P. Rivara, and Sharon K. Inouye**, “Gender Differences in Electronic Health Record Work—Amplifying the Gender Pay and Time Gap in Medicine,” *JAMA Network Open*, March 2022, 5 (3), e223940.
- Gilyarovsky, Fedor**, “Studies on the birth and mortality of children in the Novgorod province,” in “Notes of the Imperial Russian Geographical Society in the Department of Statistics; Vol. 1,” St. Petersburg: Type. K. Wulf, 1866.
- Goldin, Claudia**, “A Grand Gender Convergence: Its Last Chapter,” *American Economic Review*, April 2014, 104 (4), 1091–1119.
- , “Hours Flexibility and the Gender Gap in Pay,” Technical Report 2015.
- and **Lawrence F. Katz**, “The Power of the Pill: Oral Contraceptives and Women’s Career and Marriage Decisions,” *Journal of Political Economy*, August 2002, 110 (4), 730–770.
- Gollin, Douglas, David Lagakos, and Michael E. Waugh**, “The Agricultural Productivity Gap,” *The Quarterly Journal of Economics*, May 2014, 129 (2), 939–993.
- Goodman-Bacon, Andrew**, “Difference-in-differences with variation in treatment timing,” *Journal of Econometrics*, December 2021, 225 (2), 254–277.
- Gosteli, Martina**, “The first female medical students at the University of Zurich,” Blog der Hauptbibliothek March 2019. <https://www.uzh.ch/blog/hbz/2019/03/11/the-first-female-medical-students-at-the-university-of-zurich/?lang=en>.
- Gradstein, Mark**, “Misallocation of talent and human capital: Political economy analysis,” *European Economic Review*, September 2019, 118, 148–157.
- Greenwood, Brad N., Seth Carnahan, and Laura Huang**, “Patient–physician gender concordance and increased mortality among female heart attack patients,” *Proceedings of the National Academy of Sciences*, August 2018, 115 (34), 8569–8574. Publisher: Proceedings of the National Academy of Sciences.
- Hoehn-Velasco, Lauren**, “Explaining declines in US rural mortality, 1910–1933: The role of county health departments,” *Explorations in Economic History*, October 2018, 70, 42–72.
- Hollingsworth, Alex, Krzysztof Karbownik, Melissa A. Thomasson, and Anthony Wray**, “The Gift of a Lifetime: The Hospital, Modern Medicine, and Mortality,” *American Economic Review*, July 2024, 114 (7), 2201–2238.
- Hsiang, Solomon M.**, “Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America,” *Proceedings of the National Academy of Sciences*, August 2010, 107 (35), 15367–15372.
- Hsieh, Chang-Tai and Peter J. Klenow**, “Misallocation and Manufacturing TFP in China and India*,” *The Quarterly Journal of Economics*, November 2009, 124 (4), 1403–1448.
- , **Erik Hurst, Charles I. Jones, and Peter J. Klenow**, “The Allocation of Talent and U.S. Economic Growth,” *Econometrica*, 2019, 87 (5), 1439–1474.

- Iaria, Alessandro, Carlo Schwarz, and Fabian Waldinger**, “Gender Gaps in Academia: Global Evidence Over the Twentieth Century,” SSRN Working Paper 4150221, SSRN July 2024.
- Imperial Chancellery**, *Full Collection of the Law of the Russian Empire. 1720-1722 [Polnoe Sobranie Zakonov Rossiiskoi Imperii. 1720-1722]*, Vol. 6, Saint Petersburg: Typography of II Department of His Majesty’s Chancellery, 1830.
- , *Full Collection of the Law of the Russian Empire [Polnoe Sobranie Zakonov Rossiiskoi Imperii]*, Vol. 19 of *Series II*, Saint Petersburg: Typography of II Department of His Majesty’s Chancellery, 1867. Statute 40457.
- Ivanov, Ivan Savvich**, *Movement of Population in Kostroma Province in 1880 and Causes of Death*, Kostroma: Kostroma Province Statistical Committee, 1883.
- Jakobson, Sergej A.**, *Zemstvo Physician Alexandra Arkhangelskaya: 1851-1905 [Zemskij vrach Aleksandra Gavrilovna Arhangel’skaja: 1851-1905]*, Moscow: Medgiz, 1958.
- Jayachandran, Seema, Adriana Lleras-Muney, and Kimberly V. Smith**, “Modern Medicine and the Twentieth Century Decline in Mortality: Evidence on the Impact of Sulfa Drugs,” *American Economic Journal: Applied Economics*, April 2010, 2 (2), 118–146.
- and **Ilyana Kuziemko**, “Why do mothers breastfeed girls less than boys? Evidence and implications for child health in India,” *The Quarterly Journal of Economics*, 2011, 126 (3), 1485–1538.
- Jerant, Anthony, Klea D. Bertakis, Joshua J. Fenton, Daniel J. Tancredi, and Peter Franks**, “Patient-provider sex and race/ethnicity concordance: a national study of healthcare and outcomes,” *Medical Care*, November 2011, 49 (11), 1012–1020.
- Kane, Thomas J. and Douglas O. Staiger**, “Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation,” NBER Working Paper No. w14607 December 2008.
- Kessler, Gijs and Andrei Markevich**, “Electronic Repository of Russian Historical Statistics, 18th - 21st centuries, <https://ristat.org/>, Version I (2020): Russian Empire Historical GIS Maps (1897),” IISH Data Collection February 2020. <https://datasets.iisg.amsterdam/dataset.xhtml?persistentId=hdl:10622/DN9QDM>.
- Kesztenbaum, Lionel and Jean-Laurent Rosenthal**, “Sewers’ diffusion and the decline of mortality: The case of Paris, 1880–1914,” *Journal of Urban Economics*, March 2017, 98, 174–186.
- Kissel, Helen, Tamar Matiashvili, and Helena Roy**, “Worth the Wait? Willingness to Pay for Gender-Concordant Physicians Across Medical Specialties,” Working Paper 2026.
- Klasen, Stephan**, “What Explains Uneven Female Labor Force Participation Levels and Trends in Developing Countries?,” *World Bank Research Observer*, 2019, 34 (2), 161–197.
- Krähenmann-Müller, Simone, Vanessa S. Virgini, Manuel R. Blum, Bruno R. da Costa, Tinh-Hai Collet, Yonas Martin, Jacques Cornuz, Lukas Zimmerli, Jean-Michel Gaspoz, Douglas C. Bauer, Eve A. Kerr, Drahomir Aujesky, and Nicolas Rodondi**, “Patient and physician gender concordance in preventive care in university primary care settings,” *Preventive Medicine*, October 2014, 67, 242–247.

- Kviatkovskaia, Iulia**, *Memories of Doctors Iulia Al. Kvyatkovskaya and Maria P. Rashkovich and short biographies of N.A. and S.P. Doroshevskis, E.P. Dzhunkovskaya and E.I. Kristi-Sitsinskaya [Vospominaniya vrachej Julii Al. Kyjatkovskoy i Marii P. Rashkovich i kratkija biografii N.A. i S.P. Doroshevskih, E.P. Dzhunkovskoi i E.I. Kristi-Sicinskoj]*, Dom Knigi, 1937. Google-Books-ID: 0_HTAAAAMAAJ.
- Lagakos, David and Michael E. Waugh**, “Selection, Agriculture, and Cross-Country Productivity Differences,” *American Economic Review*, April 2013, 103 (2), 948–980.
- Lau, Emily S., Sharonne N. Hayes, Annabelle Santos Volgman, Kathryn Lindley, Carl J. Pepine, and Malissa J. Wood**, “Does Patient-Physician Gender Concordance Influence Patient Perceptions or Outcomes?,” *JACC*, March 2021, 77 (8), 1135–1138. Publisher: American College of Cardiology Foundation.
- Lautenberger, Diana M. and Valerie M. Dandar**, “The State of Women in Academic Medicine 2023–2024: Progressing Toward Equity,” Technical Report, Association of American Medical Colleges, Washington, DC November 2024.
- Lee, Hanna, Mary K. Tan, Andrew T. Yan, Paul Angaran, Paul Dorian, Claudia Bucci, Jean C. Gregoire, Alan D. Bell, Martin S. Green, Peter L. Gross, Allan Skanes, Charles R. Kerr, L. Brent Mitchell, Jafna L. Cox, Vidal Essebag, Brett Heilbron, Krishnan Ramanathan, Carl Fournier, Bruce H. Wheeler, Peter J. Lin, Murray Berall, Anatoly Langer, Lianne Goldin, Shaun G. Goodman, and FREEDOM AF and CONNECT AF Investigators**, “Association Between Patient and Physician Sex and Physician-Estimated Stroke and Bleeding Risks in Atrial Fibrillation,” *The Canadian Journal of Cardiology*, February 2019, 35 (2), 160–168.
- Lee, Munseob**, “Allocation of Female Talent and Cross-Country Productivity Differences,” *The Economic Journal*, June 2024, 134 (664), 3333–3359.
- Liebert, Helge and Beatrice Mäder**, “Physicians and the Production of Health: Returns to Health Care during the Mortality Transition,” *The Review of Economics and Statistics*, February 2026, 108 (2), 452–469.
- Lorentzon, Louise and Per Pettersson-Lidbom**, “Midwives and Maternal Mortality: Evidence from a Midwifery Policy Experiment in 19th-century Sweden,” *Journal of the European Economic Association*, August 2021, 19 (4), 2052–2084.
- Loudon, Irvine**, “Ignaz Phillip Semmelweis’ studies of death in childbirth,” *Journal of the Royal Society of Medicine*, November 2013, 106 (11), 461–463.
- Lubczyk, Moris and Petra Moser**, “Rosie the Researcher: How Female WWII Entrants Influenced Science in US Firms,” Working Paper 2025.
- López, José Joaquín and Jesica Torres**, “Size-dependent policies, talent misallocation, and the return to skill,” *Review of Economic Dynamics*, October 2020, 38, 59–93.
- Malein, Viktor and Tamar Matiashevili**, “Improvements in Urban Sanitation and Child Mortality Decline: The Case of the Russian Empire,” Working Paper 2024.

- , —, and **Francisco Beltrán Tapia**, “Culture, Economic Stress, and Missing Girls,” CEPR Discussion Paper No. DP18761 2024.
- Marein, Brian**, “Public health departments and the mortality transition in Latin America: Evidence from Puerto Rico,” *Journal of Development Economics*, January 2023, 160, 102980.
- Matiashvili, Tamar**, “Did Doctors Matter? Physician Value-Added and Medical Innovations,” Working Paper 2026.
- McDevitt, Ryan and James W. Roberts**, “Market Structure and Gender Disparity in Health Care: Preferences, Competition, and Quality of Care,” *The RAND Journal of Economics*, 2014, 45 (1), 68–93.
- Meigs, Charles D.**, *Females and their diseases: a series of letters to his class*, Philadelphia: Lea and Blanchard, 1848.
- Ministry of Internal Affairs**, *Russian Medical List*, Saint Petersburg: Ministry of Internal Affairs Typography, 1885.
- , “Report of the Medical Department of the Ministry of Internal Affairs,” Technical Report, Typography of the Ministry of Internal Affairs, St. Petersburg 1889.
- , *Russian Medical List*, Saint Petersburg: Ministry of Internal Affairs Typography, 1890.
- , *Russian Medical List*, Saint Petersburg: Ministry of Internal Affairs Typography, 1911.
- Moehling, Carolyn M., Gregory T. Niemesh, and Melissa A. Thomasson**, “Shut down and shut out: women physicians in the era of medical education reform,” *Cliometrica*, April 2026.
- Mokyr, Joel**, *The Gifts of Athena: Historical Origins of the Knowledge Economy*, Princeton N.J.: Princeton University Press, January 2005.
- Nafziger, Steven**, “Did Ivan’s vote matter? The political economy of local democracy in Tsarist Russia,” *European Review of Economic History*, December 2011, 15 (3), 393–441.
- Natkhov, Timur and Natalia Vasilenok**, “Ethnic-specific infant care practices and infant mortality in late Imperial Russia,” *The Economic History Review*, 2023, 76 (3), 783–806.
- Nimura, Janice P.**, *The doctors Blackwell: how two pioneering sisters brought medicine to women—and women to medicine*, first ed., New York, N.Y: W.W. Norton & Company, 2021.
- Nolen, Haley A., Justin Xavier Moore, Joel B. Rodgers, Henry E. Wang, and Lauren A. Walter**, “Patient Preference for Physician Gender in the Emergency Department.,” *The Yale Journal of Biology and Medicine*, 6 2016, 89 (2), 131–42.
- Novoselskii, Sergei Aleksandrovich**, *Mortality and Life Expectancy in Russia [Smertnost i Prodolzhitel'nost Zhizni v Rossii]*, Petrograd: Typography of the Ministry of Internal Affairs, 1916.
- Okeke, Edward N.**, “When a Doctor Falls from the Sky: The Impact of Easing Doctor Supply Constraints on Mortality,” *American Economic Review*, March 2023, 113 (3), 585–627.
- Pigeard-Micault, Natalie**, “A History of Women’s Entrance Into Medicine — Medica — BIU Santé, Paris,” October 2007. <https://www.biusante.parisdescartes.fr/histoire/medica/presentations/entree-femmes-en-medecine-en.php> (accessed April 14, 2025).

- Preston, Samuel, Patrick Heuveline, and Michel Guillot**, *Demography: Measuring and Modeling Population Processes*, Malden, MA: Wiley-Blackwell, 2001.
- Public Commissariat for Healthcare, Russian Soviet Federative Socialist Republic**, *Directory of Medical Doctors of the USSR (as of January 1, 1924)*, Saint Petersburg: Tipografia Gosud. Izdat. imeni Gutenberga. Leningrad, Stremiannaia, 12, 1925.
- Ramakrishnan, Aditi, Dana Sambuco, and Reshma Jagsi**, “Women’s Participation in the Medical Profession: Insights from Experiences in Japan, Scandinavia, Russia, and Eastern Europe,” *Journal of Women’s Health*, November 2014, 23 (11), 927–934.
- Rambachan, Ashesh and Jonathan Roth**, “A More Credible Approach to Parallel Trends,” *The Review of Economic Studies*, October 2023, 90 (5), 2555–2591.
- Reeder, Lydia**, *The Cure for Women: Dr. Mary Putnam Jacobi and the Challenge to Victorian Medicine That Changed Women’s Lives Forever*, New York: St. Martin’s Press, 2024.
- Restuccia, Diego and Richard Rogerson**, “The Causes and Costs of Misallocation,” *Journal of Economic Perspectives*, August 2017, 31 (3), 151–174.
- Ruggles, Steven, Catherine A. Fitch, Ronald Goeken, Josiah Grover, J. David Hacker, Matt Nelson, Jose Pacas, Evan Roberts, and Matthew Sobek**, “IPUMS Restricted Complete Count Data,” 2020. Minneapolis: University of Minnesota.
- Saaritsa, Sakari, Eero Simanainen, and Markus Ristola**, “Nurses, doctors, and mortality: the effectiveness of early health professionals in rural Finland, 1880–1938,” *European Review of Economic History*, February 2024, 28 (1), 91–119.
- Schaede, Ursina and Ville Mankki**, “Quota versus Quality? Long-Term Gains from an Unusual Gender Quota,” *American Economic Review*, 2026, 116 (6), 1996–2037.
- Scheid, Teresa L. and Galen H. Smith**, “Is physician-patient concordance associated with greater trust for women of low socioeconomic status?,” *Women & Health*, July 2017, 57 (6), 631–649.
- Schmittdiel, Julie A., Ana Traylor, Connie S. Uratsu, Carol M. Mangione, Assiamira Ferrara, and Usha Subramanian**, “The Association of Patient-Physician Gender Concordance with Cardiovascular Disease Risk Factor Control and Treatment in Diabetes,” *Journal of Women’s Health*, December 2009, 18 (12), 2065–2070.
- Smith, Cecelia**, “Admitted to Medical School as a Joke,” <https://medicine.yale.edu/news-article/its-national-women-physicians-day-admitted-joke/> February 2022.
- Sun, Liyang and Sarah Abraham**, “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects,” *Journal of Econometrics*, December 2021, 225 (2), 175–199.
- Suschinsky, Petr**, *Zhenshhina-Vrach v Rossii: Ocherk Desjatiletija Zhenskikh Vrachebnykh Kursov 1872-1882 [Female Doctor in Russia: Essay on Ten Years of Female Doctors’ Courses 1872-1882]*, St. Petersburg: Typography of N. A. Levedev, 1883.
- Tarnovskaya, Proskovya**, “Female Medical Institute and Female Medical Courses,” *European Herald [Vestnik Evropei]*, 1903, 4 (8), 497–517.

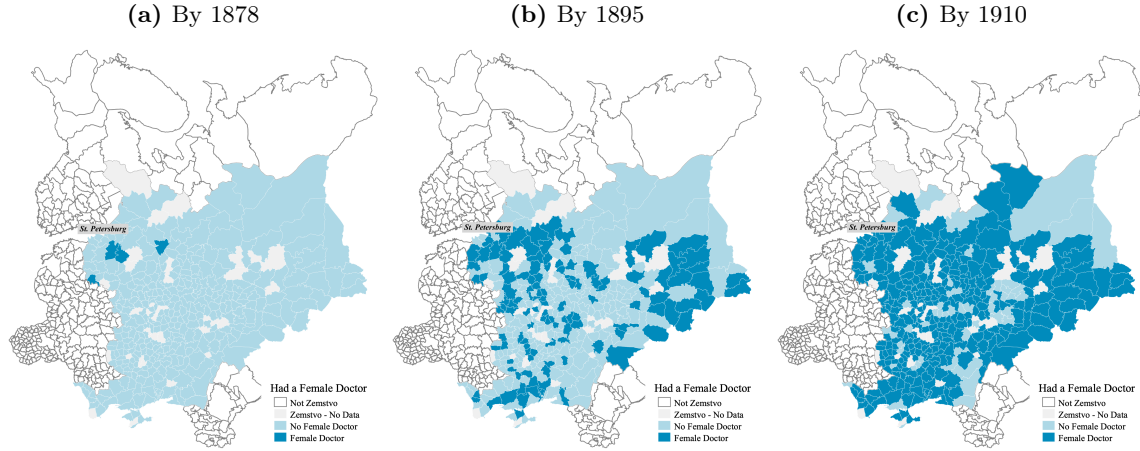
- Varadarajulu, Shyam, Carol Petruff, and William H. Ramsey**, “Patient Preferences for Gender of Endoscopists,” *Gastrointestinal Endoscopy*, 2002, 56 (2), 170–173.
- Veretennikova, A. I.**, *Notes of a Zemstvo Doctor [Zapiski zemskogo vracha]* Bashkirskaiia Lenini-
ana, Ufa: Bashkirskoe knizhnoe izd-vo, 1984.
- Vladislavlev, S. V. and V. V. Podvisotskii**, “Zemstvo Board of Poshekhonskii District Job Advertisement,” *Russkii Vrach*, July 1910, p. 16. Vol. 27.
- Winegarden, C. R. and John E. Murray**, “Effects of early health-insurance programs on European mortality and fertility trends,” *Social Science & Medicine*, May 2004, 58 (10), 1825–1836.
- World Health Organization**, “National Health Workforce Accounts (NHWA) Data Portal,” <https://apps.who.int/nhwportal/> 2025. Accessed: 2025-06-30.
- Yang, Hsing-Yu, Hsin-Jen Chen, Jill A. Marsteller, Lan Liang, Leiyu Shi, and Youfa Wang**, “Patient-health care professional gender or race/ethnicity concordance and its association with weight-related advice in the United States,” *Patient Education and Counseling*, February 2016, 99 (2), 271–278.
- Zhbankov, Dmitrii Nikolaevich**, *Zemstvo-Medical Collection [Zemsko-Meditsinskii Sbornik]*, Moscow: Tipografia D.I. Inozemtseva, Arbat, d. Karinskoi, 1891.

8 Exhibits

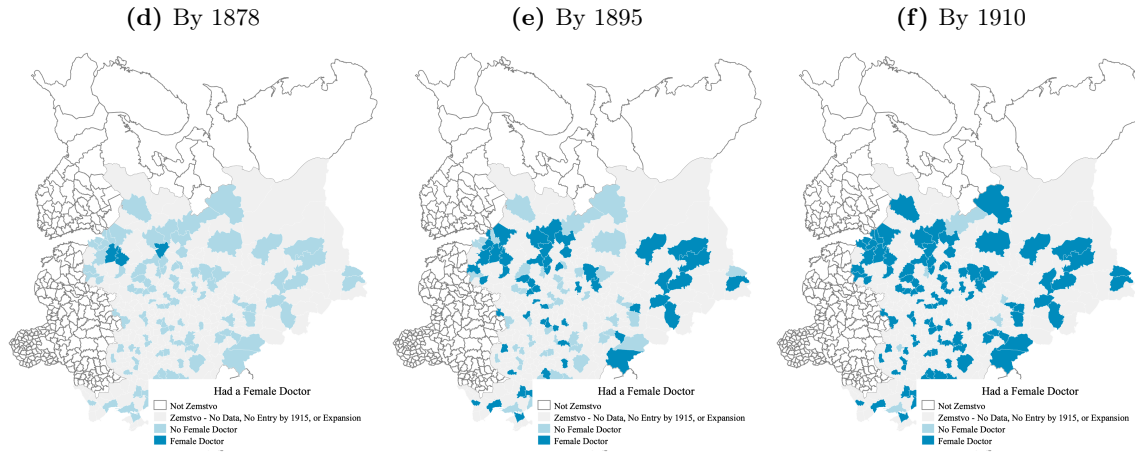
Figures

Figure 1: Geographic and Temporal Diffusion of Female Physician Entry

Panel A - Full Sample

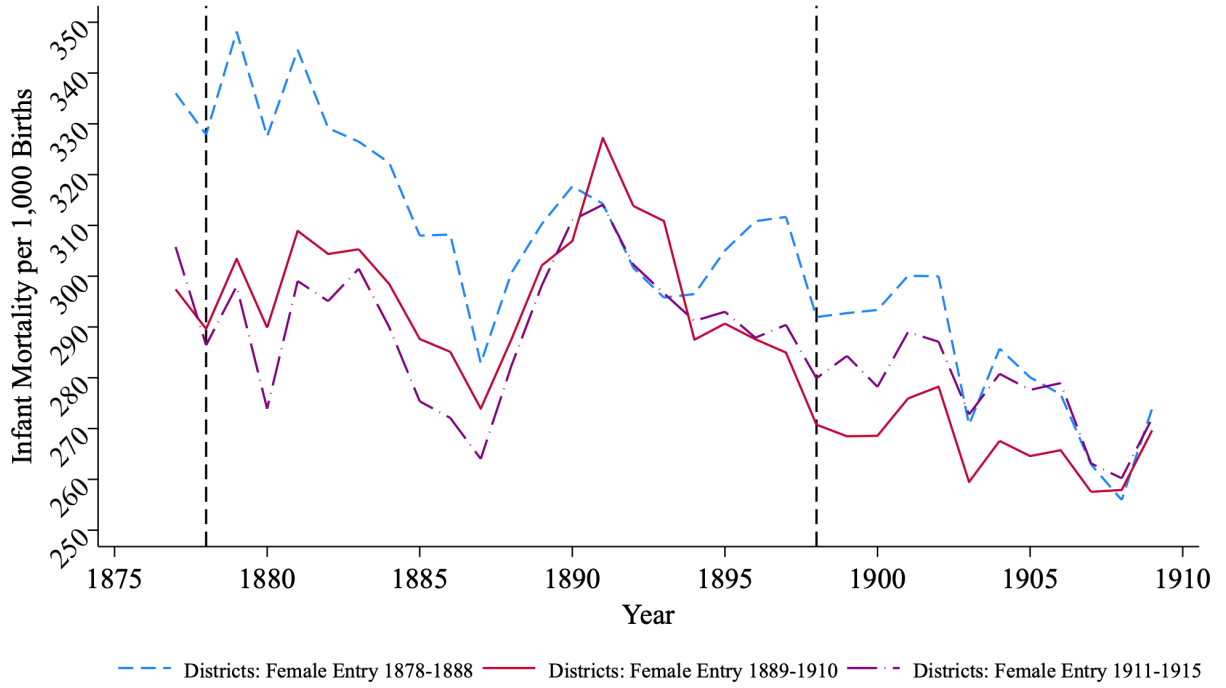


Panel B - Analysis Sample



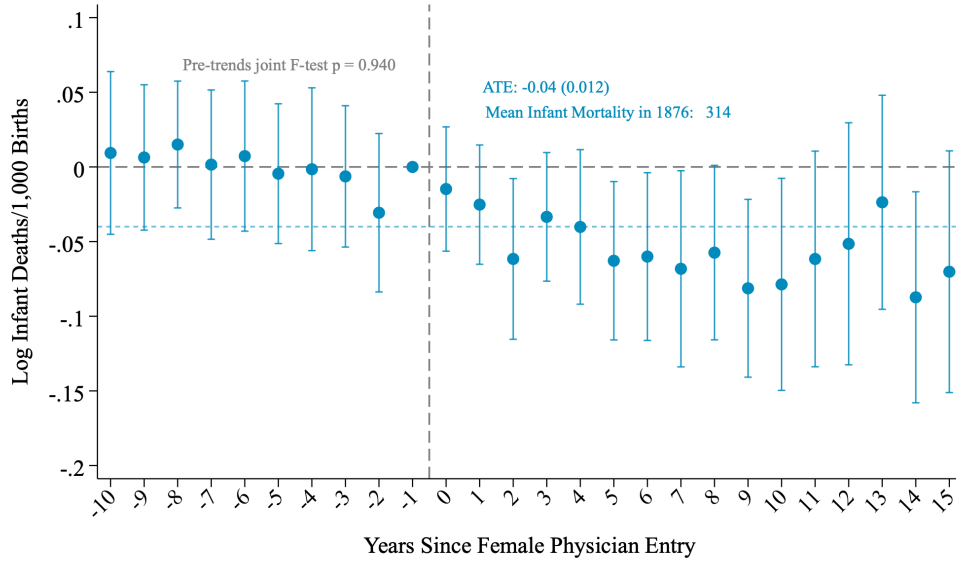
Notes: Panels (a)–(c) show the geographic diffusion of first female physician entry across the full sample of 336 Zemstvo districts, while panels (d)–(f) restrict to the 107-district analysis sample. The analysis sample includes districts whose first observed female physician hire through 1915 is classified as a replacement hire, defined as the total physician employment not increasing at the time of entry. Each panel indicates whether a district had hired its first female physician by that year. Physician-entry data come from the Zemstvo Medical Collection for 1876–1888 and the annual Medical Directories of the Russian Empire for 1889–1915; see Section 3 for details. Historical district boundaries are from [Kessler and Markevich \[2020\]](#).

Figure 2: Infant Mortality Trends by Timing of Female Physician Entry



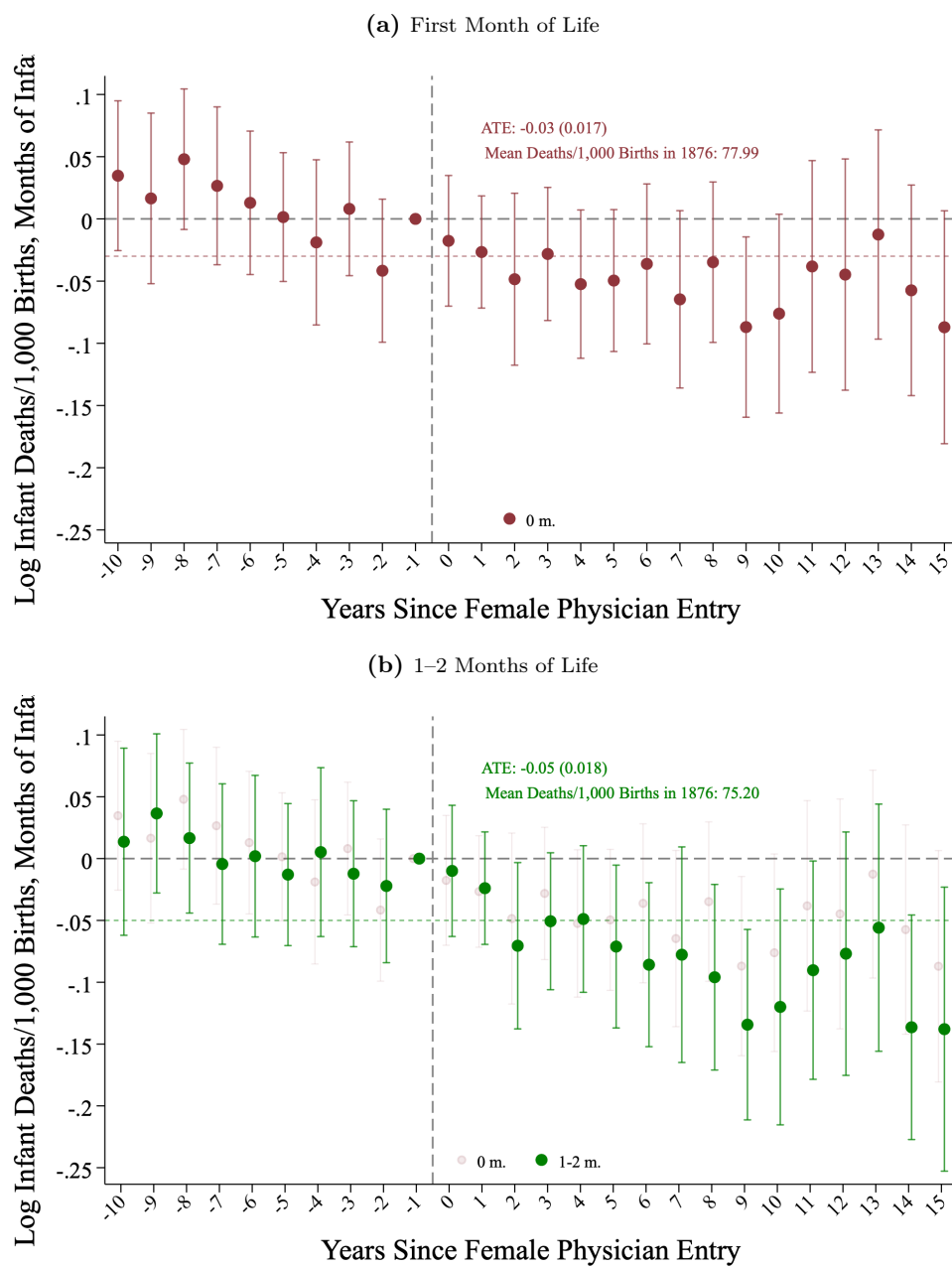
Notes: The underlying unit of observation is a district-year. Each series plots average infant mortality across districts grouped by the timing of their first female physician entry within the analysis sample. Early adopters first hired a female physician in 1878–1888 and account for 45.8% of the sample. Late adopters first hired one in 1889–1910 and account for 48.6% of the sample. Future adopters first hired one in 1911–1915, after the mortality outcome panel ends, and account for 5.6% of the sample. The figure reports centered three-year moving averages and is restricted to the 103 districts with non-missing infant mortality in every year 1876–1910. The underlying data were digitized for this project, as described in Section 3.

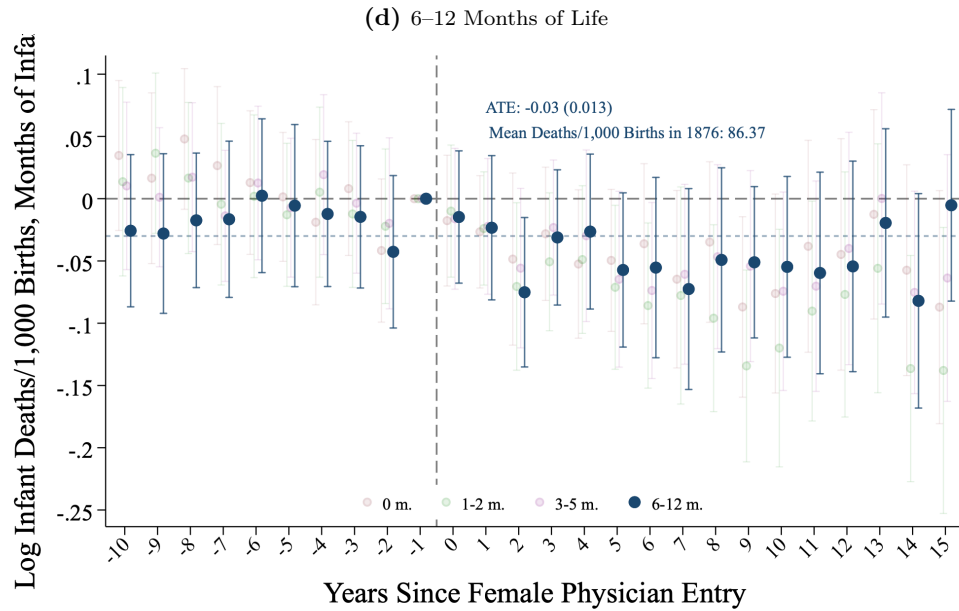
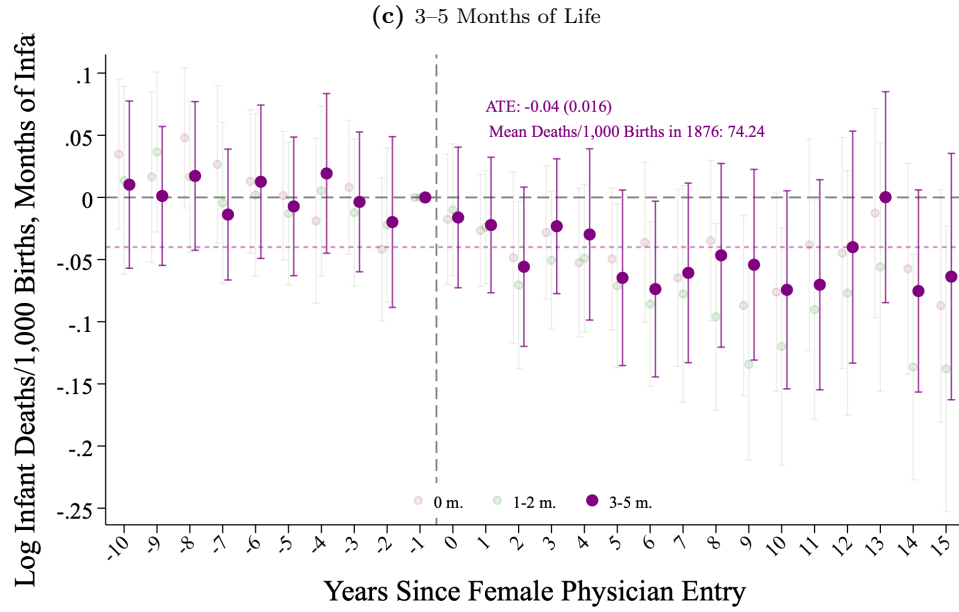
Figure 3: Female Physician Entry and Infant Mortality



Notes: This figure reports event-study estimates of equation 3, with log of infant deaths/1,000 births as the dependent variable. The sample contains 107 districts whose first female physician hire through 1915 is classified as a replacement hire and covers 3,679 district-year observations from 1876–1910. Six districts first hire a female physician in 1911–1915 and therefore contribute only untreated observations during the mortality panel. All specifications include district and year fixed effects and control for the total number of physicians. Event time is measured relative to the first female physician entry in a district. The omitted category is the year immediately preceding that entry. The edge years pool event times at or before -10 and at or after +15. Points report coefficient estimates. Vertical bars report 95% confidence intervals. Standard errors are clustered at the district level. The horizontal treatment-effect line and reported ATE correspond to the average treatment effect estimated from equation 1.

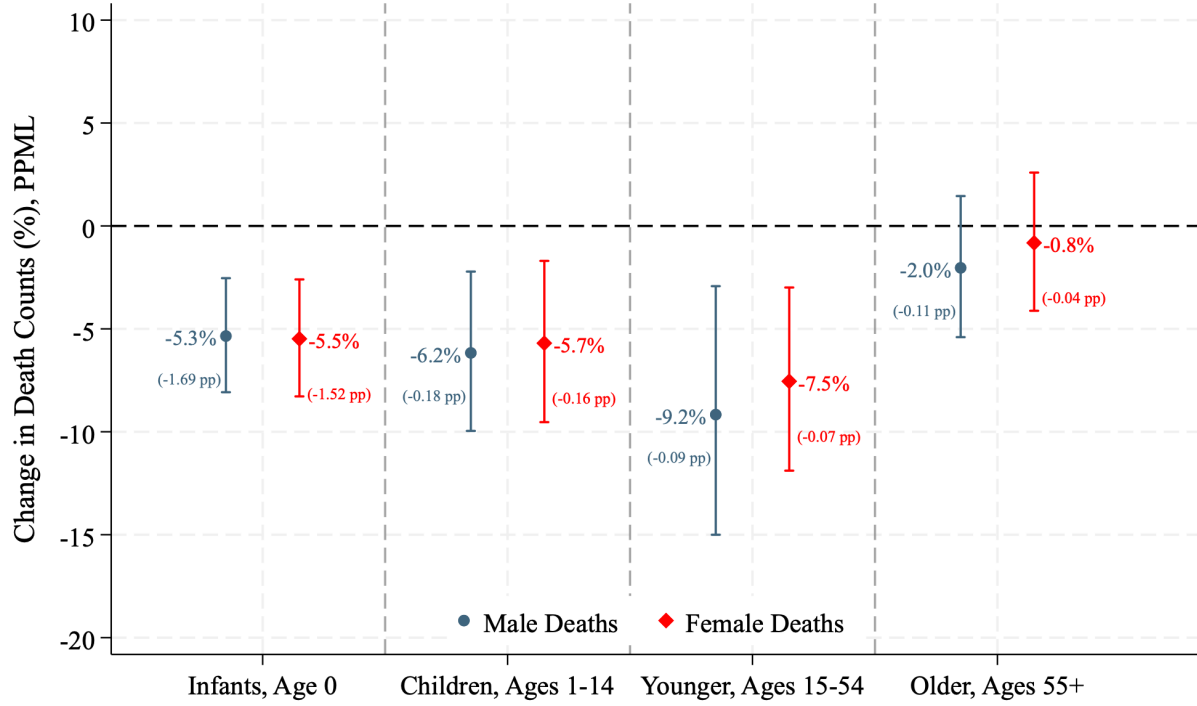
Figure 4: Female Physician Entry and Infant Mortality by Month of Life





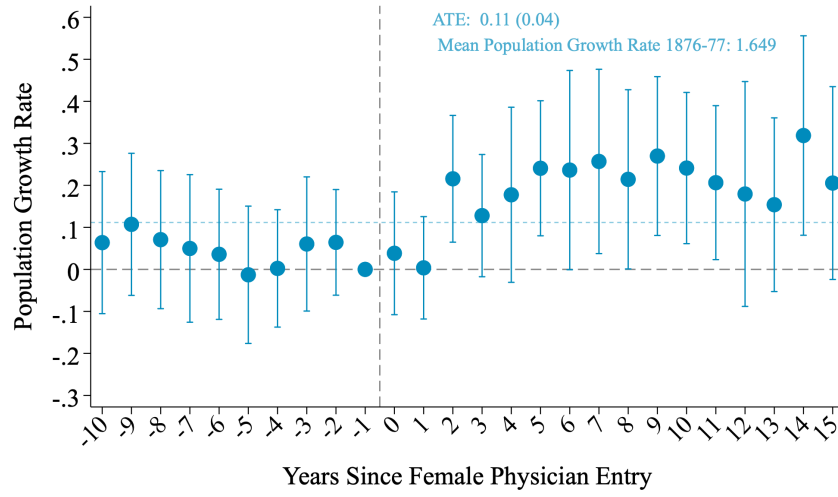
Notes: The figure reports event-study estimates for mortality within four intervals of infancy: the first month of life, months 1–2, months 3–5, and months 6–12 - age groups reported in the original data source. In each case, the dependent variable is the log of deaths in the specified age interval divided by total births in the district-year. Panels successively add each age interval, with previously displayed series retained in lighter colors for comparison. The sample and specification follow Figure 3: all regressions include district and year fixed effects and control for the total number of physicians, the year immediately preceding first female physician entry is omitted, and event times at or before -10 and at or after +15 are pooled. Vertical bars report 95% confidence intervals based on standard errors clustered at the district level. The reported ATE in each panel is the corresponding average treatment effect from equation 1.

Figure 5: Female Physician Entry and Mortality by Age and Sex



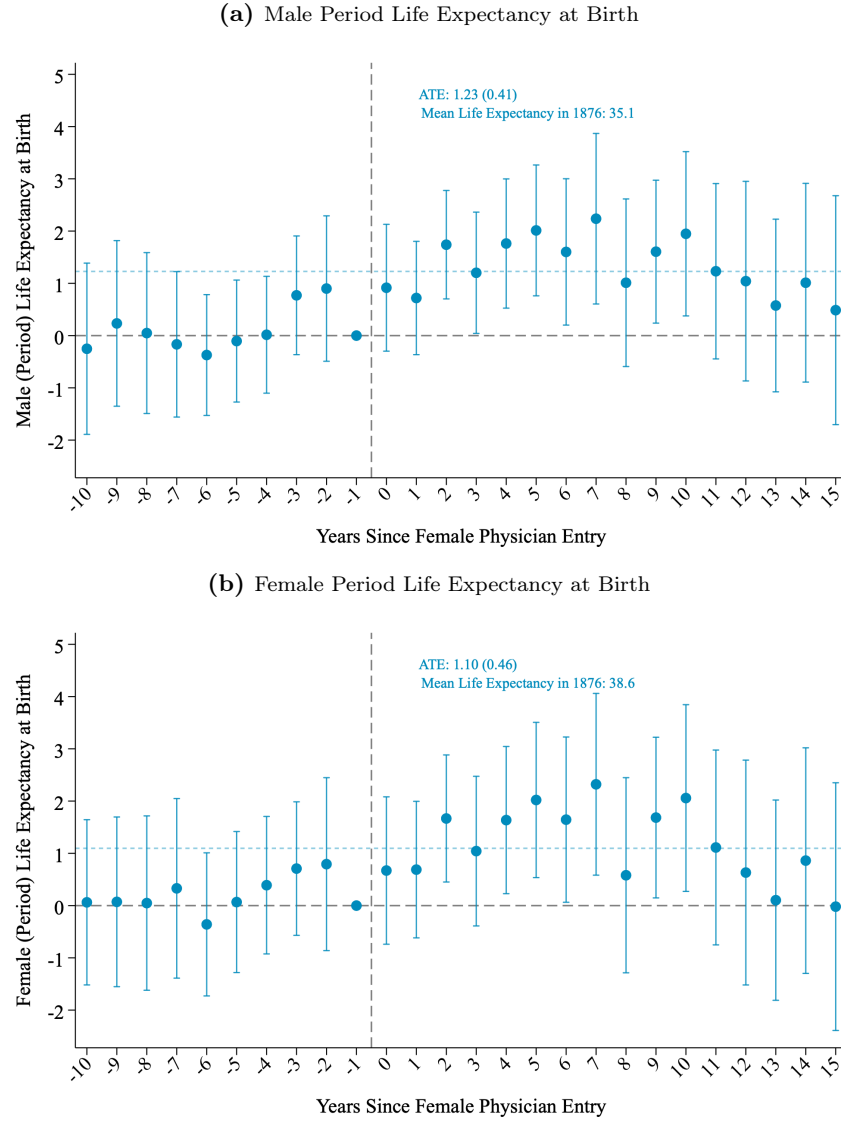
Notes: The figure reports average treatment effects of first female physician entry on annual death counts by age and sex, estimated using the PPML specification in equation 2. Age groups are infancy (age 0), childhood (ages 1-14), youth and early adulthood (ages 15-54), and older adulthood (ages 55 and above). Markers report the PPML estimates transformed into percent changes in deaths, $100[\exp(\beta) - 1]$, and vertical bars report transformed 95% confidence intervals. Values in parentheses report the implied percentage-point change in annual mortality obtained by applying each estimated proportional effect to the corresponding sex- and age-specific 1896 mortality rate, constructed using population counts from the 1897 Census. The sample is the replacement-hire analysis sample, with outcomes covering 1876-1910. All specifications include district and year fixed effects and control for the total number of physicians. Standard errors are clustered at the district level.

Figure 6: Female Physician Entry and Population Growth Rate



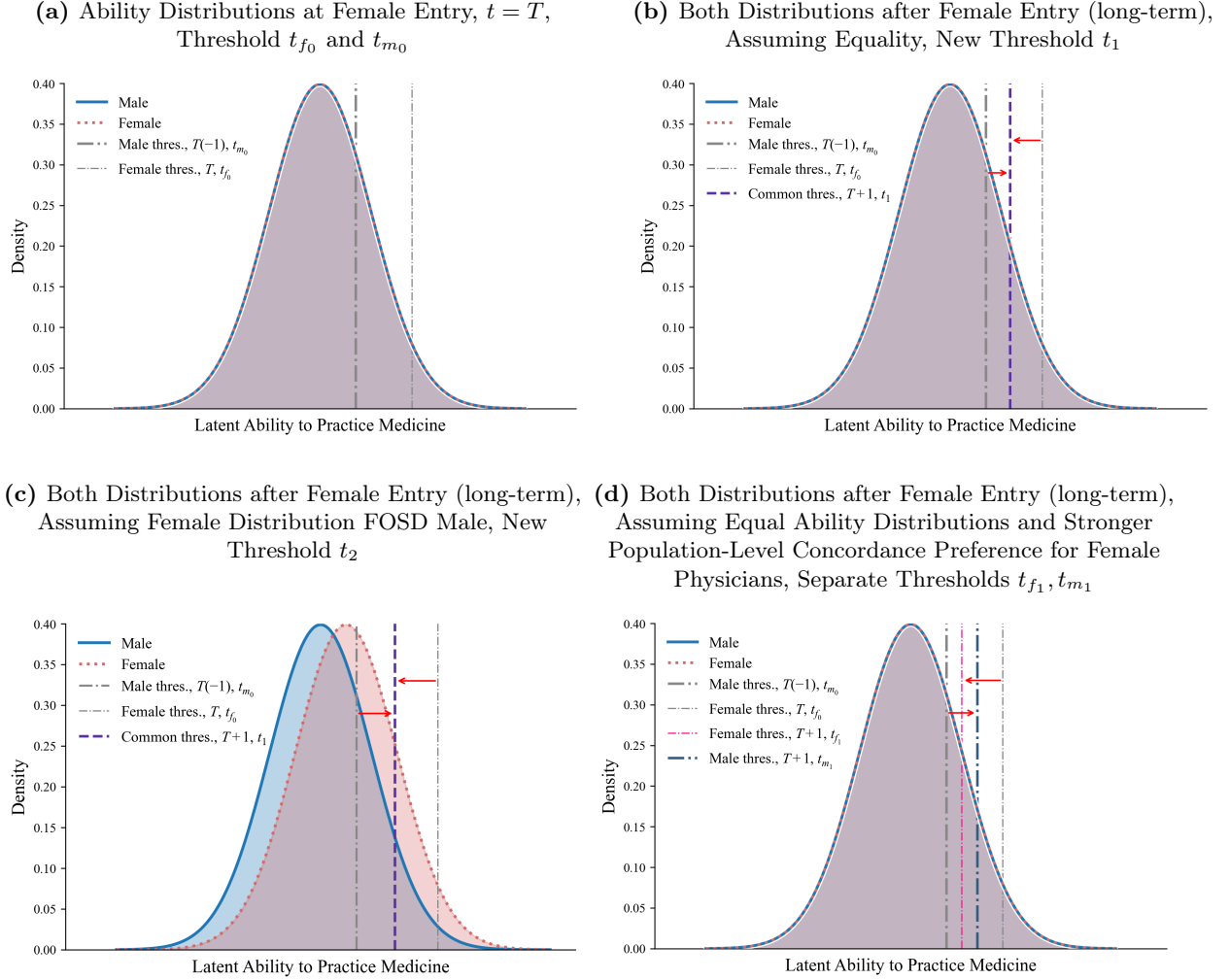
Notes: The figure reports event-study estimates of the effect of first female physician entry on the annual natural population growth rate. Natural population growth is defined as births minus deaths divided by the lagged district population. District populations are anchored to the January 1897 Census, treated as representing end-of-1896 population, and reconstructed backward and forward using annual births and deaths. Births are linearly interpolated in the 2% of district-years in which they are missing; all such observations occur between 1876 and 1883. The sample and event-study specification otherwise follow Figure 3, including district and year fixed effects, a control for the total number of physicians, an omitted event year of -1, endpoint bins at -10 and +15, and district-clustered standard errors. Vertical bars report 95% confidence intervals. The average treatment effect is 0.11 percentage points ($SE = 0.04$), equivalent to 6.7% of the pre-treatment mean of 1.65%.

Figure 7: Female Physician Entry and Life Expectancy at Birth



Notes: The figure reports event-study estimates of the effect of first female physician entry on approximate period life expectancy at birth, separately for men and women. Annual district-level life expectancy is constructed by indirect standardization using age- and sex-specific population counts from the January 1897 Census and annual age- and sex-specific death counts; Appendix C.3 provides the full construction. The event-study specification follows equation 3 and includes district and year fixed effects and a control for the total number of physicians. The omitted category is the year immediately preceding first female physician entry, and the endpoints pool event times at or before -10 and at or after +15. Vertical bars report 95% confidence intervals based on standard errors clustered at the district level. The average treatment effects are 1.2 years ($SE = 0.41$; 3.5%) for men and 1.1 years ($SE = 0.46$; 2.8%) for women.

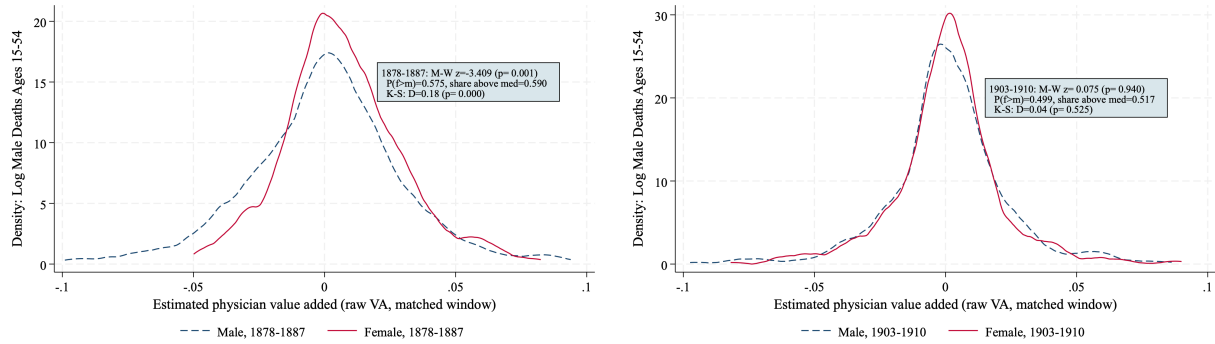
Figure 8: Evolution of Population Distributions of Latent Ability to Practice Medicine and Physician Entry Thresholds



Notes: Each panel illustrates the evolution of the population distributions of potential physician ability and the corresponding entry thresholds across three conceptual time periods: $T - 1$ (before female physician entry), T (the short-run effects of opening medical education to women), and $T + 1$ (the long-run equilibrium without barriers to entry). Panel (a) shows the initial distributions for women and men, respectively. The male physician threshold t_{m_0} reflects the prevailing entry cutoff prior to liberalization. The female threshold $t_{f_0} > t_{m_0}$ captures that the first female physicians entered from the upper tail of the ability distribution. Panel (b) depicts the long-run equilibrium when male and female ability distributions are assumed identical, yielding a common threshold t_1 such that $t_{f_0} > t_1 > t_{m_0}$. Panel (c) shows the case where the female ability distribution first-order stochastically dominates the male one. The new equilibrium threshold t_2 lies above t_1 , producing a higher overall talent threshold and a greater share of women among physicians, while maintaining parity in marginal quality across genders. Panel (d) presents the case with equal ability distributions and stronger concordance preference for female physicians compared to male physicians. This yields a greater share of women among physicians and two separate ability thresholds that satisfy: $t_{m_1} > t_{f_1}$.

Figure 9: Convergence in Estimated Value Added of Male and Female *Zemstvo* Physicians, 1889-1910

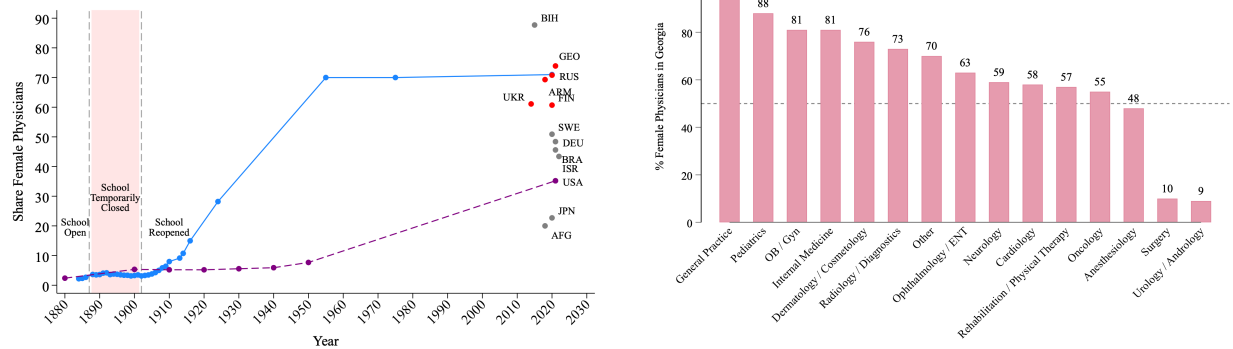
- (a) Deaths of Men Aged 15-54, by Physician Sex, Early MD Graduation Cohorts (b) Deaths of Men Aged 15-54, by Physician Sex, Late MD Graduation Cohorts



Notes: The figure plots kernel density estimates of raw physician value added based on mortality among men ages 15–54, separately by physician sex and graduation cohort. Value added is estimated from the log number of male deaths ages 15–54 in the linked physician panel for 1889–1910. District and year fixed effects are first removed from the outcome, and the resulting district-year residual is allocated equally across physicians practicing in that district-year. Physician value added is then constructed from leave-one-out spell-level residual allocations, as described in Appendix E.1; higher value added corresponds to lower mortality. Panel (a) compares physicians graduating in 1878–1887 and panel (b) compares physicians graduating in 1903–1910. Empirical Bayes shrinkage is not applied in this figure. To make the amount of observed career information more comparable across cohorts, the early cohort is restricted to its first six years in the panel, matching the average observation window of the late cohort. The Mann–Whitney statistic tests equality in the ranks of the male and female distributions; the reported probability is the share of all female–male physician pairs in which the female physician has higher estimated value added, with 0.5 indicating no difference. The Kolmogorov–Smirnov statistic tests equality of the two distributions, and “share above median” reports the fraction of female physicians above the male median. The x-axis is truncated to [−0.1, 0.1] for readability.

Figure 10: Female Share of Physicians over Time and by Specialty

- (a) Share of Female Physicians in the Russian Empire (1884–1916), USSR (1924–1975), the Russian Federation (2020), and Select Modern Countries (2010s–2020s). (b) Share of Female Physicians in the Republic of Georgia, by Specialty, 2025.



Notes: Panel (a) documents the female share of physicians over time. The solid blue series follows the Russian Empire, Soviet Union, and Russian Federation; the dashed US series provides a comparison, and the modern country markers report recent female physician shares, with highlighted markers identifying selected countries formerly part of the Russian Empire. Russian Empire shares for 1884–1916 are approximated from the number of pages devoted to male, female, and late-reporting physicians in the annual *Medical Directories*. The 1924 Soviet share is constructed analogously from the *Directory of Medical Doctors of the USSR* [Public Commissariat for Healthcare, *Russian Soviet Federative Socialist Republic*, 1925]; the 1955 estimate is from Ramakrishnan, Sambuco, and Jagsi [2014], the 1970 estimate is drawn from <https://chapaev.media/articles/11425>; and modern country estimates are from World Health Organization [2025]; US historical shares are calculated from complete-count decennial censuses Ruggles et al. [2020]. The shaded interval marks the period during which the women’s medical school was closed. Panel (b) reports female shares by medical specialty in the Republic of Georgia in 2025, using the near-universe of physicians listed on www.tsamali.ge. Physician gender is assigned from first names and specialties are taken from the platform.

Tables

Table 1: Summary Statistics

	Analysis Sample				
	(1)	(2)	(3)	(4)	(5)
	Median	Mean	SD	Districts	District-Years
Panel A: Physicians and Vital Statistics, 1876-1910					
<i>District Zemstvo:</i>					
Hired Its First Female Physician (Year)	1,893	1,894.60	11.87	107	3,745
Doctors	4	5.52	3.70	107	3,745
Births	7,349	9,283.36	6,441.84	107	3,745
Infant Deaths (Age 0) per 1,000 Births	286	293.80	89.59	107	3,679
<i>Within-Infancy Deaths per 1,000 Births, Months:</i>					
0	72	75.22	25.04	107	3,679
1-2	64	67.22	29.07	107	3,679
3-5	66	67.38	24.99	107	3,679
6-12	82	83.97	25.70	107	3,679
<i>Number Male Deaths, Ages:</i>					
0	1,126	1,417.23	984.42	107	3,680
1-14	674	903.31	760.65	107	3,680
15-54	386	553.38	819.68	107	3,679
55+	427	507.76	322.91	107	3,677
All	2,651	3,385.83	2,581.79	107	3,711
<i>Number Female Deaths, Ages:</i>					
0	945	1,199.12	853.70	107	3,680
1-14	652	873.13	738.12	107	3,568
15-54	398	529.65	492.21	107	3,678
55+	462	539.56	358.55	107	3,569
All	2,500	3,149.05	2,226.77	107	3,711
Panel B: Healthcare Provision, 1876-1888					
<i>District Zemstvo:</i>					
Midwives	2	2.83	2.58	103	1,336
Nurse-Midwives	0	0.89	1.68	102	1,326
Male Nurses	10	11.85	7.41	105	1,349
Female Nurses	0	0.38	2.03	105	1,365
Smallpox Vaccinators	4	8.86	10.45	71	873
Clinic Beds	42	49.76	32.61	103	1,330
Annual Medical Budget (Rubles, 1,000s)	18	20.61	12.53	106	1,327
Clinic Patient Death Rate, %	5	6.00	3.22	95	1,033
In-clinic Patients	525	650.10	843.31	103	1,281
Share of Births in Clinics	0	0.00	0.00	30	369
<i>Total Districts</i>				107	

Notes: This table reports summary statistics for the 107-district analysis sample, defined as districts whose first observed female physician hire through 1915 is classified as a replacement hire. Panel A pools district-year observations from 1876–1910 for physician and vital-statistics outcomes. Panel B reports healthcare-provision outcomes available from the *Zemstvo Medical Collection* for 1876–1888. Columns report the median, mean, standard deviation, the number of districts with non-missing observations, and the number of district-year observations. Infant and within-infancy mortality rates are expressed per 1,000 births; the clinic patient death rate is expressed in percent; and the annual medical budget is reported in thousands of rubles. Share of Births in Clinics is the number of births recorded in rural Zemstvo clinics divided by the total number of births in the district, with the denominator taken from the vital statistics. Physician counts and first-entry dates come from the *Zemstvo Medical Collection* for 1876–1888 and the annual *Medical Directories of the Russian Empire* for 1889–1915. Birth and death data come from annual editions of *The Movement of Population in European Russia*. Healthcare-provision measures come from the *Zemstvo Medical Collection*. Differences in sample size across outcomes reflect missing information in the original historical sources.

Table 2: Mapping Mechanisms to Outcomes

	Treatment Effect Difference ($\Delta\tau$)	Demand-Side Concordance Preference ($p_{f,f} > p_{f,m}$)
<i>Mortality Outcomes</i>		
Infant mortality	✓	✓
Adult female mortality	✓	✓
Adult male mortality	✓	×
<i>Substitute Labor Demand</i>		
Midwives employed	×	✓
<i>Healthcare Provision</i>		
In-clinic Patients	✓	✓
Share of Births in Clinics	✓	✓

Notes: This table summarizes the empirical predictions of the two mechanisms developed in Sections 4.1 and 6. The treatment effect difference mechanism refers to differences in the clinical effectiveness of female and male physicians. The demand-side concordance mechanism refers to greater willingness of female patients to seek formal medical care when the assigned physician is female, due to the physician's gender alone. A checkmark indicates that the corresponding mechanism can generate a change in this framework. A cross indicates that the mechanism does not directly predict such a change.

Table 3: Balance on Baseline Trends, 1876-1878

	1878-1888 Adopters			1889-1910 Adopters			1911-1915 Adopters			F-test p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Mean	SD	N	Mean	SD	N	Mean	SD	N	
Panel A: Physicians and Vital Statistics										
Doctors	.0045	.11	49	.027	.083	50	-.014	.057	6	0.41
Births	-.026	.046	43	-.035	.077	45	.035	.1	6	0.06
Infant Deaths (Age 0) per 1,000 Births	.069	.12	43	.061	.13	45	-.0062	.14	6	0.37
<i>Within-Infancy Deaths per 1,000 Births, Months:</i>										
0	.04	.16	43	.026	.14	45	-.033	.083	6	0.52
1-2	.043	.19	43	.023	.14	45	-.054	.072	6	0.39
3-5	.057	.1	43	.071	.17	45	-.059	.16	6	0.11
6-12	.18	.16	43	.15	.17	45	.16	.27	6	0.66
<i>Number Male Deaths, Ages:</i>										
0	.032	.14	44	.022	.13	45	.0074	.12	6	0.88
1-14	.17	.2	44	.21	.24	45	.1	.23	6	0.47
15-54	.034	.099	44	.051	.094	45	.0059	.062	6	0.47
55+	.051	.1	43	.022	.1	45	.027	.069	6	0.39
All	.048	.091	44	.064	.12	45	.024	.13	6	0.63
<i>Number Female Deaths, Ages:</i>										
0	.037	.15	44	.04	.15	45	.018	.16	6	0.94
1-14	.19	.22	44	.22	.26	45	.12	.29	6	0.58
15-54	.016	.1	44	.023	.1	45	.021	.045	6	0.95
55+	.055	.13	44	.018	.097	45	.014	.028	6	0.26
All	.048	.1	44	.067	.13	45	.022	.14	6	0.57
Panel B: Healthcare Provision (conditional on positive values)										
Midwives	.099	.22	37	.13	.35	40	.21	.74	5	0.77
Nurse-Midwives	-.033	.075	5	.14	.35	7	0	.	1	0.55
Male Nurses	.016	.095	48	.043	.093	47	.041	.057	6	0.37
Female Nurses	.083	.66	6	0	0	3	.	.	0	0.84
Smallpox Vaccinators	.056	.2	20	-.0045	.24	23	-.21	.37	3	0.17
Clinic Beds	.086	.25	43	.059	.12	43	.095	.17	6	0.77
Annual Medical Budget (Rubles, 1,000s)	.16	.22	34	.36	1.7	34	.041	.046	5	0.73
Clinic Patient Death Rate, %	.19	.4	29	-.00062	.35	28	.19	.26	3	0.16
In-clinic Patients	.21	.58	38	.12	.2	35	.13	.13	5	0.68
Share of Births in Clinics	-.47	.75	2	1.1	.00068	2	.	.	0	0.10

Notes: This table reports baseline trends in observables by timing of the first female physician hire. Baseline trends are calculated as mean annual growth rates over 1876–1878. Although three districts in the analysis sample had already hired a female physician by 1878, I retain the full 1876–1878 window. This is done so that the baseline trend measure averages two annual growth rates instead of relying on a single year-to-year change. Early adopters are districts with a first hire in 1878–1888 (45.8% of the sample), late adopters in 1889–1910 (48.6%), and future adopters in 1911–1915 (5.6%). Share of Births in Clinics is the number of births recorded in rural Zemstvo clinics divided by the total number of births in the district, with the denominator taken from the vital statistics. Columns 1–9 report group means, standard deviations, and numbers of districts; column 10 reports the p-value from a joint F-test of equality of means across groups. Variation in the number of districts across cells in Panel A reflects missing data, which can prevent calculation of the baseline growth measure. In Panel B, cells with especially few observations reflect either missing reporting or baseline values equal to zero, for which proportional growth rates are undefined. See the notes to Table 1 for data sources.

Table 4: Female Physician Entry and Infant Mortality

	Infant Deaths/1,000 Births				Log Infant Deaths/1,000 Births			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female Physician Entry	-13.665*** (4.035)	-13.659*** (4.039)	-0.562 (8.086)	-12.692** (4.876)	-0.038*** (0.012)	-0.038*** (0.012)	-0.001 (0.029)	-0.034** (0.015)
Female Physician Entry X Above-Median Share at Entry			-20.254* (10.790)				-0.058 (0.035)	
Female Physician Entry X Infant Mortality in 1876				-0.360*** (0.070)				-0.001*** (0.000)
N	3,679	3,679	3,679	3,414	3,679	3,679	3,679	3,414
Districts	107	107	107	98	107	107	107	98
Mean Y_{1876}	313.793	313.793	313.793	313.793	-1.220	-1.220	-1.220	-1.220
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
District FE	✓	✓	✓	✓	✓	✓	✓	✓
Controls		✓		✓		✓		✓
R-squared	0.619	0.619	0.620	0.609	0.672	0.673	0.673	0.676

Notes: This table reports average treatment effects of first female physician entry on infant mortality over 1876–1910 in the 107-district replacement-hire analysis sample. Columns (1)–(4) use infant deaths before age one per 1,000 births as the dependent variable; columns (5)–(8) use the logarithm of infant deaths divided by births. All specifications include district and year fixed effects. Columns (1) and (5) omit time-varying controls; columns (2)–(4) and (6)–(8) control for the total number of physicians. Columns (3) and (7) interact female physician entry with an indicator for whether the first female physician represented an above-median share of the district physician workforce at entry. Infant mortality in 1876 is mean-centered across districts, so the coefficient on Female Physician Entry in columns (4) and (8) gives the estimated treatment effect at the sample mean of 1876 infant mortality. “Mean Y_{1876} ” reports the mean dependent variable in 1876. Standard errors, clustered at the district level, are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Female Physician Entry and Healthcare Provision

	(1)	(2)	(3)
	In-clinic Patients	Share of Births in Clinics	Midwives
Panel A: Healthcare Provision and Access			
Female Physician Entry	0.205** (0.093)	0.002* (0.001)	-0.206** (0.094)
N	1,274	369	1,206
Districts	102	30	93
Raw Mean Y_{1876}	561.544	0.000	2.630
Year FE	✓	✓	✓
District FE	✓	✓	✓
Controls	✓	✓	✓
Estimator	PPML TWFE	TWFE	PPML TWFE
	(1)	(2)	(3)
	Male Nurses	Female Nurses	Nurse-Midwives
Panel B: Healthcare Personnel			
Female Physician Entry	-0.051 (0.047)	0.494 (0.353)	-0.140 (0.183)
N	1,336	390	793
Districts	104	30	61
Raw Mean Y_{1876}	10.165	0.800	0.443
Year FE	✓	✓	✓
District FE	✓	✓	✓
Controls	✓	✓	✓
Estimator	PPML TWFE	PPML TWFE	PPML TWFE

Notes: This table reports changes in healthcare provision following first female physician entry using outcomes available for 1876-1888. The sample follows the replacement-hire design used in the main analysis. Treatment is equal to one beginning in the year of first female physician entry. All specifications include district and year fixed effects and control for the total number of physicians. Standard errors are clustered at the district level. Count outcomes are estimated using Poisson Pseudo-Maximum Likelihood (PPML), while the Share of Births in Clinics is estimated using a conventional two-way fixed-effects specification. PPML coefficients are semi-elasticities. The proportional effect is given by $\exp(\beta) - 1$. For example, the coefficient of -0.206 on midwives in Panel A corresponds to an 18.6% decline, or approximately 0.49 fewer midwives relative to the 1876 mean of 2.63. The coefficient on Share of Births in Clinics is a level change in the share, so a coefficient of 0.002 corresponds to an increase of 0.2 percentage points. Share of Births in Clinics is the number of births recorded in rural *Zemstvo* clinics divided by the total number of births in the district, with the number of births taken from the vital statistics. “Raw Mean Y_{1876} ” reports the mean of the untransformed dependent variable in 1876. Sample sizes differ across outcomes because healthcare-provision measures are incompletely reported in the original *Zemstvo Medical Collection*. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

APPENDIX

A Historical Background

A.1 Social Origins of the First Female Physicians

The tabulation below is based on the table provided by [Suschinsky, 1883, p. 34], based on the first 8 years of admissions (1872-1879):

Table A.1.1: Social Origin Breakdown

Category	Number	Percent
Wives and daughters of civil officials	214	30.0%
Merchants and honorary citizens	133	18.7%
Military	107	15.0%
Townspeople (Meshchane)	105	14.7%
Clergy	59	8.3%
Wives and daughters of doctors and students	41	5.8%
Persons of free origin	20	2.8%
Daughters of lower ranks	13	1.8%
Peasants, colonists, and domestic servants	9	1.3%
Foreigners	8	1.1%
Guild members and artisans	4	0.6%

A.2 Costs to Patients, Examples

The table below is based on accounts of individual district operation descriptions found in Zhbankov [1891].

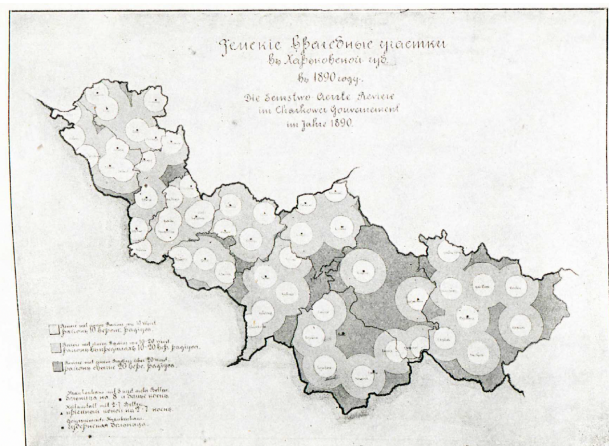
Patient Category	Arzamasskii, Nizhegorod	Kazanskii, Kazan	Gorokhovetskii, Vladimir
<i>Inhospital</i>			
Local	Free	Free	3 rubles/month
Non-local	20 kopecks/day	8 rubles 10 kopecks/month	3 rubles/month
<i>Ambulatory</i>			
Local poor	Free	Free	Free
Local not poor	Pay for medicine	Free	Free
Non-local poor	Free	8 rubles 10 kopecks/month	Free
Non-local non-poor	Pay for medicine	8 rubles 10 kopecks/month	Free

Table A.2.1: Comparison of patient charges across three district hospital examples

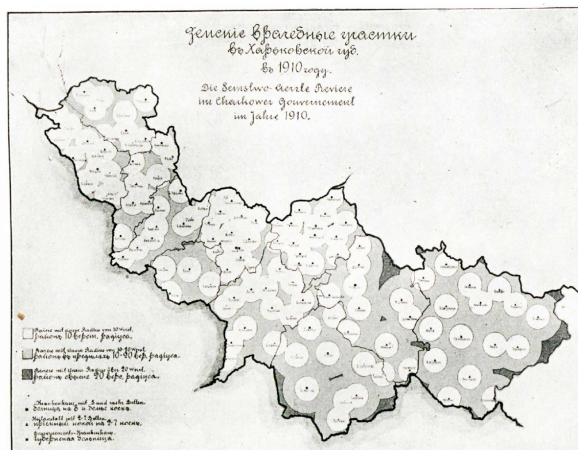
A.3 Doctors' Catchment Areas

Figure A.3.1: Example of Catchment Areas, Kharkiv Province, 1890 and 1910

(a) Zemstvo Doctor Catchment Areas, Kharkiv Province 1890



(b) Zemstvo Doctor Catchment Areas, Kharkiv Province 1910



Source: [Frenkel \[1913\]](#)

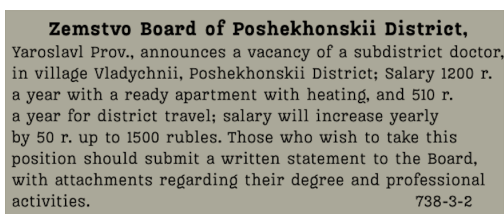
A.4 Job Advertisement Example

Figure A.4.1: Example of a Job Advertisement, 1910 [[Vladislavlev and Podvisotskii, 1910](#)]

(a) Job Ad, Original



(b) Job Ad, English Translation



A.5 Causes of Death

Table A.5.1: % of Causes of Death by Age Group and Disease in Kostroma Province, 1880 based on [Ivanov \[1883\]](#)

	0-6 mo	6-12 mo	1-15 y F	1-15 y M	16-49 y F	16-49 y M	50+ y F	50+ y M
Abdominal Diseases	50.5	38.3	14.6	14.7	2.6	3	1.5	2.2
Inherent Weakness	31.3	1.2	0	0	0	0	0	0
Other Infectious Diseases	12.2	46.3	72	70.4	46.4	49.9	14.9	15.9
Acute Respiratory Diseases	1.9	4.2	2.7	2.2	2.7	2.5	1.4	1.7
Other	.5	1.7	1.2	1.9	10.4	8.8	71.6	62.3
Fever (unknown underlying dis.)	3.4	8	8.4	8.3	19.5	18.7	9.1	11.4
Apoplexia, Violent, Accidental Deaths	.1	.2	1.1	2.6	3.8	17.1	1.6	6.5
Childbirth	0	0	0	0	14.7	0	0	0

B Data and Descriptive Statistics

B.1 *Zemstvo Medical Collection* Data Example

Figure B.1.1: Zemstvo District Medical Time-Series, example table

(a) Zemstvo District Medical Time-Series, example table, original

Екатеринославский уездъ.	1866	1867	1868	1869	1870	1871	1872	1873	1874	1875	1876	1877	1878	1879	1880	1881	1882	1883	1884	1885	1886	1887	1888
Число врачебныхъ участковъ съ мѣстожителемъ, врачей въ работѣ ихъ	—	—	2	2	2	2	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	6
Число мелкихъ лечебныхъ заведений съ живущимъ при нихъ врачомъ:																			1	1	1	1	2
а) въ крошечной (лабугатори) въ селѣхъ	—	—	1	2	2	2	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
б) безъ крошечной (лабугатори) въ селѣхъ	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Число фельдшерскихъ пунктовъ	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Число врачебныхъ пунктовъ (безъ постоянного персонала)	—	—	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18
Число всѣхъ врачей, состоящихъ на службѣ земствъ:	—	—	2	2	2	2	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	6
Изъ коихъ:																							
а) Слѣдующихъ поспѣло земской службѣ	—	—	2	2	2	2	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	6
б) Слѣдующихъ въ сѣмѣхъ	—	—	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
в) Женщинъ-врачей	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Число:																							
а) Фельдшеровъ	—	—	2	2	2	2	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	6
б) Акушерокъ	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
в) Осмотрѣвающихъ (подлѣжныхъ)	—	—	17	17	17	17	21	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20

Source: [Zhbankov \[1891\]](#)

(b) Zemstvo District Medical Time-Series, example table, translated

Ekatrinoslav District	1866	1867	1868	1869	1870	1871	1872	1873	1874	1875	1876	1877	1878	1879	1880	1881	1882	1883	1884	1885	1886	1887	1888
Number of medical districts with doctors' residences in their area	—	—	2	2	2	2	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	6
Number of medical clinics with doctors living nearby:																			1	1	1	1	2
(1) With beds	—	—	1	2	2	2	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
(2) Without beds	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Number of nurse stations	—	—	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18
Number of medical facilities (without regular personnel)	—	—	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18
Number of all doctors working for Zemstvo	—	—	2	2	2	2	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	6
Including:																							
(1) employed only by Zemstvo	—	—	2	2	2	2	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	6
(2) living in their catchment area	—	—	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
(3) female doctors	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Number of:																							
(1) Nurses	—	—	2	2	2	2	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	6
(2) Midwives	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
(3) Smallpox Vaccinators	—	—	17	17	17	17	21	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20

B.2 *Medical Directories* Data Example

Figure B.2.1: Medical List, example table

(a) *Medical Directories* example page, 1890, original

Алфавитный списокъ женщинъ-врачей.
Ангелова (по мужу Рейнъ) Анна Кирилл. 51 ж.-вр. 79, впр. С. Сосновка, Моршан., Тамбов.
Андреева Ольга Як. 60 ж.-вр. 87, впр. город. учил. Москва.
Антусевичъ Викторія Іос. 58 ж.-вр. 88, Орл. зем. госп. Извѣстскій зав., Саранпу., Вытск.
Антошкина Доротея Ісаак. 51 ж.-вр. 77, впр. Ростовъ на Д.
Арнольдъ Евгѣнія Ант. 62 ж.-вр. 87, впр. Севастополь.
Аронова Вѣра Евг. 56 ж.-вр. 86, Зем. вр. С. Поповка, Алексин., Тулск.
Архангельская Ал-дра Гавр. 51 ж.-вр. 82, Зем. вр. С. Петровка, Вереиск., Москв.
Архангельская Вѣра Ив. 61 ж.-вр. 87, впр. Тихвинской богад. Бронница, Моск.

Source: [Ministry of Internal Affairs \[1890\]](#).

(b) *Medical Directories* example page, 1890, translated

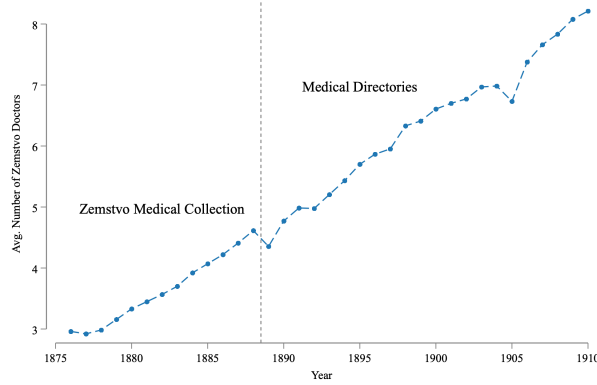
Alphabetical list of women-doctors
Angeleva (married name Rein) Anna Kirill. 51 wom.-dr. 79, priv. V. Sosnovka, Morshan., Tambov.
Andreeva Olga Yak. 60 wom.-dr. 87, Dr. city school, Moscow.
Antushevich Viktoria Ios. 58 wom.-dr. 88, Ord. Zem. hosp. Ishevskii factory, Sarapul., Vyatsk.
Aptekman Dorothea Isaak. 51 wom.-dr. 77, priv. Rostov on D.
Arnold Evgenia Ant. 62 wom.-dr. 87, priv. Sevastopol.
Aronova Vera Evg. 56 wom.-dr. 86, Zem. Dr. V. Popovka, Aleksin., Tulsk.
Arkhangelskaya, Al-dra Gavr. 51 wom.-dr. 82, Zem. Dr. V. Petrovka, Vereisk., Moskov.
Arkhangelskaya Vera Iv. 61 wom.-dr. 87, Dr. clin. Tikhvin char. Bronnitsa, Mosk.

Notes: In the blue frames, I highlight the observations I identify as Zemstvo female doctors. Notice that we observe all doctors' full names, birth and graduation years, and exact work locations. Sometimes, the district is not specified, so villages need to be geolocated.

The main challenge with the Medical Directories data is their textual format. First, I must identify all *Zemstvo* physicians while excluding other doctors employed by *Zemstvos*, such as sanitary physicians responsible for epidemic management and physicians working in large city-based *Zemstvo* hospitals. Precise classification is important given the research design, which focuses on replacement hires and requires accurate measurement of physician counts. I rely on structured text searches to identify relevant physicians. Specifically, the string combination “Зем.” (Zem.) and “вр.” (doctor) captures most *Zemstvo* physicians. I supplement this approach with additional inclusion and exclusion rules to address spelling variation and OCR errors, and I manually verify the resulting list to minimize false positives. While I cannot entirely rule out false negatives, I reviewed thousands of entries that were likely to include false negatives, where I have found none.

A second challenge is geolocating physicians. Most entries list the district of practice, allowing reliable text-based matching within provinces. Some entries, however (~10%), report only a village

Figure B.3.1: Average Number of Zemstvo Doctors



and province.⁵⁷ In these cases, I geolocate villages using historical geographic records. When multiple villages bear the same names within a province, I leave those observations unassigned. A small number of *Zemstvo* physicians served railroad networks and therefore operated across districts, making district assignment infeasible. In addition, some physicians, particularly in later years, do not report precise locations. As a result, between 1889 and 1910, I am unable to geolocate 0.6% (0.64% of male and 0.60% of female) of the 51,082 physician-years (8,430 unique physicians). Of course, despite my best efforts, given the vastness and heterogeneous quality of the data, some assignment errors may remain.

B.3 Continuity Across Physician Data Sources

Because the *Medical Directories* relied on physician self-reporting and required more extensive digitization and cleaning, I assess the continuity of the physician series around the transition from the *Zemstvo Medical Collection* to the *Medical Directories*. The data source changes in 1889, with the 1890 *Medical Directories* reflecting the workforce as of 1889. Figure B.3.1 plots the average number of *Zemstvo* doctors by year. Although there is a small decline in 1889, the levels recover immediately in 1890, suggesting a compression in *Zemstvo* staffing in 1889, rather than a systematic difference between the data sources, given that a systematic difference would affect the entire 1889-1910 period. Notice also that a similar dip takes place again in 1905, likely resulting from the Russo-Japanese War and the First Russian Revolution. Similarly, Figure B.5.1 shows no discontinuity in female physician entry around 1889.

⁵⁷I assign districts based on the presence of certain string combinations (e.g. if the location place contains "Vyaz" and is in Smolensk province, this row is assigned to Vyazemskii district). This improves efficiency vis-a-vis OCR errors and original typos, but may introduce errors if, in fact, the location cell is a village that also starts with "Vyaz" and is in a different district. The current implementation of this assignment ensures no such errors are made, after careful checks of all village-province-only entries.

B.4 Vital Statistics Data Example

Figure B.4.1: Vital Statistics Data, example table

(a) Deaths by District, Age Group and Gender, 1890, original

Губернии и уезды. Governments et districts.	I.	Возрасты. Ages.																			
		До 1 года. Au dessous d'un mois.					1-3 м.					3-6 м.					от 6 м. до 1 г.				
		м.	ж.	м.	ж.	м.	ж.	м.	ж.	м.	ж.	м.	ж.	м.	ж.	м.	ж.	м.	ж.	м.	ж.
		м.	ж.	м.	ж.	м.	ж.	м.	ж.	м.	ж.	м.	ж.	м.	ж.	м.	ж.	м.	ж.	м.	ж.
Архангельская	1. Архангельск. уезд.	72	64	51	63	62	39	83	68	16	56	23	31	18	38	17	11	31	21	11	10
	2. Кемский	16	10	10	9	28	15	30	28	18	21	9	9	6	8	9	5	5	6	4	5
	3. Ковской	15	9	10	9	17	13	15	8	9	10	10	6	5	3	3	6	8	5	5	4
	4. Мезенский	63	53	77	62	75	85	121	120	91	101	41	35	31	25	18	23	28	8	26	15
	5. Пинежский	71	53	50	36	49	36	75	68	46	62	20	27	25	17	9	10	17	15	8	11
	6. Пинежский	55	40	28	37	41	22	48	40	41	40	14	23	11	11	4	10	32	25	9	14
	7. Холмогорский	78	42	27	37	43	52	67	52	51	34	47	31	24	21	10	14	38	20	9	14
	8. Шенкурский	105	79	100	129	128	124	205	163	172	166	58	61	51	40	23	27	62	73	23	13
	Итого	470	344	320	370	441	291	642	520	470	501	226	244	170	176	87	101	230	233	55	120

(b) Deaths by District, Age Group and Gender, 1890, translated

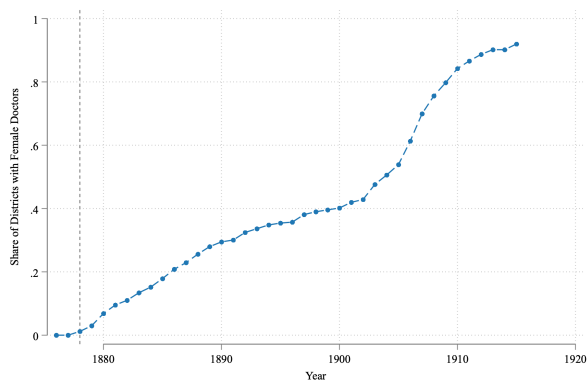
Губернии и уезды. Governments et districts.	I.	Возрасты. Ages.																			
		До 1 месяца. Au dessous d'un mois.					1-3 м.					3-6 м.					от 6 м. до 1 г.				
		м.	ж.	м.	ж.	м.	ж.	м.	ж.	м.	ж.	м.	ж.	м.	ж.	м.	ж.	м.	ж.	м.	ж.
		м.	ж.	м.	ж.	м.	ж.	м.	ж.	м.	ж.	м.	ж.	м.	ж.	м.	ж.	м.	ж.	м.	ж.
Arkhangelskaya	1. Arkhangelskii	72	64	51	63	62	39	83	68	16	56	23	31	18	38	17	11	31	21	11	10
	2. Kemskii	16	10	10	9	28	15	30	28	18	21	9	9	6	8	9	5	5	6	4	5
	3. Kovskii	15	9	10	9	17	13	15	8	9	10	10	6	5	3	3	6	8	5	5	4
	4. Mezenskii	63	53	77	62	75	85	121	120	91	101	41	35	31	25	18	23	28	8	26	15
	5. Pinezhskii	71	53	50	36	49	36	75	68	46	62	20	27	25	17	9	10	17	15	8	11
	6. Pinezhskii	55	40	28	37	41	22	48	40	41	40	14	23	11	11	4	10	32	25	9	14
	7. Kholmogorskii	78	42	27	37	43	52	67	52	51	34	47	31	24	21	10	14	38	20	9	14
	8. Shenkurskii	105	79	100	129	128	124	205	163	172	166	58	61	51	40	23	27	62	73	23	13
	Total	470	344	320	370	441	291	642	520	470	501	226	244	170	176	87	101	230	233	55	120

Source: Central Statistical Committee of the Ministry of Internal Affairs [1895].

B.5 Timing of Female Physician Entry

Figure B.5.1: Share of Districts That Had Hired a Female Physician by Year

(a) Full Sample



(b) Analysis Sample: Treated Districts Limited to Replacement Hires

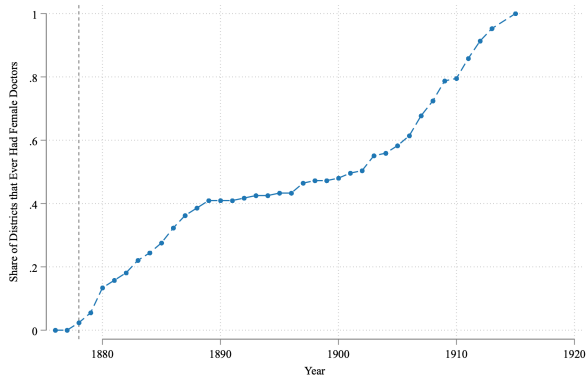


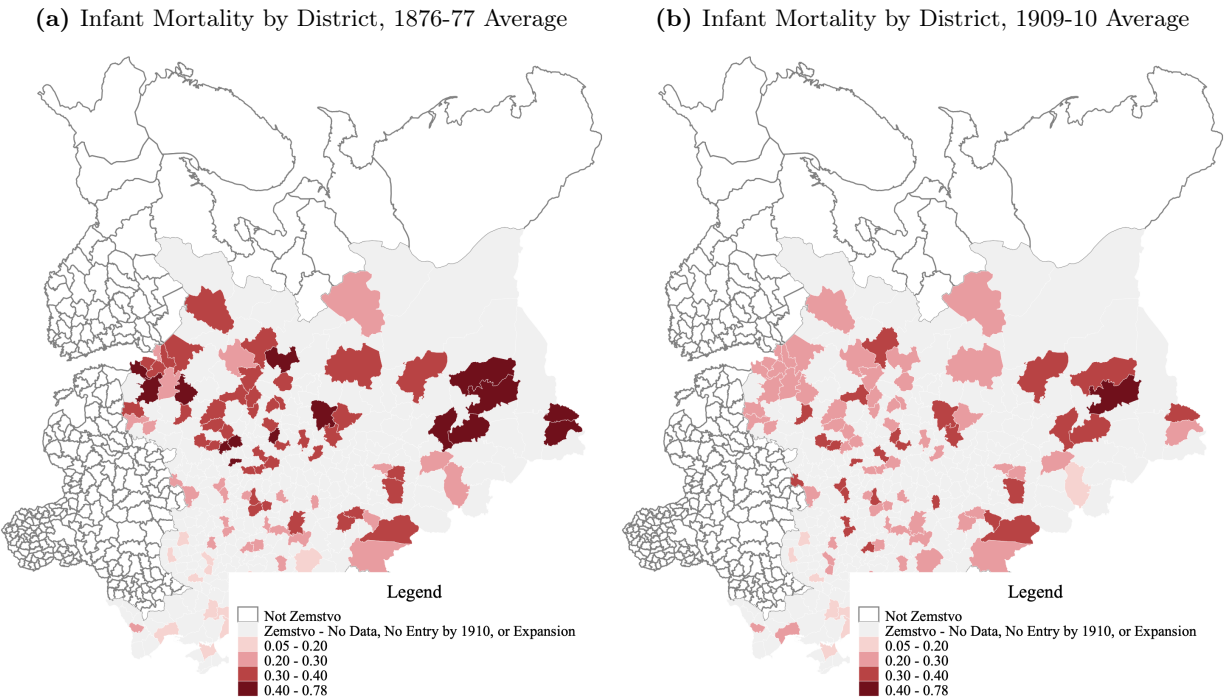
Table B.5.1: Balance on Baseline Levels, 1876

	1878-1888 Adopters			1889-1910 Adopters			1911-1915 Adopters			F-test p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Mean	SD	N	Mean	SD	N	Mean	SD	N	
Panel A: Physicians and Vital Statistics										
(per 1,000 pop (1896) unless stated otherwise)										
Doctors	.025	.021	49	.02	.011	52	.02	.012	6	0.21
Births	55	71	44	45	7.4	48	39	9	6	0.57
Infant Deaths (Age 0) per 1,000 Births	331	109	44	303	91	48	273	84	6	0.24
<i>Within-Infancy Deaths per 1,000 Births, Months:</i>										
0	80	30	44	77	26	48	70	26	6	0.70
1-2	83	41	44	69	28	48	61	28	6	0.09
3-5	78	28	44	72	27	48	62	23	6	0.30
6-12	89	25	44	85	24	48	79	19	6	0.48
<i>Number Male Deaths, Ages:</i>										
0	10	13	44	7.6	2.7	48	6.2	3.2	6	0.32
1-14	4.9	7.1	44	3.9	1.1	48	3.8	1.1	6	0.56
15-54	5.7	20	44	2.7	.77	48	2.3	.52	6	0.54
55+	3.3	5.5	44	2.5	.75	48	2.3	.96	6	0.56
All	24	45	44	17	4.1	48	15	4	6	0.47
<i>Number Female Deaths, Ages:</i>										
0	8.5	11	44	6.3	2.4	48	5.2	2.6	6	0.29
1-14	4.6	6.2	44	3.7	.96	48	4	1.1	6	0.55
15-54	4	9.1	44	2.8	.75	48	2.5	.68	6	0.63
55+	3.5	5.5	44	2.7	.95	48	2.5	1.2	6	0.53
All	21	31	44	16	3.7	48	14	3.5	6	0.47
Panel B: Healthcare Provision										
(per 1,000 pop (1896) unless stated otherwise)										
Midwives	.02	.021	46	.013	.0096	50	.015	.013	6	0.07
Nurse-Midwives	.0027	.012	47	.0024	.0067	49	0	0	6	0.80
Male Nurses	.068	.041	49	.068	.032	49	.069	.029	6	1.00
Female Nurses	.00064	.003	49	.0017	.01	50	0	0	6	0.72
Smallpox Vaccinators	.054	.058	31	.069	.071	32	.059	.038	3	0.62
Clinic Beds	.31	.25	46	.24	.17	47	.24	.16	6	0.24
Clinic Patient Death Rate, %	.058	.033	33	.073	.06	30	.057	.031	3	0.44
In-clinic Patients	750	2368	41	424	322	43	262	248	6	0.59
Share of Births in Clinics	0	0	15	0	0	9	0	.	1	.
Annual Medical Budget (Rubles, 1,000s)	.085	.075	39	.086	.039	34	.08	.02	5	0.98

This table shows baseline levels of observables by timing of the first female physician hire. Baseline levels are measured in 1876, when no female physicians had yet graduated from the Female Medical School. Units are per 1,000 population, as measured in the January, 1897 Census (representing population in 1896), unless noted otherwise. Columns 1–9 report, for each group, the mean, standard deviation, and number of districts; column 10 reports the p-value from the joint F-test of equality of group means. Please refer to the notes to Table 1 for data sources.

B.6 Trends in Infant Mortality Over Space

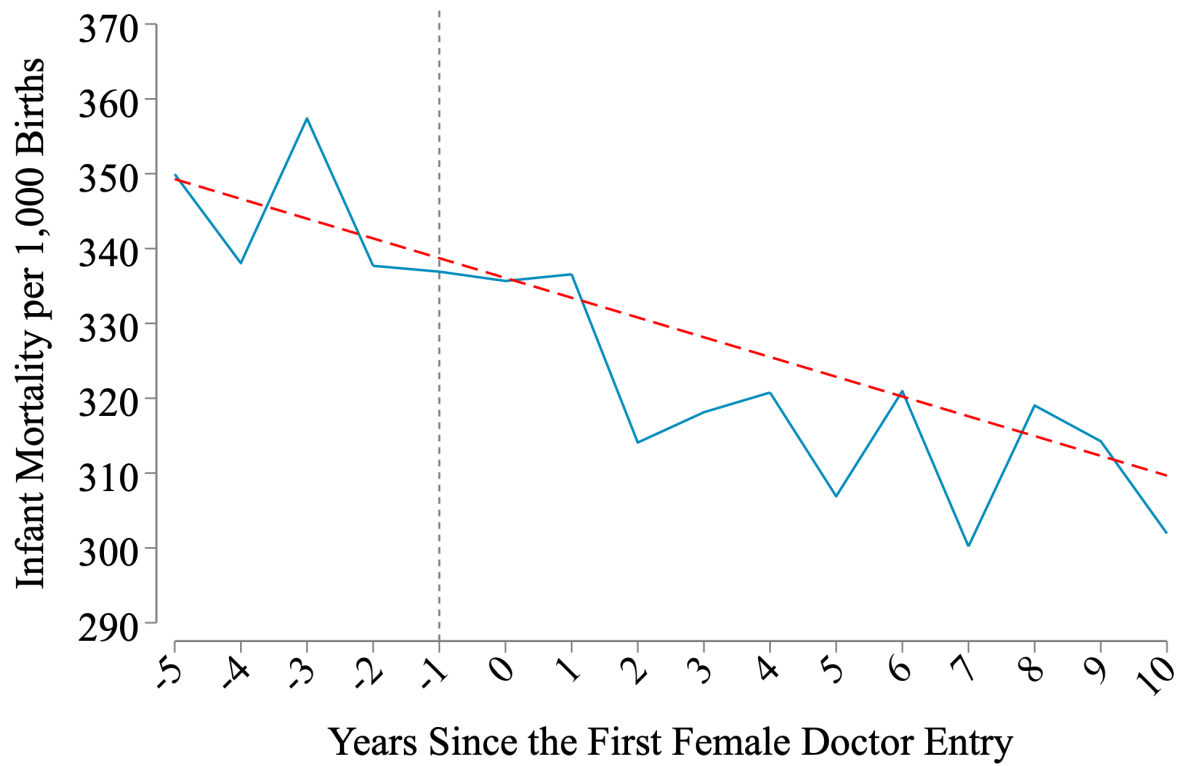
Figure B.6.1: Raw Trends in Infant Mortality



Notes: Source for historical shapefiles: [Kessler and Markevich \[2020\]](#)

B.7 Raw Data

Figure B.7.1: Infant Mortality in Treated Districts, Relative Time

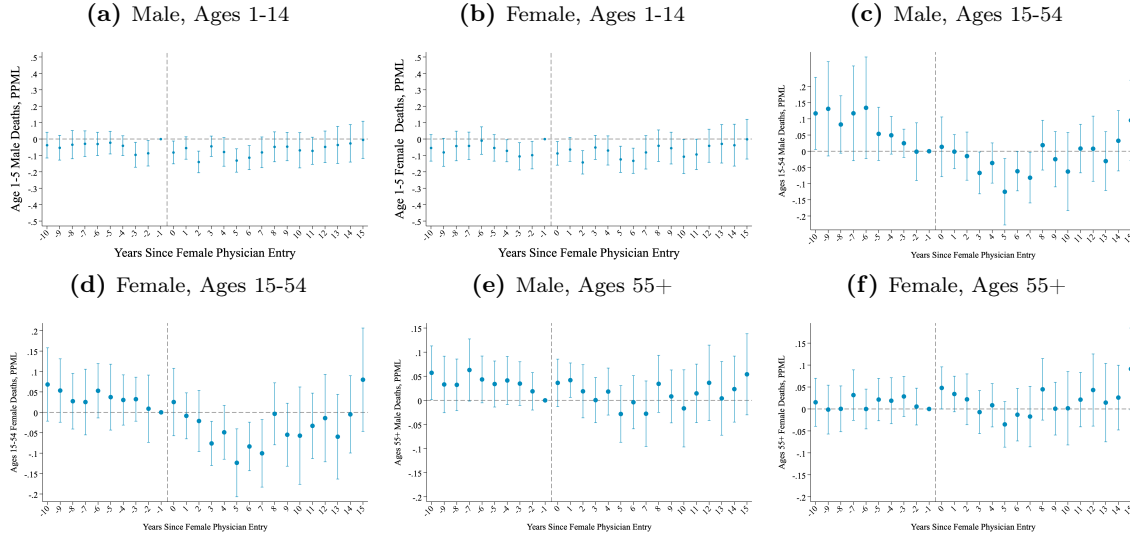


Notes: The sample is limited to 1881-1900 adopters (41 districts) to create a balanced panel. The red dashed line represents the trend extrapolated from the -5 to -1 period.

C Results

C.1 Mortality Event Studies

Figure C.1.1: Event Studies by Age Groups



Notes: Sample construction as described in Figure 5. Estimation follows equation 4.

C.2 Population Growth Rates

To examine how female physician entry affected population dynamics, I construct annual district-level natural population growth rates by combining comprehensive data on births and deaths with district population counts from the January 1897 Population Census. Natural population growth is defined using the standard population accounting identity:

$$\frac{P_{dt} - P_{dt-1}}{P_{dt-1}} = \frac{\text{Births}_{dt} - \text{Deaths}_{dt}}{P_{dt-1}},$$

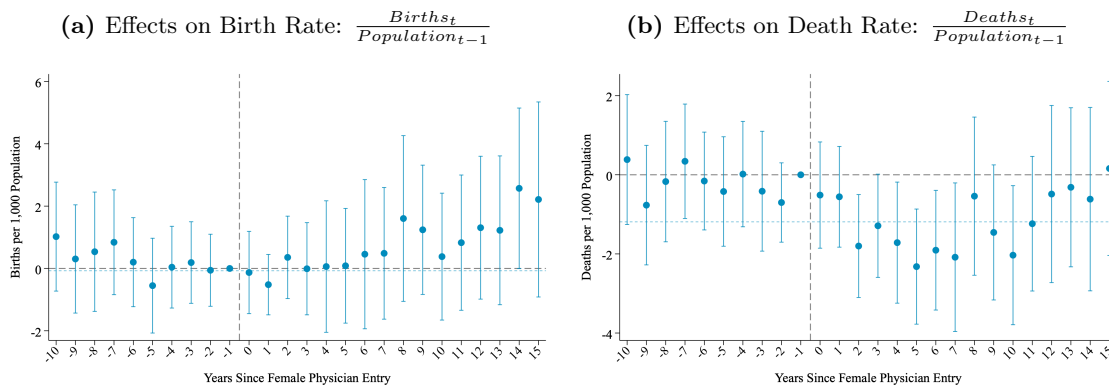
where P_{dt} denotes the population of district d in year t . This identity allows natural growth to be decomposed into birth and death rate components.

I estimate separate event-study specifications for the birth rate, $\frac{\text{Births}_{dt}}{P_{dt-1}}$, and the death rate, $\frac{\text{Deaths}_{dt}}{P_{dt-1}}$. Appendix Figure C.2.1 presents these results. The estimates show that the initial increase in natural population growth is driven by reductions in mortality. In later years following the entry of female physicians, birth rates also increase.

The suggestive increase in births is somewhat surprising in light of evidence from other settings where reductions in infant mortality have been associated with fertility decline (e.g., [Ager, Worm Hansen, and Sandholt Jensen, 2018](#)) and broader fertility transitions. One possible explanation is that female physician entry reduced not only infant mortality but also mortality among young adults, thereby

increasing the number of reproductive years within the population and the length of time spouses lived together. A second possibility is that fertility control was limited in rural Russia during this period. Although some knowledge of fertility control existed, evidence suggests that Russian peasant families rarely attempted to actively limit fertility [Malein, Matiashvili, and Beltrán Tapia, 2024]. In such settings, improvements in infant survival may have had only modest effects on fertility through biological channels such as breastfeeding-related changes in fecundity [Jayachandran and Kuziemko, 2011]. Third, the increase in births may represent an increase in *live* births, where stillbirths previously would not have been reported. Notably, this pattern diverges from the classic demographic replacement effect, in which reductions in infant mortality tend to reduce fertility as parents need fewer births to reach a target number of surviving children.

Figure C.2.1: Effects on Birth and Death Rates



Notes: Sample and estimation same as Figure 6.

C.3 Calculating Period Life Expectancy at Birth

Below, I describe the steps used in reconstructing annual period life expectancies at birth. The methodology is based on standard period life expectancy calculations in the demography literature [Preston, Heuveline, and Guillot, 2001], except for using the indirect standardization approach [Curtin and Klein, 1995], which is common in historical settings where annual population counts are not available, as is the case in this setting. Note that a more direct approach would have backward- and forward-imputed population counts by age group. However, because the time horizon is long and age groups between the population census and the deaths records are not exactly aligned, the measurement error associated with individual age groups aging in and out of the broader age groups is a large enough problem that some age groups reach negative counts the further back and forward we go. So I use the more conservative approach described below, which will somewhat underestimate the true impact of female physician entry on life expectancy.

Deaths data harmonization First, I put all death counts into a single set of age groups that are consistent every year and between deaths and Census population bins. Based on available bins, I create new groupings: Ages 0, 1-9, 10-19, 20-29, 30-39, 40-49, 50-59, and 60 plus. In the deaths data, the published age groupings are coarser than these bins in the 1876-1883 period, and more

detailed in the 1884-1910 period. For 1876–1883, I split the broader bins proportionally using sex-specific shares learned from the detailed years (so the early years inherit the later years’ observed age pattern by sex).⁵⁸ For 1884–1910, I collapse the detailed bins to our target bins; this yields, for each district and sex and year, a consistent vector of deaths across the same age groups.

Exposures (population at risk) We only observe population by age once, in the 1897 Census. I align those census counts to the same target bins I use for deaths.⁵⁹ These age-by-sex, district-specific populations serve as fixed anchors.

1896 standard age pattern. Using deaths and census exposures in 1896,⁶⁰ I construct a *sex-specific, common (sample-wide)* standard mortality schedule that serves as an anchor for all districts. For sex $s \in \{\text{Male, Female}\}$ and age bin x ,

$$M_x^{\text{std},1896,s} = \frac{D_{x,1896}^s}{P_{x,1896}^s},$$

where $D_{x,1896}^s$ is the total number of deaths in 1896 in age bin x (summed over all districts) and $P_{x,1896}^s$ is the corresponding 1896 census exposure. This schedule varies by sex but *not* by district; it is used solely as an anchor that will be scaled to obtain district–year–sex rates in subsequent steps.

Indirect standardization each year (SMR). For each district d , year t , and sex s , I compare observed deaths with the expected deaths expected, had the district experienced the 1896 standard age pattern applied to its own 1896 age structure. Let x index age bins.

$$O_{d,t}^s = \sum_x D_{d,t,x}^s$$

$$E_d^s = \sum_x P_{d,1896,x}^s M_x^{\text{std},1896,s}$$

$$\text{SMR}_{d,t}^s = \frac{O_{d,t}^s}{E_d^s}$$

The standardized mortality ratio $\text{SMR}_{d,t}^s$ is a single multiplier that adjusts the overall level of mortality in year t (for district d , sex s) while preserving the 1896 age pattern.

⁵⁸For example, the 1876-1883 years only include a 20-34 age bin. To calculate the number of deaths in the 20-29 age bin for each sex in this period, I first sum the deaths in age groups 20-24, 25-29, and 30-34 from 1884-1910 years (separately by sex) and then divide the sum of deaths in 20-24 and 25-29 age groups by the total sum. In this case, this share for women is 0.69. That is, 69% of deaths among women in the group 20-34 occur in ages 20-29. I then multiply deaths in age groups 20-34 in 1876-1883 by 0.69 to estimate the number of deaths that occurred in the 20-29 age group.

⁵⁹I split age 0-9 into age 0 (births, from vital statistics), and population aged 1-9 (observed population minus births).

⁶⁰The 1897 Census was taken in January, so those population counts more accurately reflect 1896 end-of-year population counts.

Annual age-specific rates and life tables. For each district–year–sex, we obtain annual age-specific death rates by scaling the 1896 standard schedule with the standardized mortality ratio (SMR):

$$M_{x,t} = \text{SMR}_{d,t}^s \cdot M_x^{\text{std},1896},$$

where x indexes age bins and t indexes years. From these $M_{x,t}$, I build abridged period life tables. For each closed interval of width n_x , I convert rates to probabilities using

$$q_{x,t} = \frac{n_x M_{x,t}}{1 + (n_x - a_x) M_{x,t}}, \quad \text{with} \quad a_x = \frac{n_x}{2} \quad (\text{including } a_0 = \frac{1}{2}).$$

Starting from a radix $l_{0,t} = 100,000$, survivorship proceeds via

$$d_{x,t} = l_{x,t} q_{x,t}, \quad l_{x+n_x,t} = l_{x,t} - d_{x,t}.$$

Person-years lived in closed intervals are

$$L_{x,t} = n_x l_{x+n_x,t} + a_x d_{x,t},$$

and for the open interval ω we use

$$L_{\omega,t} = \frac{l_{\omega,t}}{M_{\omega,t}}.$$

Total person-years above birth are $T_{0,t} = \sum_x L_{x,t}$, and life expectancy at birth is

$$e_{0,t} = \frac{T_{0,t}}{l_{0,t}}.$$

The life expectancy estimates rely on the assumption that the relative age pattern of mortality remained constant over time, with only the overall level changing across districts and years. If mortality improvements were concentrated in specific age groups (such as infants or young adults, as is the case in this paper) rather than uniform across all ages, this method may not fully capture the true variation in life expectancy changes. In addition, since the expected deaths denominator relies on a fixed 1896 population structure, unobserved changes in the true population at risk will affect the SMR. Crucially, because female physician entry increases natural population growth, treated districts mechanically experience higher absolute death counts over time due to a larger surviving population. This fixed-exposure method therefore artificially inflates the SMR in treated districts, further attenuating estimated treatment effects. Because both the fixed age-pattern assumption and the fixed population exposure bias the estimated gains downward, these results should be interpreted as conservative approximations that likely underestimate the true magnitude of life expectancy improvements.

D Robustness

D.1 Alternative Identification Strategies

Male versus Female Replacements. To approximate the comparison between districts hiring male versus female replacement physicians in the same year, I construct a stacked difference-in-differences design using individual physician records. These records allow me to identify replacement hires in the later part of the sample. I compare changes in mortality around female replacement hires with changes around male replacement hires occurring in the same calendar year.

Replacement events I identify an entering physician as one who is observed in a district in year t but was not observed in that district in year $t - 1$. Entry cannot be identified in the first year of the individual physician panel, which I therefore exclude from the set of potential replacement events. I classify a district-year as a replacement event when at least one physician enters and the total number of Zemstvo physicians does not increase relative to the immediately preceding calendar year.

A female replacement event involves at least one female entrant and no male entrants. A male replacement event involves at least one male entrant and no female entrants. I exclude district-years in which both male and female physicians enter. Unlike the primary specification, which focuses on the first female physician hire in each district, this exercise includes all qualifying female and male replacement events, in order to increase precision.

Stack construction I construct a separate stack for each replacement year c , comparing districts that experienced a female replacement with those that experienced a male replacement in the same year. I retain only stacks with at least one event of each type and include mortality observations from $c - 5$ through $c + 5$, where available. Districts can appear in multiple stacks if they experience replacement events in different years. Therefore, event windows can overlap, and the same district-year may appear in more than one stack.

Estimation d indexes districts, t calendar years, and s stacks, with c_s denoting the replacement year defining stack s . F_{ds} equals one if district d experiences a female replacement event in year c_s and zero if it experiences a male replacement event. In addition, let

$$P_{ts} = \mathbf{1}\{t \geq c_s\}.$$

For each age and sex group g , I estimate the following conditional mean specification using Poisson pseudo-maximum likelihood:

$$\mathbb{E}[Y_{dst}^g \mid F_{ds}, \alpha_{ds}^g, \lambda_{ts}^g] = \exp[\alpha_{ds}^g + \lambda_{ts}^g + \beta^g (F_{ds} \times P_{ts})], \quad (5)$$

where Y_{dst}^g is the number of deaths in group g , α_{ds}^g denotes district-by-stack fixed effects, and λ_{ts}^g

denotes calendar-year-by-stack fixed effects. District-by-stack fixed effects account for persistent differences across districts within each stack, while calendar-year-by-stack fixed effects account for annual shocks common to districts in the same stack. These fixed effects also absorb the main effects of F_{ds} and P_{ts} , respectively. I cluster standard errors at the district level to allow for dependence across years and across stacks involving the same district.

The coefficient β^g measures the change in mortality following female replacement hires relative to male replacement hires, expressed in log points. I report the estimates as percentage changes,

$$100 \left[\exp(\hat{\beta}^g) - 1 \right].$$

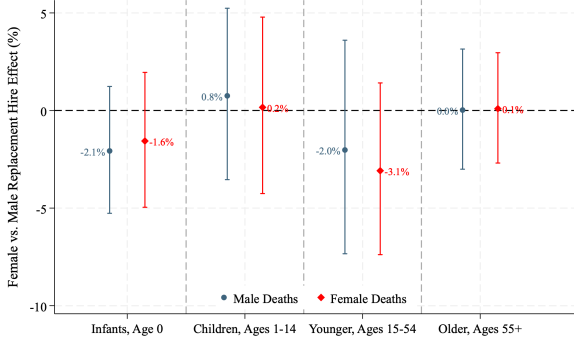
A negative estimate indicates a larger mortality decline following female replacement hires than following male replacement hires. A causal interpretation requires that, absent the difference in the gender of the replacement hires, the two groups would have experienced parallel mortality trends within each stack.

Comparison with the primary specification. To assess whether the choice of comparison strategy materially affects the results, I also re-estimate the primary specification using the female-entry cohorts over 1889-1910 and observation years 1885-1910. This comparison aligns district coverage, female-entry cohorts, and the observation period.

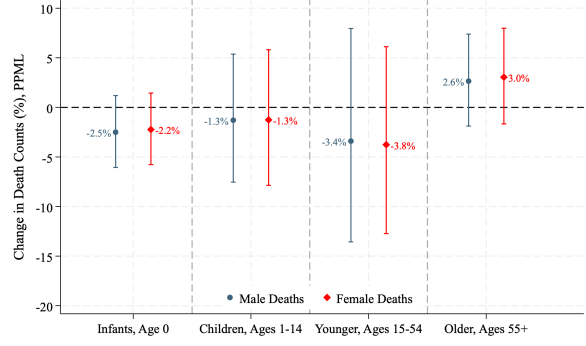
Figure D.1.1 reports the stacked replacement estimates alongside the estimates from the restricted primary specification. The two approaches produce similar negative point estimates for infants and adults ages 15–54. The estimates differ somewhat more for children ages 1–14 and adults ages 55 and older. The negative estimates for infants and younger adults are also consistent with the corresponding full-sample results. The estimates for children diverge from the full-sample results, consistent with the more cautious interpretation with those results, also stemming from the negative observed pre-trends. The results for older adults are consistent with the main specification, in that the coefficients are never negative.

Male versus Female Expansions As complementary evidence, I use expansions in physician employment as an alternative source of variation. Expansions were substantially less frequent than physician turnover: the median district expanded its physician workforce six times over the sample period (interquartile range: 4–7). I define a focal expansion as an increase of exactly one physician between consecutive years. Restricting attention to one-physician expansions provides a direct comparison between expansions that add one male physician and expansions that introduce a district’s first female physician.

For female physicians, I define an expansion when the increase of one physician coincides with the district’s first female physician entry and the number of physicians rises by exactly one. Thus, the event corresponds to a district expanding from N to $N + 1$ physicians, with the marginal physician being the district’s first female doctor. Male focal events are years in which the district similarly



(a) Stacked female-versus-male replacements



(b) Primary specification, restricted sample

Figure D.1.1: Mortality Effects: Stacked Replacement Design and Primary Specification in a Comparable Sample

Notes: Panel (a) reports estimates from stacked PPML regressions comparing female and male replacement hires occurring in the same calendar year. Each stack includes available observations within five years before and after the focal replacement year. Replacements are defined by physician entry accompanied by a non-increasing total number of Zemstvo physicians; events with both male and female entrants are excluded. Regressions include district-by-stack and calendar-year-by-stack fixed effects. Panel (b) reports estimates from the primary specification restricted to the corresponding districts and female-entry cohorts over 1889-1910, and observation years 1885-1910. Outcomes are death counts by age and sex. Estimates are expressed as $100[\exp(\hat{\beta}) - 1]$. Bars indicate 95% confidence intervals, with standard errors clustered by district. Male and female markers refer to the sex of the deceased.

expands from N to $N + 1$ physicians while the number of female physicians remains unchanged. Male expansion events must occur before the district's first female physician entry, and districts that never hire a female physician are also eligible. Unlike first female entry, a district's first observed male expansion has no particular substantive significance. I therefore use all eligible one-physician male expansions.

Because a district may experience several expansions during the sample period, fixed event windows around each expansion would cause the same district-year to enter multiple event stacks and allow the post-period of one expansion to extend through subsequent increases in physician supply. Therefore, I construct mutually exclusive event windows. All increases in physician employment, including expansions of more than one physician, define boundaries between neighboring expansion episodes. For each one-physician expansion, I assign the years between adjacent expansions to the nearest expansion, splitting the interval at its midpoint, and cap each event window at ten years before and after the focal event. As an additional restriction, a male-expansion window ends when the district receives its first female physician, even when that entry occurs through replacement. This procedure ensures that an underlying district-year contributes to at most one expansion window and that male post-expansion periods are not subsequently exposed to female physician entry.

I then stack the resulting event windows and estimate

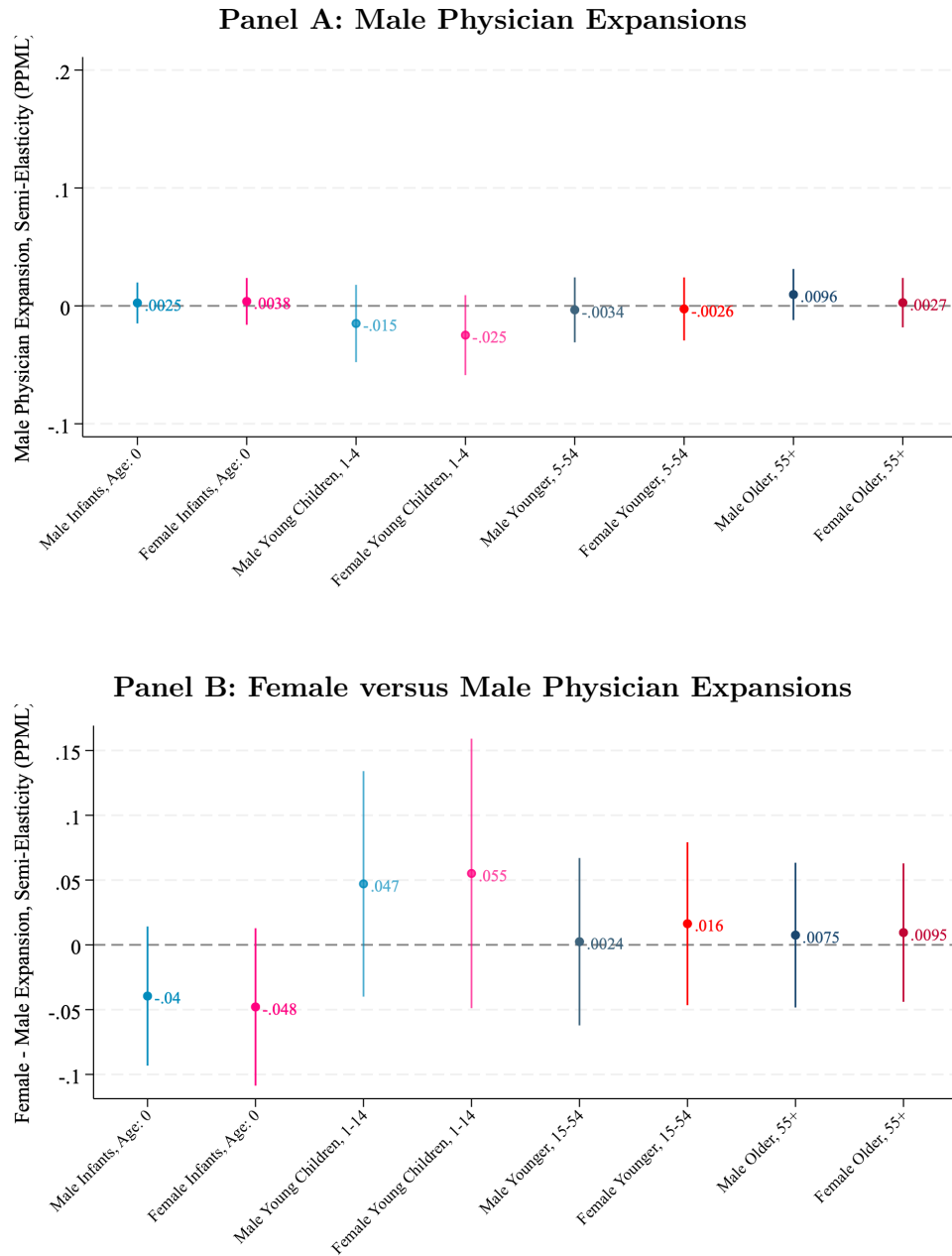
$$E[Y_{dst} | X_{dt}] = \exp(\alpha + \beta_M \text{Post}_{dst} + \beta_D (\text{Female}_s \times \text{Post}_{dst}) + \xi_s + \mu_t + \Psi X_{dt}),$$

where s indexes expansion-event stacks, Post_{dst} equals one beginning in the year of the focal expansion, and Female_s indicates that the focal expansion introduces the district's first female physician. ξ_s denotes stack fixed effects and μ_t denotes calendar-year fixed effects. As in the main analysis, I control for the total number of physicians in the district and cluster standard errors at the district level. The coefficient β_M measures the change in mortality following a male physician expansion, while β_D measures the differential change following a first-female expansion relative to a male expansion.

Figure D.1.2 reports these estimates across age and gender groups. Panel A reports the effects of male physician expansions. Because male expansions are more frequent than the first female expansions, these estimates are more precise. Except for mortality among children, the point estimates are close to zero across age groups.

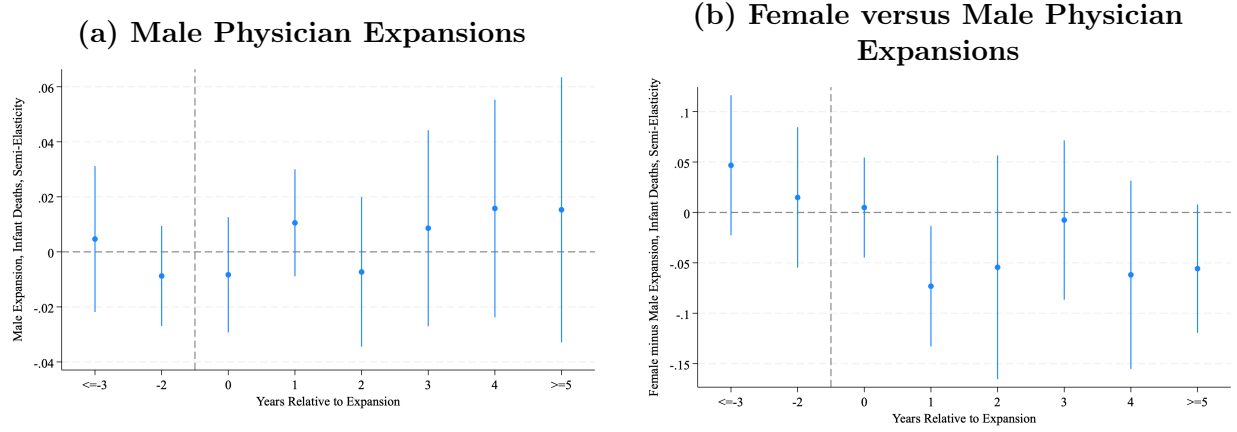
Figure D.1.3 examines the corresponding dynamics for infant deaths. Panel (a) shows little evidence of differential mortality trends around male physician expansions. The pre-expansion coefficients are close to zero, and there is no clear decline following entry. Panel (b) reports the differential coefficients for expansions introducing a district's first female physician relative to male physicians. The pre-expansion differentials are imprecisely estimated and include zero, while the post-expansion coefficients turn negative, with the largest declines arising one year after entry.

Figure D.1.2: Male and Female Physician Expansions: Effects by Age and Gender



Notes: The figure reports estimates from stacked PPML regressions around one-physician expansions occurring between 1878 and 1900. Female expansion events are years in which the district's first female physician enters and both total physician employment and female physician employment increase by exactly one. Male expansion events are one-physician expansions in which female physician employment remains unchanged and that occur before first female physician entry. Event windows are non-overlapping. All changes that increase physician employment define boundaries between expansion episodes, with intervening years assigned to the nearest expansion and each window capped at ten years before and after the entry event. Male-expansion windows additionally end at the district's first female physician entry. Panel A reports the estimated post-expansion effect for male physician expansions. Panel B reports the differential post-expansion effect of first-female relative to male physician expansions. All specifications include event-stack and calendar-year fixed effects and control for the total number of physicians. Standard errors are clustered at the district level.

Figure D.1.3: Event Studies Around Male and Female Physician Expansions: Infant Deaths



Notes: The figure reports event-study estimates from the stacked expansion design described in Appendix D.1. The outcome is the number of infant deaths. Female expansion events are years in which a district's first female physician enters and total physician employment increases by exactly one; male expansion events add exactly one male physician and occur before first female physician entry. Panel (a) reports the event-time coefficients around male physician expansions. Panel (b) reports the differential coefficients for first-female expansions relative to male expansions. The omitted period is the year immediately preceding the expansion. The first coefficient pools event times at or before -3 , and the final coefficient pools event times at or after $+5$. All specifications include event-stack and calendar-year fixed effects and control for the total number of physicians. Standard errors are clustered at the district level.

Interrupted Time Series As a complementary within-district exercise, I estimate an interrupted time-series (ITS) specification around the first female physician's entry. The ITS approach estimates whether mortality within a district departs from the trajectory predicted by that district's pre-entry trend after the first female physician arrives.

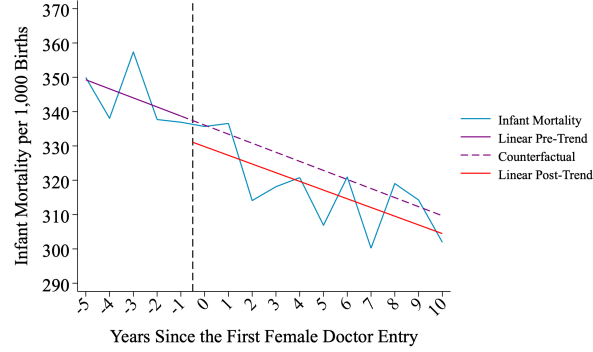
For infant mortality, I estimate

$$\log(Y_{dt}) = \delta_d + \theta_d \text{TimeTrend}_{dt} + \beta_1 \text{Post}_{dt} + \beta_2 (\text{TimeTrend}_{dt} \times \text{Post}_{dt}) + \gamma \text{NumDoctors}_{dt} + \epsilon_{dt}, \quad (6)$$

where δ_d denotes district fixed effects, $\text{TimeTrend}_{dt} = t - T^d$ is event time centered around the year of first female physician entry T^d , and θ_d is a district-specific linear trend coefficient. Post_{dt} equals one beginning in the year of female physician entry. The coefficient β_1 captures the discrete change in mortality at entry, while β_2 captures the change in the subsequent mortality trend relative to the district-specific pre-entry trajectory. I additionally control for the total number of physicians practicing in the district. The identifying assumption is that, absent female physician entry, mortality would have continued along the district-specific linear trajectory estimated from the surrounding time series.

For age-specific mortality counts, I estimate the analogous specification using Poisson pseudo-

Figure D.1.4: Interrupted Time Series Visual Example



Notes: The figure is limited to districts in which the first female physician entered as a replacement for a departing male physician and infant mortality is non-missing in every relative year from -5 through $+10$. Relative time is in relation to the district's first female physician entry. The solid pre-entry line reports the fitted linear trend over relative years -5 through -1 . The dashed line extrapolates this trend into the post-entry period and the post-entry line reports the fitted trend over relative years 0 through 10 .

maximum likelihood (PPML), with the same linear index, district fixed effects, district-specific linear trends, and physician-count control. Standard errors in all specifications are clustered at the district level. As in the primary analysis, I restrict the sample to districts in which the first female physician entered as a replacement for a departing male physician.

Figure D.1.4 illustrates the identifying variation for infant mortality. For transparency, the figure uses a balanced subset of districts with non-missing infant mortality in every relative year from five years before through ten years after first female physician entry. Because the mortality data span 1876-1910, this balanced sample is effectively restricted to entries occurring between 1881 and 1900. The solid pre-entry line shows the fitted linear trend prior to entry, while its dashed continuation illustrates the counterfactual mortality trajectory implied by that trend. The post-entry line shows the fitted mortality trajectory following female physician entry.

Table D.1.1 reports the ITS estimates across age groups. The results provide the clearest corroborating evidence at younger ages. Infant mortality falls by approximately 2.5 percent at entry, with a further negative post-entry trend. The level shift is statistically significant at the 5% level, while the trend change is suggestive but less precisely estimated. Mortality among children ages 1-14 displays a similar pattern for both sexes: deaths fall by approximately 5 percent at entry, and the post-entry mortality trend becomes significantly more negative. These results indicate that the mortality improvements at younger ages appear both as an immediate break at female physician entry and as a subsequent divergence from the district's pre-entry mortality trajectory.

The evidence for ages 15-54 is weaker. The estimated level shifts are negative for both men and women, but are imprecisely estimated. For women, I find no statistically significant change in the subsequent trend. For men, the post-entry trend is positive and statistically significant, offsetting the initially negative point estimate over time. One possible contributor is that these outcomes

are measured as death counts rather than mortality rates. Because population increased following female physician entry, growth in the population at risk may increase the death counts over time even when mortality risk does not proportionally increase. I therefore interpret the ITS estimates for this age group cautiously. Finally, I find little evidence of either a level or trend change in mortality at ages 55 and older. The ITS results provide within-district evidence of mortality reductions among infants and children, while providing less support for the young-adult effects found in the primary analysis.

Table D.1.1: Interrupted Time Series Results

	Infant Mortality	Children, Ages 1–14		Ages 15–54		Ages 55+	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All	Male	Female	Male	Female	Male	Female
Post-Female Physician Entry	-0.025** (0.012)	-0.050** (0.022)	-0.053** (0.024)	-0.031 (0.039)	-0.036 (0.029)	0.015 (0.018)	0.011 (0.018)
Linear Time Trend $\times$ Post-Female Physician Entry	-0.004* (0.002)	-0.012** (0.006)	-0.014** (0.006)	0.027** (0.011)	0.004 (0.006)	0.004 (0.004)	-0.005 (0.004)
N	3,471	3,472	3,366	3,471	3,470	3,469	3,367
Districts	101	101	101	101	101	101	101
District FE	✓	✓	✓	✓	✓	✓	✓
District-Specific Trend	✓	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓
Estimator	OLS	PPML	PPML	PPML	PPML	PPML	PPML

Notes: The sample is limited to districts in which the first female physician entered as a replacement for a departing male physician. Column (1) reports OLS estimates using log infant mortality as the dependent variable. Columns (2)–(7) report PPML estimates for mortality counts by age and sex. All specifications include district fixed effects, district-specific linear trends, and a control for the total number of physicians. Standard errors are clustered at the district level.

D.2 Shift-Share Instrument for First Female Physician Entry

To further address the possibility that districts hired female physicians alongside other contemporaneous investments in healthcare, I implement an instrumental-variables strategy based on a shift-share design. The instrument exploits variation in the regional supply of newly trained female physicians interacted with district-level physician vacancies.

The exposure component of the instrument captures the number of physician positions that became available in each district. Using the physician rosters described in Section 3, I observe the full set of doctors practicing in each district at the time of each medical directory. I define district-level vacancies as the number of physicians present in district j in year $t - 1$ who are no longer observed in that district in year t . These vacancies only count turnover, but not expansions or contractions.

The shift component captures fluctuations in the supply of newly trained female physicians in the *Zemstvo* provinces.⁶¹ Let G_t^f denote the number of female physicians who have graduated between years $t - 2$ and t . The instrument interacts the district-level share of vacancies with this aggregate

⁶¹I limit the sample of new graduates working in the studied *Zemstvo* to capture the relevant labor market. Alternatively, I could use the entire sample of all national graduates, but that pool is likely to be less relevant, as people from border territories and far east were more likely to represent ethnic minorities who would prefer and be able to effectively practice there. Using the full national supply yields near-identical results.

supply measure:

$$Z_{jt} = V_{j,t-1} \times G_t^f, \quad (7)$$

where $V_{j,t-1}$ is the share of physician vacancies per district roster in district j between years $t - 1$ and t .

Intuitively, when larger cohorts of female physicians enter the profession, districts with more vacancies are more likely to fill those positions with women. Districts with few or no vacancies are less impacted by fluctuations in the regional supply of female physicians. Therefore, the interaction generates differential variation across districts in the likelihood that a vacancy is filled by a female physician.

The shift-share instrument is used to predict the timing of first female physician entry, and thereby generates variation in whether a district is observed in the post-entry period. Let T_j denote the first year in which district j hires a female physician. I define the treatment indicator as

$$Treat_{jt} \equiv \mathbf{1}\{t \geq T_j\},$$

which equals one in all years at and after first female physician entry and zero otherwise.

I estimate the following first-stage specification:

$$Treat_{jt} = \pi Z_{jt} + \alpha_j + \lambda_t + X_{jt}\Gamma + u_{jt}, \quad (8)$$

where $Z_{jt} = V_{j,t-1} \times G_t^f$ is the shift-share instrument defined above, α_j are district fixed effects, and λ_t are year fixed effects. The vector X_{jt} includes the number of all doctors. Standard errors are clustered at the district level.

The corresponding second-stage specification is

$$Y_{jt} = \beta \widehat{Treat}_{jt} + \alpha_j + \lambda_t + X_{jt}\Gamma + \varepsilon_{jt}, \quad (9)$$

where Y_{jt} denotes the mortality outcome in district j and year t , and $\widehat{Treat}_{jt}$ is the predicted post-entry treatment status from the first stage. This compares outcomes before and after first female physician entry, using the component of treatment variation induced by regional fluctuations in the supply of newly trained female physicians interacted with district-level vacancies.

The shift-share identification strategy relies on the interaction between local district vacancies and fluctuations in the supply of newly graduated female physicians in *Zemstvo* provinces. The iden-

tifying assumption is that, conditional on district and year fixed effects and the total number of practicing physicians, this interaction affects district-level mortality only through its impact on the likelihood that district vacancies are filled by a female physician. The shift component reflects regional supply conditions in the medical profession rather than local healthcare demand or district-level investments. A major source of variation in the supply of female physicians is the reopening of the female medical school in 1897, which had been closed since 1887. Because medical training lasted five years, the first cohort from the reopened school graduated in 1903, leading to a large increase in the number of newly trained female physicians entering the profession.⁶² Importantly, the number of new graduates in a given year reflects the size of the matriculating cohort several years earlier and, therefore, is unlikely to respond to contemporaneous mortality shocks. To ensure the local vacancy component does not violate the exclusion restriction, I control for total physician employment, reducing concern that the instrument operates through contemporaneous changes in medical capacity.

One potential concern is that the increase in female physician graduates after 1903 coincided with the Russo–Japanese War and the revolutionary unrest of 1904–05. These events increased mortality among young adults⁶³. For this reason, I focus the IV analysis on infant mortality, which is unlikely to be directly affected by wartime and revolutionary mortality shocks.

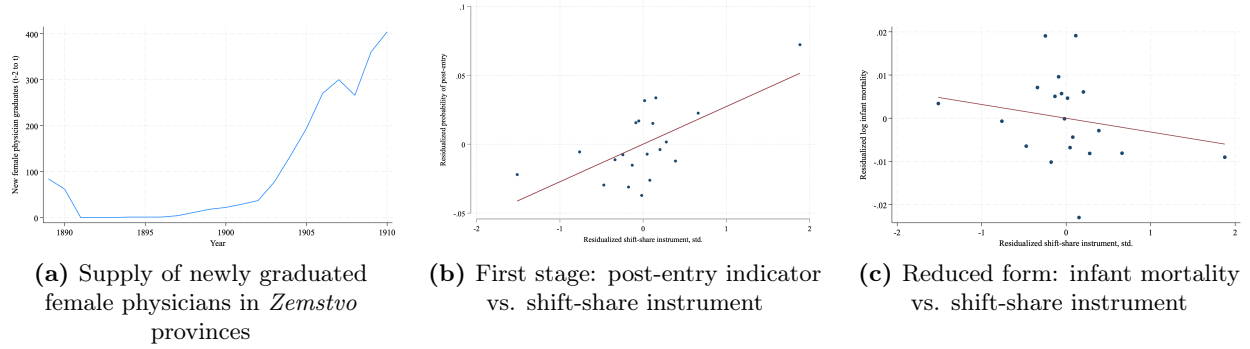
The IV estimate identifies the effect of being in the post–female physician entry period for districts where timing of entry responds to regional fluctuations in the supply of newly trained female physicians interacted with local vacancies. The estimate can be interpreted as a local average treatment effect for districts that hire female physicians when larger cohorts of women enter medicine.

Figure D.2.1 illustrates the variation underlying the shift-share instrumental-variables strategy and reports the corresponding 2SLS estimate. Panel (a) plots the supply of newly graduated female physicians in *Zemstvo* provinces, defined as graduates within three years of completing training. The supply of female physicians increases sharply after 1903, when the female medical school resumed graduating cohorts. Panel (b) shows the first-stage relationship between the shift-share instrument and the post-entry treatment indicator. Districts with higher exposure to the regional supply shock - i.e, those with a larger share of physician vacancies - are more likely to enter the post–female physician period when cohorts of female graduates are larger. Panel (c) presents the reduced-form relationship between the instrument and infant mortality. The corresponding 2SLS estimate (-0.12 [SE 0.17]) is negative and generally similar in magnitude to the baseline difference-in-differences estimates, though less precise.

⁶²A small number of female physicians entered the profession during the intervening years, likely graduating from European universities.

⁶³Deaths among both young men and women increase dramatically in this period.

Figure D.2.1: Shift-share instrument: regional female graduate supply, first stage, and reduced form

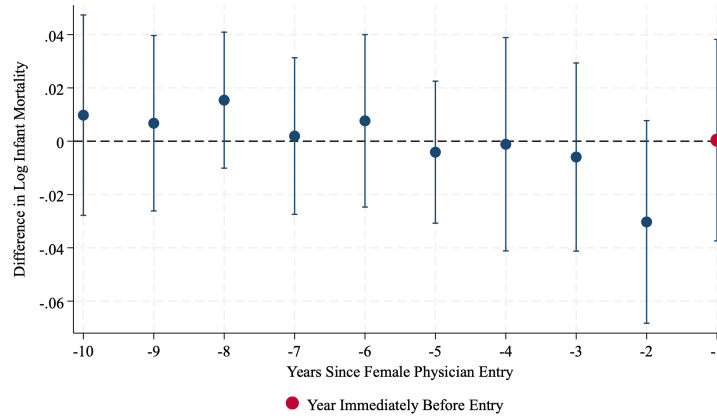


Notes: Panel (a) plots the supply of newly graduated female physicians in *Zemstvo* provinces (defined as graduates within three years of completing training). Panel (b) plots binned averages of the residualized post-entry treatment indicator against the residualized shift-share instrument, after partialling out district and year fixed effects and controls. Panel (c) shows the reduced-form relationship between log infant mortality and the shift-share instrument, both residualized with respect to the controls and fixed effects. The fitted lines come from OLS regressions using these residualized variables. The corresponding 2SLS estimate is $\hat{\beta} = -0.12$ (SE = 0.17), with first-stage $F = 9.8$.

D.3 Difference-in-Differences Design

Omitted Period

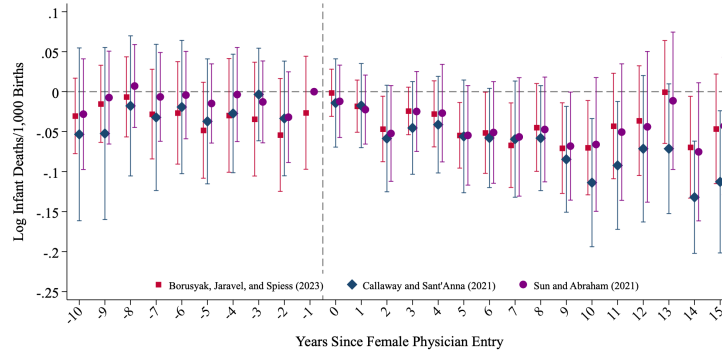
Figure D.3.1: Pre-Treatment Infant Mortality Around Female Physician Entry



Notes: The figure plots pre-treatment event-study coefficients for log infant mortality, re-centered so that the average of the pre-treatment event-time coefficients through $t = -2$ equals zero. Therefore, the estimate at $t = -1$ measures mortality in the year immediately preceding female physician entry relative to the other pre-treatment periods. The $t \leq -10$ point pools all earlier event times. Vertical bars show 95% confidence intervals. Standard errors are clustered at the district level.

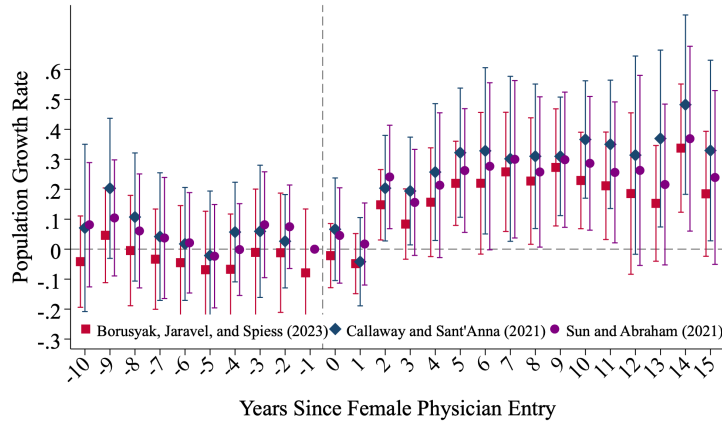
Alternative Two-Way Fixed Effects Estimators

Figure D.3.2: Event-Study: First Female Physicians and Infant Mortality. Alternative Staggered Difference-in-Differences Estimators



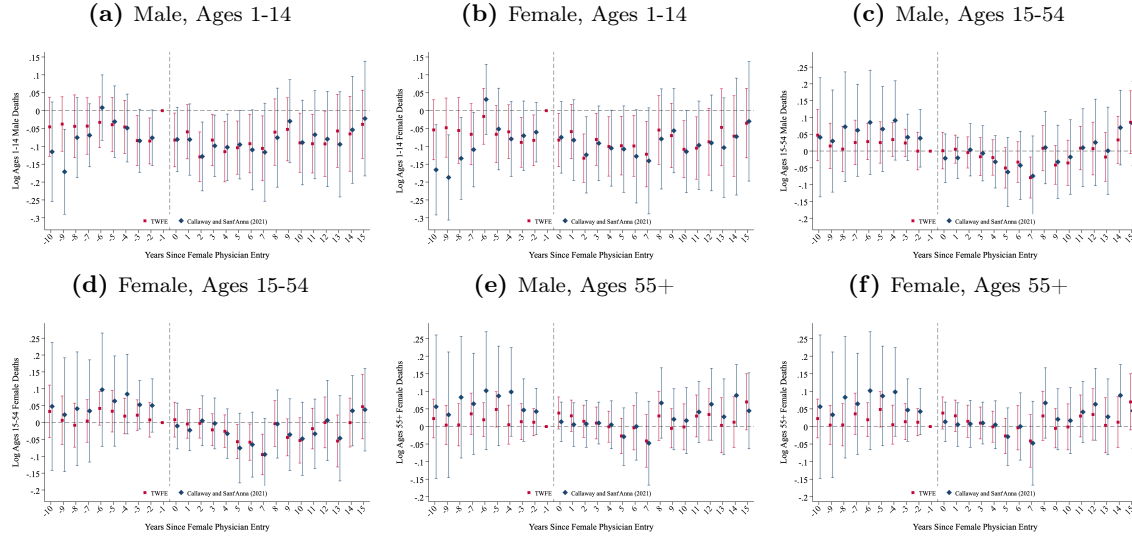
Notes: Sample construction follows Figure 3. The figure compares event-study estimates using the imputation estimator of Borusyak, Jaravel, and Spiess [2024], the estimator of Callaway and Sant'Anna [2021], and the interaction-weighted estimator of Sun and Abraham [2021]. All three estimators allow for heterogeneous treatment effects under staggered adoption. I use entry dates observed through 1915 to determine treatment timing, so districts where female physicians first entered between 1911 and 1915 serve as future-treated controls.

Figure D.3.3: Event-Study: First Female Physicians and Population Growth. Alternative Staggered Difference-in-Differences Estimators



Notes: Sample and outcome construction follow Figure 6. The figure compares event-study estimates using the imputation estimator of Borusyak, Jaravel, and Spiess [2024], the estimator of Callaway and Sant'Anna [2021], and the interaction-weighted estimator of Sun and Abraham [2021]. All three estimators account for heterogeneous treatment effects under staggered adoption. First female physician entry is treated as an absorbing state. Treatment timing is defined using entry dates observed through 1915. Districts where first female physician entry occurs between 1911 and 1915 remain untreated throughout the 1876-1910 outcome panel.

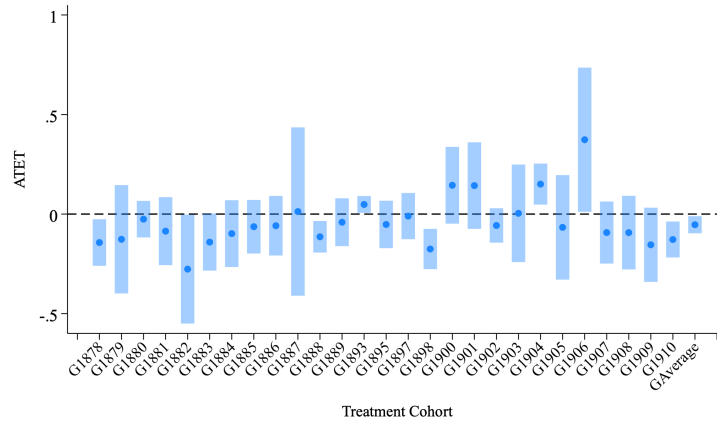
Figure D.3.4: Event Studies: Age Groups, Robustness Checks with Logged Outcomes



Notes: Sample construction as described in Figure 5.

Cohort-Specific Effects

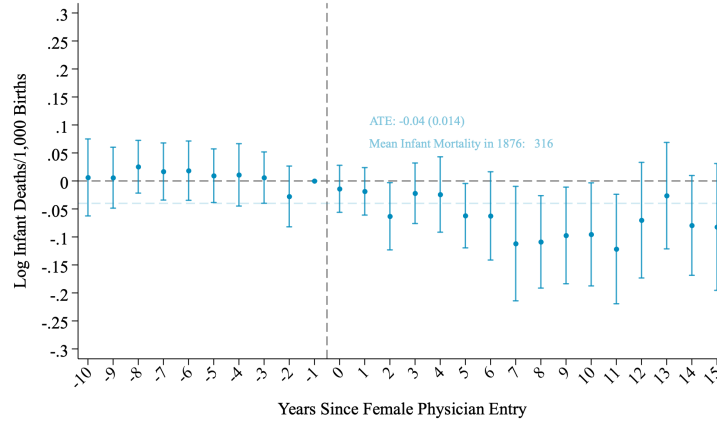
Figure D.3.5: ATET by Treatment Cohort



Notes: Sample construction and estimation follow Figure 3. Estimates based on Callaway and Sant'Anna [2021]

Restricting Post-Entry Years to Contemporaneous Female Physician Presence As a robustness check, I depart from the baseline absorbing-treatment definition and exclude post-entry district-years in which no female physician is present.

Figure D.3.6: Restricting Post-Entry Years to Contemporaneous Female Physician Presence



Notes: Sample and estimation construction same as in Figure 3 except I exclude post-entry district-years in which no female physician is present.

Table D.3.1: Robustness to Weighting and Sample Restriction

	Log Infant Deaths/1,000 Births							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female Physician Entry	-0.038*** (0.012)	-0.027** (0.012)	-0.045* (0.023)	-0.031* (0.018)	-0.042*** (0.013)	-0.040*** (0.015)	-0.058** (0.024)	-0.027*** (0.007)
N	3,679	3,414	1,893	1,994	3,670	3,679	2,907	11,485
Districts	107	98	55	58	107	107	107	336
Mean Y_{1876}	-1.220	-1.220	-1.199	-1.253	-1.209	-1.220	-1.220	-1.225
Year FE	✓	✓ x 1876 mortality	✓	✓	✓	✓	✓	✓
District FE	✓	✓	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓	✓
Weights					1/Doctors	Births	Population	
Sample	Restricted	Restricted	Restricted; Early	Restricted; Late	Restricted	Restricted	Restricted	Full
R-squared	0.673	0.708	0.722	0.640	0.670	0.695	0.675	0.698

Notes: Column (1) replicates Column (6) of Table 4. Column (2) interacts 1876 mortality levels with fixed effects. Column (3) restricts the sample to pre-1889 and 1911-1915 adopters, while Column (4) restricts it to post-1889 and 1911-1915 adopters. Columns (5) - (7) add weights scaled by: the inverse of the number of doctors; annual births; 1897 Population Census population counts, respectively. Column (8) does not restrict the sample to replacement hires.

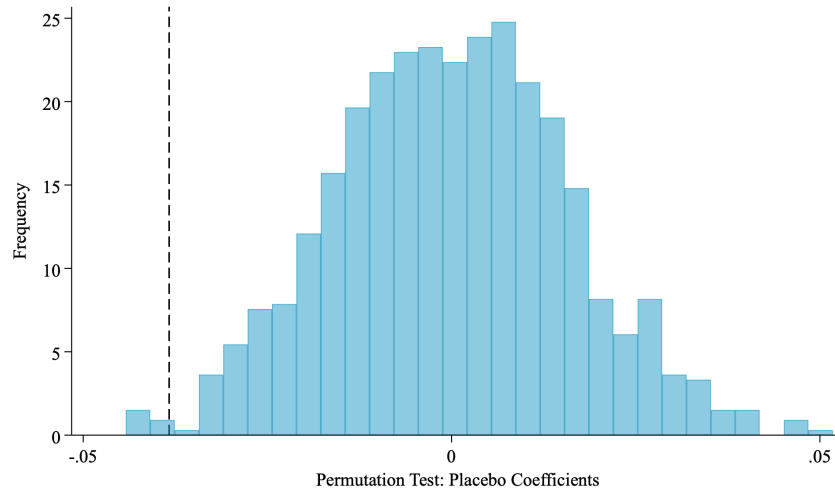
Robustness to Alternative Specifications

Permutation Tests. I conduct a permutation test to assess whether an effect as large as the estimated effect could plausibly arise from the allocation of female physician entry timing across districts. I randomly permute the observed first female physician entry years across the 107 districts in the analysis sample. This procedure preserves the observed distribution of treatment timing, including the six districts in which female physicians first entered between 1911 and 1915, that is, after the end of the 1876–1910 vital statistics panel.

For each of 1,000 permutations, I reconstruct the treatment indicator using the randomly assigned entry year and re-estimate equation 1. Appendix Figure D.3.7 plots the resulting distribution of placebo treatment coefficients with the coefficient estimated using the true allocation of treatment years. The two-sided permutation p -value is calculated as $(B_{\text{extreme}} + 1)/(B + 1)$, where B_{extreme} is the number of placebo coefficients at least as large in absolute value as the observed coefficient

and $B = 1,000$. The two-sided permutation p -value is 0.016. Therefore, the observed estimate lies in the far tail of the distribution generated under random reassignment of treatment timing.

Figure D.3.7: Permutation Test

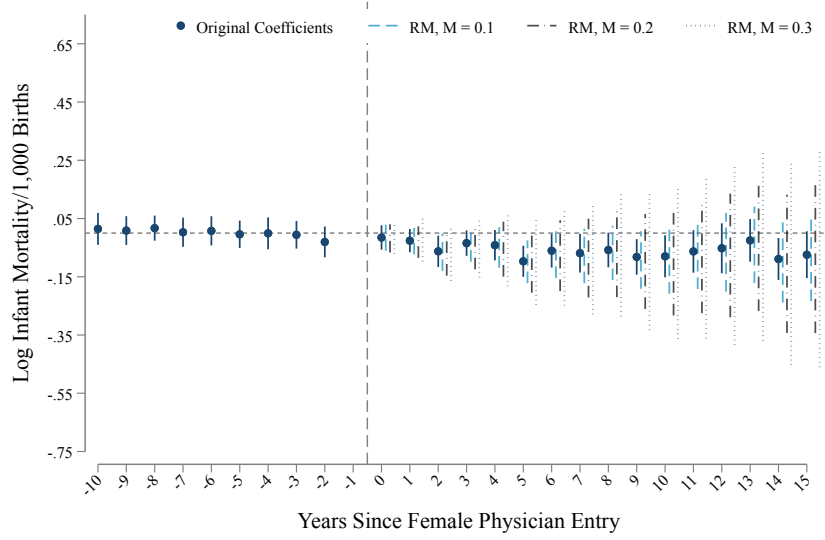


Notes: The figure plots coefficients from 1,000 permutations of the observed first-female-physician entry years across the 107 districts in the analysis sample. Each permutation preserves the empirical distribution of treatment timing. I then re-estimate Equation 1 using the baseline specification, treating female physician entry as an absorbing state beginning in the assigned year. The vertical dashed line indicates the coefficient estimated using the observed allocation of treatment years. The two-sided permutation p -value is 0.016.

HonestDiD sensitivity analysis I follow [Rambachan and Roth \[2023\]](#) and implement the relative-magnitudes HonestDiD procedure (`delta(rm)`) to assess the sensitivity of the infant mortality event-study coefficients to violations of the parallel trends assumption. The parameter M bounds the size of post-treatment deviations from parallel trends relative to the largest absolute deviation observed in the pre-period. For example, $M = 0.3$ allows post-treatment deviations up to 30% as large as the largest pre-period deviation. I apply this procedure to each post-treatment event time coefficient (equation 3), using pre-period coefficients for event times -10 to -2 and post-period coefficients for event times 0 to 15 . Appendix Figure D.3.8 plots the original 95% confidence intervals together with the HonestDiD intervals for $M = 0.1, 0.2$, and 0.3 .

The HonestDiD results show that the estimated short-run declines in infant mortality remain similar in magnitude for small relaxations of the parallel-trends assumption ($M = 0.1$), although the confidence intervals become wider. For moderate relaxations ($M \geq 0.2$), the honest confidence intervals widen and generally include zero in the post-treatment period. The combination of a limited number of districts, volatility of mortality rates, and the long event-study horizon likely explains this pattern.

Figure D.3.8: HonestDiD Sensitivity of Event-Study Estimates



Notes: Figure shows event-study coefficients for log infant mortality with 95% conventional confidence intervals, alongside HonestDiD confidence sets computed using the relative-magnitudes framework of [Rambachan and Roth \[2023\]](#). The parameter M bounds the size of post-treatment deviations from parallel trends relative to the largest absolute deviation observed in the pre-period. Results are shown for $M = 0.1, 0.2$, and 0.3 . Figure produced from the specification in equation 3.

Conley Standard Errors Accounting for Spatial Correlation To assess the robustness of the main results to spatial and serial correlation in unobservables, this appendix reports standard errors based on [Conley \[1999\]](#). Health outcomes and medical practices may be correlated across neighboring districts and over time, which is not captured by the clustered standard errors used in the paper. Following [Conley \[1999\]](#), I allow regression residuals to be correlated across districts within a specified geographic radius and across nearby years, using a Bartlett kernel to down-weight correlations with distance and time.

Table [D.3.2](#) reports estimates using alternative spatial distance cutoffs and temporal correlation windows. Columns (2)–(4) allow for spatial correlation within 200, 500, and 1000 km, respectively, and serial correlation up to one year. Columns (5)–(7) extend the temporal window to two years. The point estimates remain statistically significant across the various specifications.

Table D.3.2: Conley Standard Errors

	Baseline (1) Clustered	(2) 200 km	Lag = 1 (3) 500 km	(4) 1000 km	(5) 200 km	Lag = 2 (6) 500 km	(7) 1000 km
Female Physician Entry	-0.0393*** (0.0123)	-0.0393*** (0.0114)	-0.0393*** (0.0118)	-0.0393*** (0.0117)	-0.0393*** (0.0119)	-0.0393*** (0.0123)	-0.0393*** (0.0122)
<i>N</i>	3646	3646	3646	3646	3646	3646	3646

Notes: Each column reports the estimated effect of first female physician entry on log infant deaths per 1,000 births. Column (1) reports two-way fixed-effects estimates with standard errors clustered at the province-district level. Columns (2)–(4) report [Conley \[1999\]](#) standard errors allowing for spatial correlation within 200, 500, and 1000 km, respectively, and serial correlation up to one year. Columns (5)–(7) report Conley standard errors allowing for serial correlation up to two years. All specifications include province-district and year fixed effects. Spatial standard errors are computed using district centroid latitude and longitude with a Bartlett kernel. The estimation sample is held fixed across columns. One district included in the main estimation (e.g., Table 4) did not have centroid data and so is excluded from this sample. The data on district centroids come from [Natkhov and Vasilenok \[2023\]](#). The standard errors are calculated using the program from [Hsiang \[2010\]](#).

D.4 Robustness Related to Training and Entry Timing

Training Recency and Characteristics of Entering and Departing Physicians In Table D.4.1, I draw on the linked physician panel to identify the male physicians who departed from each replacement district in the year prior to first female physician entry. I define the replacement year as the year following the departing physician’s final observed year in the district and merge these records to replacement events by province, district, and first female physician entry year. When multiple male physicians depart in the same replacement year, I average their characteristics. I calculate each departing physician’s age at exit and graduation year and merge the graduation year of the entering female physician from the Medical Directories. Column (1) uses the difference between the entering female and departing male physicians’ graduation years, while column (2) enters their graduation years separately. Column (3) examines the age of the departing male physician.

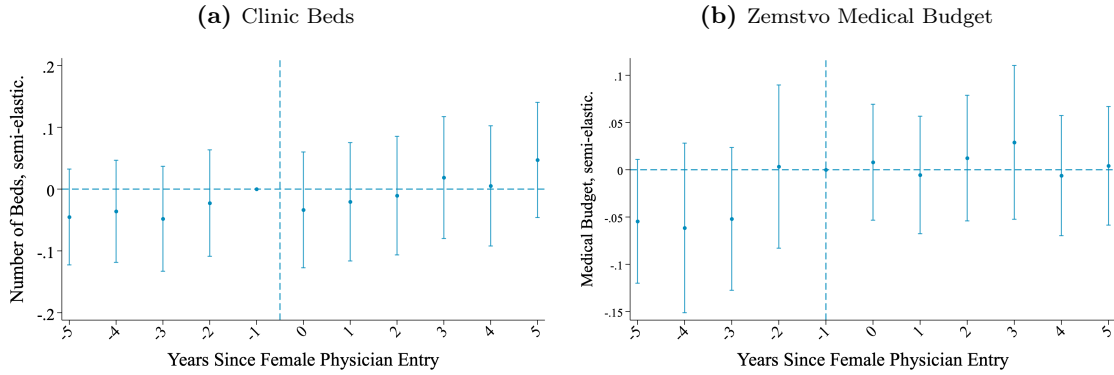
Table D.4.1: Training Recency and Characteristics of Entering and Departing Physicians

	Log Infant Mortality		
	(1)	(2)	(3)
Female Physician Entry	-0.030 (0.023)	-0.021 (0.022)	-0.025 (0.020)
Female Physician Entry $\times \Delta$ Graduation Year (F – M)	0.001 (0.002)		
Female Physician Entry $\times$ Entering Female MD Year		0.001 (0.002)	
Female Physician Entry $\times$ Departing Male MD Year		0.000 (0.003)	
Female Physician Entry $\times$ Age of Departing Physician			-0.002 (0.003)
N	1,553	1,553	1,658
Districts	45	45	48
Mean in 1889	-1.152	-1.152	-1.179
Year FE	✓	✓	✓
District FE	✓	✓	✓
Controls	✓	✓	✓
R-squared	0.605	0.605	0.583

Notes: This table examines whether the effect of female physician entry on log infant mortality varies with the training recency of the entering female and departing male physicians or with characteristics of the departing male physician. The sample is restricted to replacement districts whose first female physician entry falls within the 1889–1910 period covered by the linked physician panel. Column (1) interacts female physician entry with the difference in graduation years between the entering female and departing male physician; positive values indicate that the female physician graduated more recently. Column (2) instead interacts female physician entry separately with the graduation years of the entering female and departing male physicians. These graduation years are mean-centered within the replacement-event sample. Column (3) interacts female physician entry with the departing physician’s age at exit. In districts where multiple male physicians departed in the same replacement year, their characteristics are averaged.

Contemporaneous Investments Figure D.4.1 plots event-study estimates for observable margins of healthcare provision to assess whether the timing of the first female physician hire coincided with other potentially mortality-reducing investments, such as expansions in clinic beds, or medical budget allocations. I find no evidence of contemporaneous changes in these variables at the time of female physician entry. The data are drawn from the *Zemstvo Medical Collection* and cover the 1876–1888 period.

Figure D.4.1: Robustness: Contemporaneous Medical Investments



Notes: Source: *Zemstvo Medical Collection*, 1876–1888. The sample follows the main analysis sample and is restricted to districts whose first female physician hire through 1915 was classified as a replacement hire. Because healthcare provision outcomes are observed only through 1888, districts whose first female physician entry occurs after 1888 remain untreated in this outcome period and contribute only pre-treatment observations. Treatment is defined as an absorbing state beginning with the first female physician entry. Standard errors are clustered at the district level.

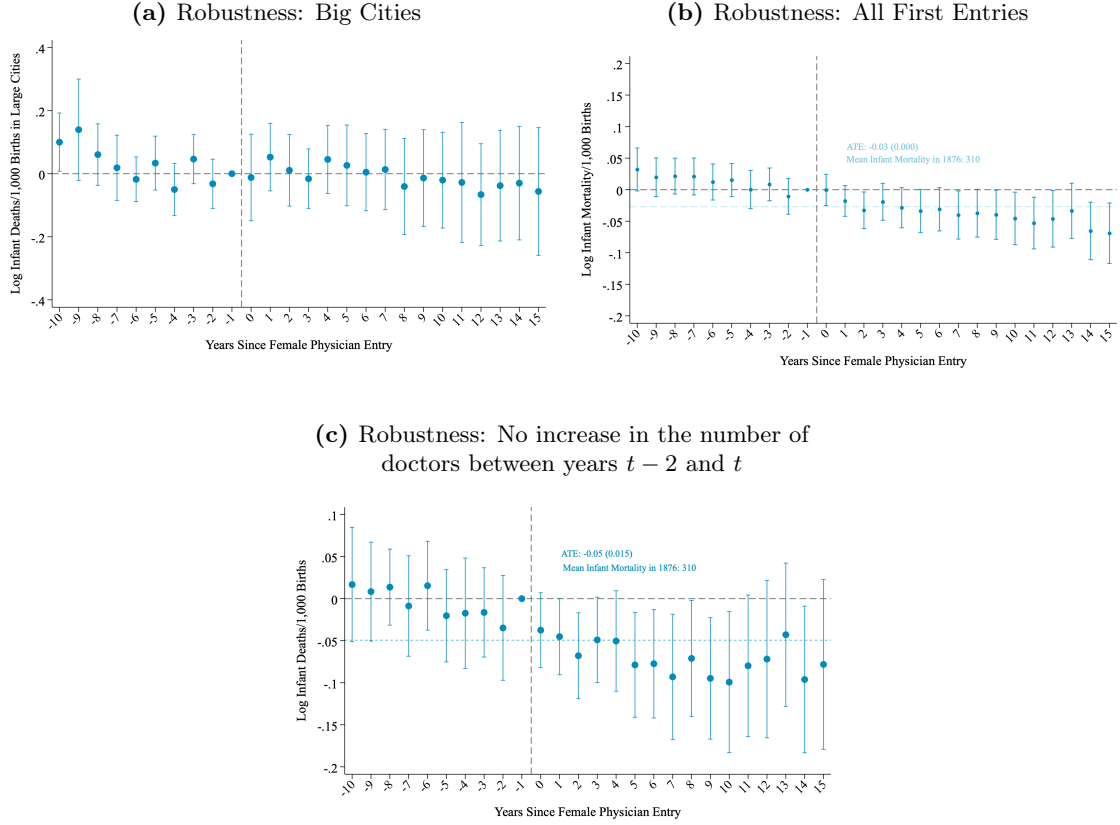
Big Cities *Zemstvo* medicine was designed to serve rural populations within districts. While smaller towns were typically included in these programs, large cities maintained separate medical provision systems. I therefore use mortality in large cities as a placebo outcome for broader local changes that may have coincided with female physician entry. Restricting this exercise to the replacement-hire analysis sample leaves only 12 cities, so I instead include all 41 major cities located in districts that eventually hired a female physician, regardless of whether the first female physician entered as a replacement or an expansion. Treatment timing is defined by the first female physician entry in the surrounding *Zemstvo* district. Figure D.4.2a presents the corresponding event-study estimates using birth and death data from *The Movement of Population in European Russia* volumes, originally digitized by Malein and Matiashvili [2024].

Full Sample The first female physicians were hired either when the *Zemstvo* boards were expanding care by adding catchment areas and doctor vacancies, or when they replaced a departing male doctor. The timing of the hiring decision was plausibly more exogenous in the latter scenario, so I limit the sample to these hires in the main analysis. Figure D.4.2b brings back all the first female entries, including those hired via medical care expansions. I find very similar results, although, as expected, the pre-period trends suggest that the parallel trends assumption may be violated in this full sample.

Entry Timing An important assumption underlying the main estimation sample is that the timing of first female physician entry is measured accurately. For 1889–1910, when I rely on the *Medical Directories*, some imprecision is possible. These directories were typically published in the first quarter of year t , and I assign each roster to year $t - 1$, assuming it reflects the prior year’s workforce. This could introduce error if, for example, a district hired its first female physician in January of year t , but I assign her entry to $t - 1$ and classify the hiring as a replacement based on physician counts between $t - 2$ and $t - 1$. To address this concern, Appendix Figure D.4.2c restricts

the sample to districts that experienced no net change in the number of physicians between $t - 2$ and t . The estimated reductions in infant mortality remain very similar to the main specification, with slightly larger average treatment effects.

Figure D.4.2: Additional Robustness Checks

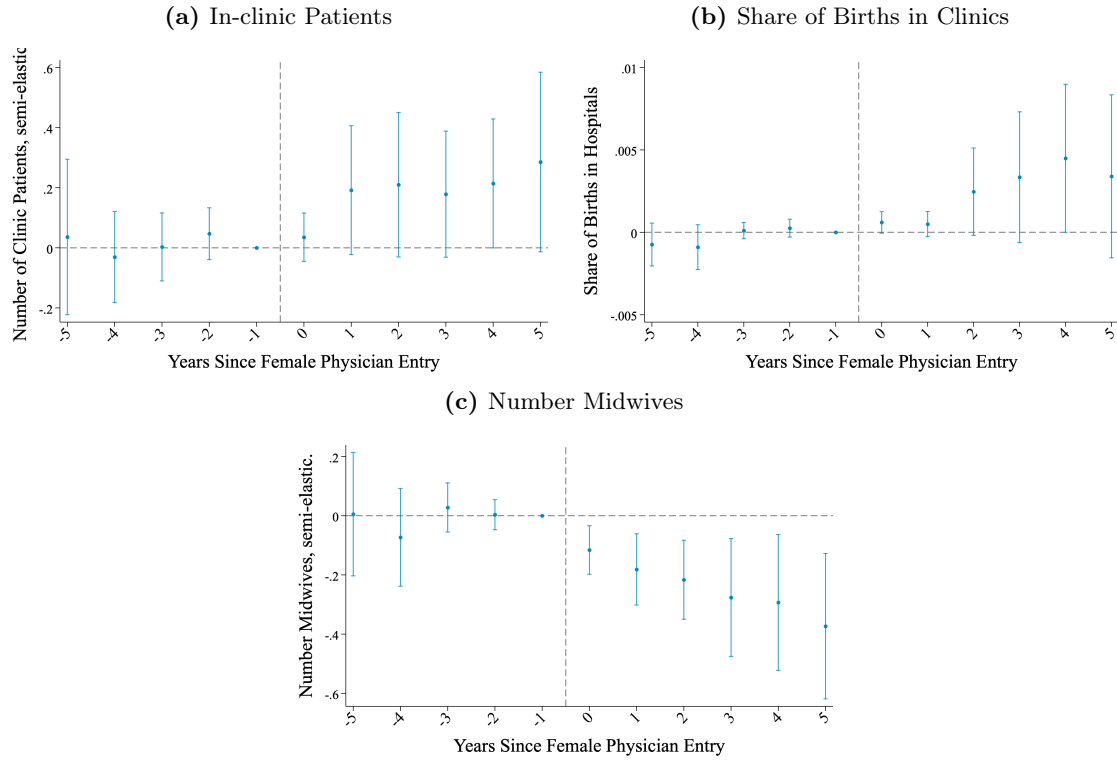


Notes: Panel (a) reports event-study estimates for 41 major cities located in districts that eventually hired a female physician. Unlike the main analysis, this placebo exercise is not restricted to first entries classified as replacement hires. Treatment timing is defined by the first female physician entry in the surrounding *Zemstvo* district. Panel (b) expands the main analysis to all first female physician entries, including entries associated with expansions in physician employment. Panel (c) restricts the main sample to districts with no net increase in physician employment between years $t - 2$ and t .

D.5 Changes in Healthcare Provision

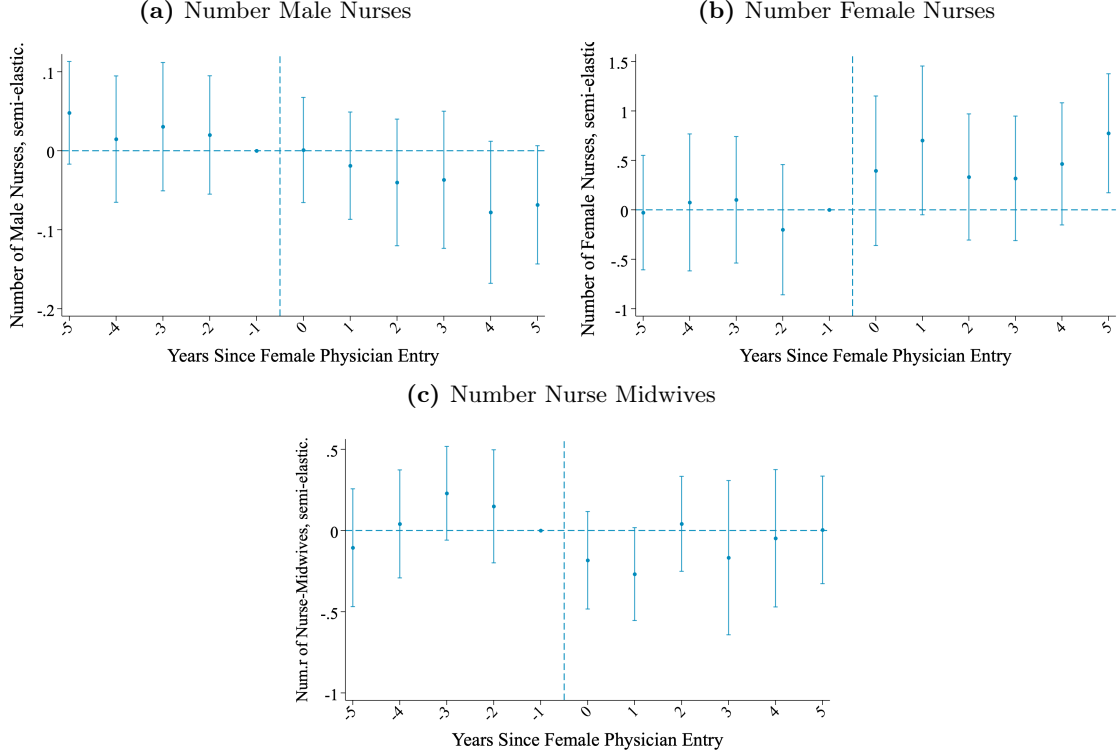
D.5.1 Event Studies

Figure D.5.1: Event Studies to Table 5, Panel A



Notes: Sample and estimation as specified in the notes to Table 5.

Figure D.5.2: Event Studies to Table 5, Panel B



E Conceptual Framework

Long-Run Talent Selection with Symmetric Gender-Concordance Preferences Consider a setting in which male and female physicians are drawn from identical talent distributions, and both male and female patients exhibit the same preference for being treated by a physician of their own gender. Let $p_{f,f} = p_{m,m}$ and $p_{f,m} = p_{m,f}$ denote the probabilities of seeking care when treated by a gender-concordant versus gender-discordant doctor.

Assume a representative catchment area with exactly one physician, and equal numbers of local male and female patients: $N_f = N_m = N$. In this single-physician local market, expected patient demand for the assigned physician depends only on concordance probabilities and population composition.

The demand for female physicians is:

$$D_f = N_f \cdot p_{f,f} + N_m \cdot p_{m,f}$$

and for male physicians:

$$D_m = N_f \cdot p_{f,m} + N_m \cdot p_{m,m}$$

Given the symmetry assumption $p_{f,f} = p_{m,m}$ and $p_{f,m} = p_{m,f}$, it follows that:

$$D_f = D_m$$

In long-run equilibrium, marginal entrants of each gender must receive the same payoff from entering the physician workforce. If $\tau_f(\theta_f)$ and $\tau_m(\theta_m)$ denote the treatment effectiveness of the threshold (marginal) female and male physicians, respectively, then:

$$\tau_f(\theta_f) \cdot D_f = \tau_m(\theta_m) \cdot D_m$$

Since $D_f = D_m$, we must have $\tau_f(\theta_f) = \tau_m(\theta_m)$, implying:

$$\theta_f = \theta_m$$

However, in the short run, when women face historical entry barriers and only highly talented women enter, we expect $\theta_f > \theta_m$ and thus $\tau_f(\theta_f) > \tau_m(\theta_m)$, producing an observable short-run performance advantage for female physicians. Over time, as barriers to entry disappear, this talent gap will close. The common threshold θ^* must exceed the original male-only threshold $\theta_m^{(0)}$. With total patient demand unchanged, the entry of women expands the pool of potential physicians, allowing patients to be served by a more selective set of providers. Female entry therefore raises the minimum talent required of subsequent entrants and increases average physician quality in equilibrium.

Male-Favoring Asymmetric Gender-Concordance Preferences Now consider a setting where patients exhibit stronger concordance preferences for male physicians. Specifically, assume:

$$p_{m,m} > p_{f,f} \quad \text{and} \quad p_{f,m} = p_{m,f}$$

This implies that male physicians receive more patient demand than female physicians in equilibrium. Formally, the expected demand is:

$$D_f = N_f \cdot p_{f,f} + N_m \cdot p_{m,f}$$

$$D_m = N_f \cdot p_{f,m} + N_m \cdot p_{m,m}$$

With $p_{m,m} > p_{f,f}$ and symmetric discordance terms ($p_{f,m} = p_{m,f}$), we have:

$$D_m > D_f$$

To satisfy the zero-profit equilibrium condition:

$$\tau_f(\theta_f) \cdot D_f = \tau_m(\theta_m) \cdot D_m$$

It must be that:

$$\tau_f(\theta_f) > \tau_m(\theta_m) \quad \Rightarrow \quad \theta_f > \theta_m$$

Thus, even in the long run, female physicians must be positively selected relative to male physicians in order to compensate for lower patient demand, resulting in a persistent talent advantage among women in equilibrium.

Female-Favoring Asymmetric Gender-Concordance Preferences Finally, consider a setting where female patients exhibit stronger concordance preferences for female physicians. Specifically, assume:

$$p_{f,f} > p_{m,m} \quad \text{and} \quad p_{f,m} = p_{m,f}$$

This implies that female physicians receive more patient demand than male physicians in equilibrium. Expected patient demand becomes:

$$D_f = N_f \cdot p_{f,f} + N_m \cdot p_{m,f}$$

$$D_m = N_f \cdot p_{f,m} + N_m \cdot p_{m,m}$$

Given $p_{f,f} > p_{m,m}$ and symmetric discordant probabilities, we have:

$$D_f > D_m$$

To satisfy the zero-profit equilibrium condition:

$$\tau_f(\theta_f) \cdot D_f = \tau_m(\theta_m) \cdot D_m$$

This implies:

$$\tau_f(\theta_f) < \tau_m(\theta_m) \quad \Rightarrow \quad \theta_f < \theta_m$$

Thus, in contrast to the male-favoring case, marginal female physicians can be of lower talent than marginal male physicians and still enter the workforce, due to greater expected demand. Over time, this generates a persistent equilibrium where female physicians are, on average, less positively selected than male physicians. This would predict convergence - and potentially reversal - of the short-run performance advantage observed when female physicians first enter the market.

Summary of Equilibrium Predictions These three cases yield distinct equilibrium predictions for the composition and quality of the physician workforce:

- **Symmetric Preferences:** Equal demand across genders leads to equal talent thresholds in equilibrium, which are higher than the historical male-only threshold, improving overall physician quality.
- **Male-Favoring Preferences:** Male physicians attract greater demand, so marginal female physicians must be more skilled to enter, generating a persistent positive selection advantage among women.
- **Female-Favoring Preferences:** Female physicians attract more demand, allowing marginal female entrants to have lower talent levels than men, leading to eventual convergence or reversal of any initial female advantage.

E.1 Physician Value-Added Estimation

This appendix describes the construction of physician-level value-added estimates used in Section 6. I estimate value added using the linked physician panel for 1889–1910, restricted to physicians employed in *Zemstvo* positions. Because individual physicians cannot be linked to individual patients, the procedure follows the logic of value-added estimators that use group-level outcomes, as in Biasi [2021]. The goal is to estimate individual physicians’ contributions to mortality outcomes using variation in district-level mortality associated with their presence, entry, exit, and movement across districts.

Define y_{dt} as the mortality outcome in district d and year t . I first residualize the outcome with respect to district and year fixed effects

$$y_{dt} = \alpha_d + \lambda_t + \varepsilon_{dt}, \quad (10)$$

where α_d captures persistent differences across districts and λ_t captures common shocks across years. The residual ε_{dt} measures the difference between observed mortality in district d and year t and the level predicted by district and year effects.

I first divide each district-year residual equally among the N_{dt} physicians practicing in that district and year. The resulting physician-year residual allocation is:

$$r_{pdt} = \frac{\varepsilon_{dt}}{N_{dt}}. \quad (11)$$

I then divide each physician’s career into spells, defined as consecutive years of employment in the same district. A change in district or a gap in observed employment marks the beginning of a new

spell. Within each spell, I average the annual residual allocations:

$$\bar{r}_{ps} = \frac{1}{T_{ps}} \sum_{t \in s} r_{pdt}, \quad (12)$$

where T_{ps} denotes the number of years physician p is observed in spell s . This averaging reduces year-to-year noise in the residual allocations.

Finally, I construct physician value added using a leave-one-out average across spells:

$$VA_{p,-s} = \frac{1}{S_p - 1} \sum_{s' \neq s} \bar{r}_{ps'}, \quad (13)$$

where S_p is the number of observed spells for physician p . Thus, the value-added estimate assigned to a physician in a given spell draws exclusively on mortality outcomes from that physician's other spells.

For physicians observed in only one spell, I cannot construct the leave-one-out value added. For these physicians, value added is identified from the variation in district-level mortality residuals around physician entry into and exit from the district. This variation provides weaker identification than explicit cross-spell leave-one-out comparisons and may be more affected by the district fixed effect when a spell spans much of the district's observation period. The concern is smaller for shorter spells. In the data, spells last 3.5 years on average, or approximately 16% of the 1889-1910 panel.

I then apply empirical Bayes shrinkage, following [Kane and Staiger \[2008\]](#), to account for differences in the precision of physician value-added estimates. Estimates with greater sampling uncertainty are pulled more strongly toward zero, while more precisely estimated effects receive less adjustment.

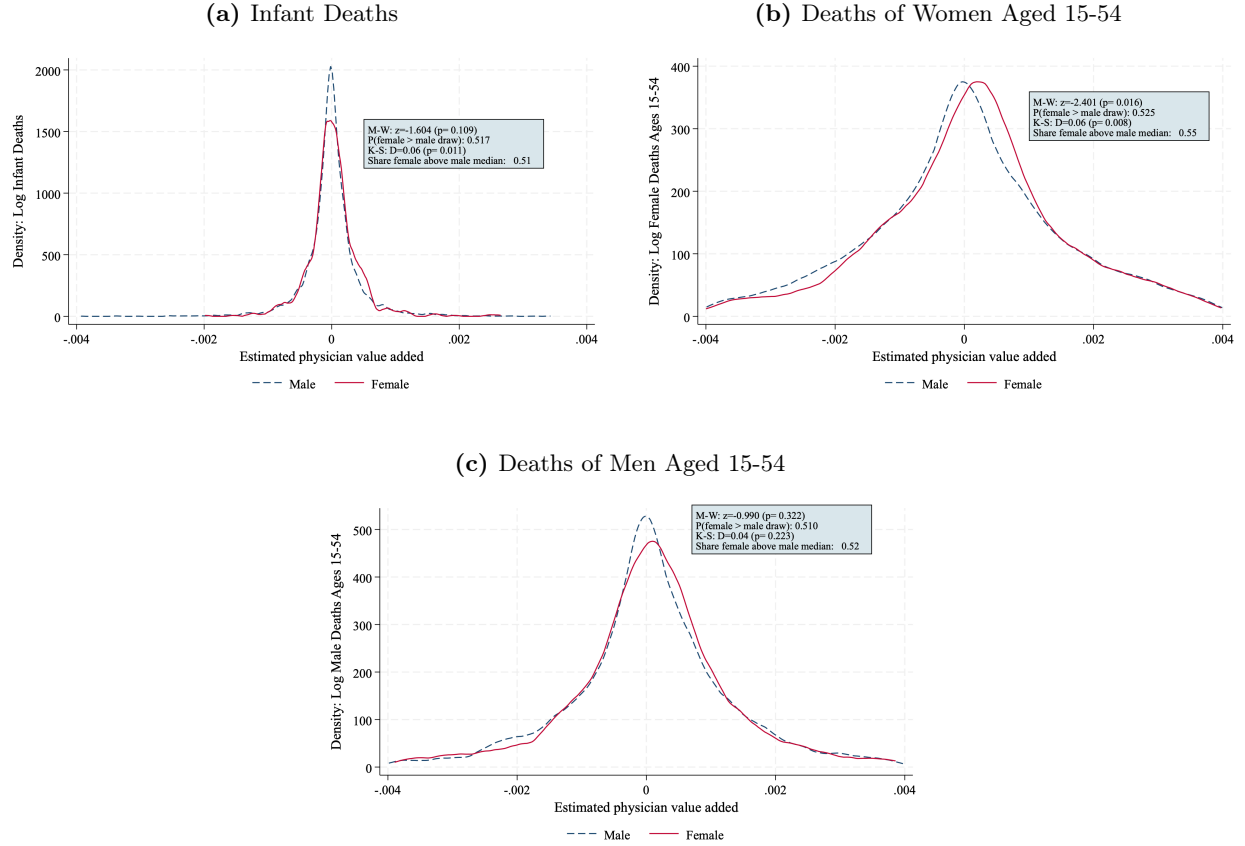
I construct these measures separately for the different mortality outcomes. The value added estimates are multiplied by -1 so that higher value added corresponds to lower mortality. The resulting physician-level estimates are used to compare the value-added distributions of male and female physicians and to examine how these differences evolve across graduation cohorts. In total, I estimate value added for 8,430 unique physicians, including 7,602 male physicians and 828 female physicians, over the 1889-1910 period. The results are presented in Appendix Figure [E.1.1](#).

Validation I validate the value-added estimator using simulated data in which true physician quality is known. The data-generating process follows the model implicit in the estimator. Log infant deaths in district d and year t equal district fixed effects, year fixed effects, the mean value added of physicians practicing in d at time t , and a district-year idiosyncratic shock. Formally,

$$y_{dt} = \alpha_d + \gamma_t + \bar{\mu}_{dt} + \varepsilon_{dt},$$

where $\mu_i \sim \mathcal{N}(0, \sigma_a^2)$ is physician i 's true value added and $\varepsilon_{dt} \sim \mathcal{N}(0, \sigma_e^2)$ is a district-year idiosyncratic shock.

Figure E.1.1: Value added of Male and Female *Zemstvo* Physicians, 1889-1910

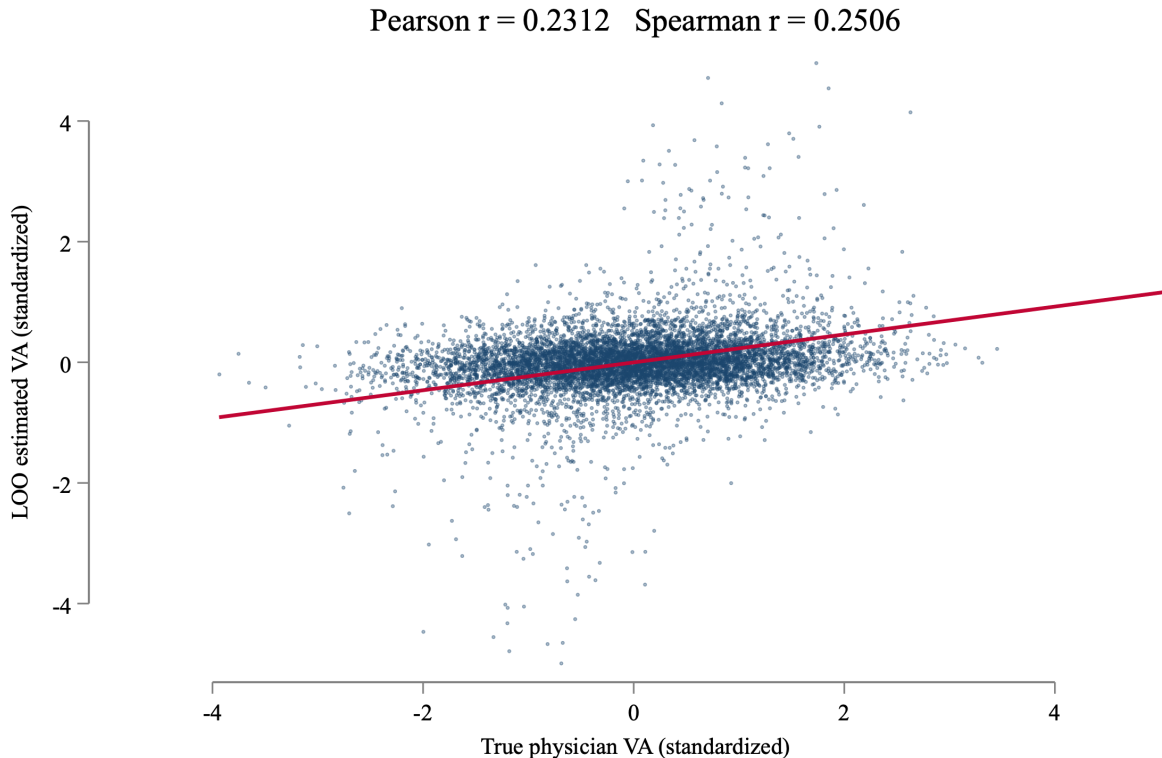


Notes: The figure plots kernel density estimates of physician value added separately for male and female physicians using the Epanechnikov kernel function. Value added is estimated from residualized district-year mortality measured as the log number of infant, 15–54-year-old male, and 15–54-year-old female deaths, using the linked physician panel for 1889–1910 [Matiashevili, 2026]. District and year fixed effects are first removed from the outcome. The remaining district-year residual is allocated equally across all physicians practicing in that district-year, following Biasi [2021]. Physician value added is constructed as the leave-one-out mean of spell-level residual allocations, where a spell is a contiguous run of years in the same district. For physicians with a single spell, value added is identified by entry into and exit from the district. Estimates are shrunk toward zero using cohort-specific empirical Bayes shrinkage following the Kane and Staiger [2008] approach. Higher values indicate lower mortality. The x-axis is truncated for exposition. The Mann-Whitney z-statistic tests the null hypothesis that a randomly drawn female physician is equally likely to rank above or below a randomly drawn male physician in the value-added distribution. The reported probability is the share of all possible female-male pairs in which the female physician has higher estimated value added. A value of 0.5 indicates no difference. The Kolmogorov–Smirnov statistic tests equality of the male and female value-added distributions. The reported share indicates the fraction of female physicians whose estimated value added exceeds the median value added among male physicians.

Parameters are calibrated to match the actual data. I simulate 8,000 physicians across 200 districts and over 22 years, matching the observed spell structure (67% of physicians observed in a single district spell, 33% in two or more, mean spell length 3.5 years) and physician density. The noise parameter $\sigma_e = 0.02$ is set so that the root mean squared error of the district-year regression matches the observed value of 0.158 log points. True physician value added is drawn with standard deviation $\sigma_a = 0.44$, implied by the empirical Bayes signal estimate from the actual data.

I then apply the leave-one-out estimator exactly as in the main analysis, residualizing log infant deaths on district and year fixed effects. I then allocate the residual equally across co-located physicians, define career spells, and compute leave-one-out averages across spells. Comparing estimated to true physician value added, I recover a Pearson correlation of 0.23 and a Spearman rank correlation of 0.25 (Appendix Figure E.1.2). The positive Pearson and Spearman correlations indicate that, under the calibrated data-generating process, the estimator contains meaningful but noisy information about underlying physician quality.

Figure E.1.2: Simulating Value-Added Estimation



Notes: The figure plots standardized leave-one-out estimated physician value added against standardized true value added for 8,000 simulated physicians across 200 districts over 22 years. The data-generating process follows $y_{dt} = \alpha_d + \gamma_t + \bar{\mu}_{dt} + \varepsilon_{dt}$, where $\mu_i \sim \mathcal{N}(0, \sigma_a^2)$ is physician i 's true value added, $\bar{\mu}_{dt}$ is the mean value added of physicians active in district d at time t , and $\varepsilon_{dt} \sim \mathcal{N}(0, \sigma_e^2)$ is a district-year shock. Parameters are calibrated to the data: $\sigma_a = 0.44$ (implied by the empirical Bayes signal estimate), $\sigma_e = 0.02$ (set to match the observed root mean squared error of 0.158 log points), and the physician career structure matches the observed spell distribution (67% single-spell, mean spell length 3.5 years). The LOO estimator is applied as in the main analysis. The red line is a linear fit estimated on the full sample. The scatter is displayed for observations within ± 5 standard deviations on both axes. Pearson and Spearman correlations are computed on the full sample.

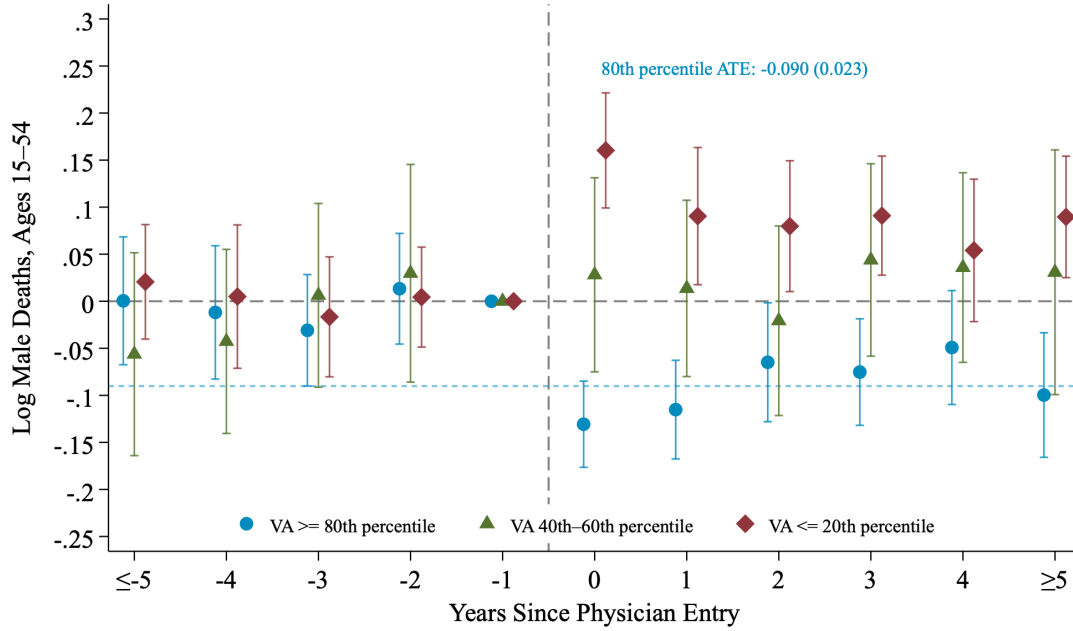
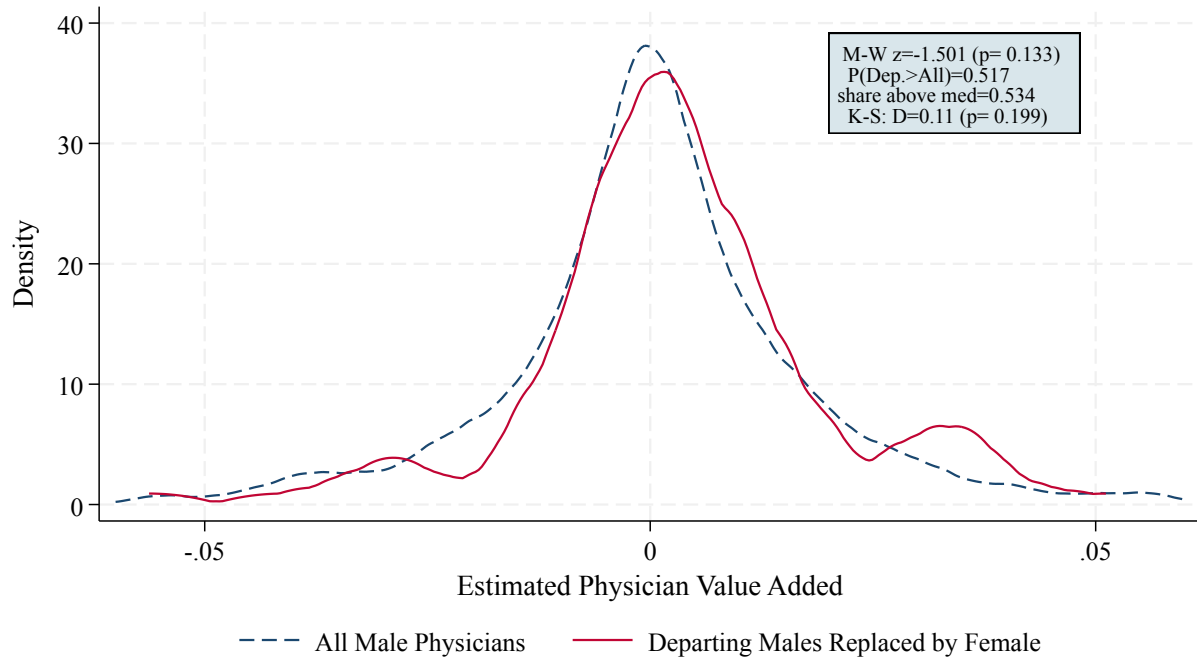


Figure E.1.3: Mortality Around Top-Quintile Physician Entry

Notes: The figure plots event-study estimates of log male deaths aged 15-54 around the first observed entry of a *Zemstvo* physician in three groups of estimated value added: the top quintile (at or above the 80th percentile), the third quintile (40th-60th percentiles), and the bottom quintile (at or below the 20th percentile). For each group, entry is defined as the first observed arrival of a physician in that value-added range. Entries are restricted to replacements, defined by an unchanged total number of *Zemstvo* physicians from the preceding year. Regressions include district and year fixed effects and control for the number of physicians. Year -1 is the reference period; the endpoints pool years ≤ -5 and ≥ 5 . Bars indicate 95% confidence intervals, with standard errors clustered by district. The annotation reports the post-entry coefficient and standard error for top-quintile physician entry, estimated on the same sample as the corresponding event study.

E.1.1 Robustness

Figure E.1.4: Departing Male Physicians versus All Male Physicians



Two-sample t-test p-value: 0.34

Notes: The departing male physician is identified as the male *Zemstvo* physician whose last observed year in a replacement district is the year immediately preceding the first female physician entry. Where multiple male physicians departed in the same district-year, value added is averaged across departing physicians. The comparison group comprises all male *Zemstvo* physicians with an estimated value added, regardless of whether they departed. I use raw value added estimates here, given the small sample size of the departing male physicians, by construction.